2024

Functional Family Therapy: A Multi-Year Statewide Program Evaluation

Natalie Finn
Virginia Commonwealth University

Functional Family Therapy: A Multi-Year Statewide Program Evaluation

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University

by
Natalie Finn

M.S., Virginia Commonwealth University, 2019
B.S., Cornell University, 2013

Dissertation Chair:
Michael Southam-Gerow, Ph.D.
Professor, Department of Psychology

Virginia Commonwealth University
Richmond, Virginia

Dissertation approved by committee: September 27, 2024

Acknowledgements

I would like to thank my dissertation chair Dr. Michael Southam-Gerow for his support and guidance throughout this project and my graduate training at VCU. I would also like to thank my committee for their valuable feedback. I would like to thank my classmates for all their help and inspiration throughout graduate school, as well as my friends and family for their continued encouragement.

Table of Contents

Abstract

FUNCTIONAL FAMILY THERAPY: A MULTI-YEAR STATEWIDE PROGRAM EVALUATION

By: Natalie Finn, M.S.

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2024

Major Director: Michael Southam-Gerow, Ph.D., Professor, Department of Psychology

Functional Family Therapy (FFT) is a family-based treatment program for youth with externalizing behavior problems, often involved in juvenile justice and child welfare systems. FFT is an evidence-based program and has been implemented across many states, including the state of Virginia. **Objective.** The present study performed a program evaluation and observational study of the implementation and sustainment of FFT in Virginia, during the first five years of implementation. **Method.** Participants included all youth who participated in FFT in Virginia between October 2017 and June 2022. Secondary data analysis was conducted using data entered by therapists and supervisors into the FFT data tracking system. Data related to the quality of implementation, including team size, average caseload, treatment duration, fidelity, treatment completion, and youth placement status at termination, were compiled to assess trends over time. The composition of participants in the FFT program is described and differences in outcomes based on race/ethnicity, gender, referral source, primary referral reason, age, and

participation pre- or post- the start of the COVID-19 pandemic were examined. Modifications to FFT were identified and characterized. **Results.** At the end of the study time period, all teams were still providing FFT. Trends in implementation and sustainment, participant differences in completion rate and outcomes, and modifications to FFT that occurred during the study period are discussed.

List of Tables

List of Figures

Introduction

Mental health problems are widespread among youth, with externalizing problems being among the most common and impairing (Bronsard et al., 2016, Underwood & Washington, 2016). Youth with these problems and their families are often involved a web of state and local systems (e.g., Department of Social Services (DSS), Department of Juvenile Justice (DJJ); Bronsard et al., 2016, Underwood & Washington, 2016). Considerable progress has been made in identifying interventions that can improve functioning among these youth (McCart & Sheidow, 2016; Epstein et al., 2015). Among these, family-based therapies are common (Henggeler & Sheidow, 2012; Baldwin, et al., 2012; McCart & Sheidow, 2016). Such approaches are complex and ensuring fidelity to the models represents a major focus of training and ongoing quality assurance/quality improvement efforts (Hogue et al., 2013; Wandersman, Chien, & Katz, 2012).

The present study is focused on one family-based model, Functional Family Therapy (FFT; Alexander & Parsons, 1982; Alexander et al., 2013), in the context of a multi-year statewide effort to improve capacity to help youth and their families in the juvenile justice and child welfare systems. The study addresses questions related to quality of implementation, across the first 5 years of implementation, describes the participant population served by FFT in Virginia, examines differences in outcomes, and identifies modifications to FFT in Virginia, including a modification in response to the Covid-19 pandemic, to allow for FFT to be delivered via teletherapy.

Specifically, the study will answer the following questions related to quality of implementation, FFT participation, and modifications to FFT in Virginia: (1) *Quality of implementation:* How does the implementation of FFT in Virginia compare to national

benchmark standards of implementation of FFT and what, if any, quality assurance targets would benefit from focused quality improvement efforts? Given the organizational and individual impacts of the Covid-19 pandemic, how was the quality of implementation of FFT in Virginia impacted by the Covid-19 pandemic? (2) *Participation:* Given the applicability of FFT to multiple populations, (for example, youth and families involved in the child welfare, juvenile justice, and public mental health systems), who actually participated in FFT in its first five years of implementation, and did completion rates and outcomes differ for these groups? Additionally, did completion rates and outcomes differ based on race/ethnicity, gender, age, or primary referral reason? (3) *Modifications:* Modifications are a common, and sometimes expected facet of implementing evidence-based practices in real-world settings. For example, previous research indicated the feasibility of FFT delivered via teletherapy, in response to the Covid-19 pandemic. In the present study, did this shift to teletherapy impact completion rates, youth outcomes, treatment duration, and number of sessions youth completed? What, if any, additional modifications to FFT were identified during the first five years of implementation of FFT in Virginia?

Before describing the method in detail, this paper will review past work related to (a) family-based approaches to treatment of youth externalizing behaviors, (b) research on FFT effectiveness and fidelity, and (c) frameworks and methods for describing and evaluating the implementation and sustainment of evidence-based practices, like FFT. Then, the specific aims and research questions will be stated.

Family Treatments for Youth Externalizing Behavior

Over 20% of adolescents in the United States experience at least one mental health problem by the time they are 18 years old (Merikangas et al., 2010; US Department of Health

and Human Services, and Centers for Disease Control and Prevention, 2013). Externalizing behaviors, including conduct disorders, are the most common form of mental health problem in children and adolescents and are characterized by repetitive and persistent patterns of antisocial, aggressive, or defiant behavior that exceed age-appropriate expectations (American Psychiatric Association, 2013; NICE, 2013). It is estimated that 5–10% of children and adolescents display externalizing behaviors that are clinically significant and persistent in duration (Hill, 2002; Tonge, 2007). Externalizing behaviors are known to impact a youth's life in a number of ways, including affecting attendance at school, educational achievement and future health and life chances (Liu, 2004; Colman et al., 2009).

The nation's child welfare and juvenile justice systems were built to address specific issues: abuse, neglect, and serious delinquency (Merkel-Holguin, 2017; Bartollas, 2010). Today, many youth are placed in these systems for "challenging behaviors," such as defying their parents, being truant from school, running away, abusing alcohol and drugs or engaging in risky sexual or other activities that threaten their well-being or safety (Annie E. Casey Foundation, 2015). Externalizing behavior disorders such as disruptive and conduct disorders are the most common mental health diagnoses among youth involved with the child welfare and juvenile justice systems (Bronsard et al., 2016, Underwood & Washington, 2016). Out of home placements for these youth are expensive, and placement in residential treatment and incarceration, have been shown to be ineffective (Henggeler & Schoenwald, 2011; Annie E. Casey, 2012).

To address the inadequacies of out of home placements, efforts have shifted towards funding prevention and intervention efforts that keep children in their homes and communities. For example, in FY2018, Virginia expended $145.6 million in federal (Medicaid and Children's

Health Insurance Program) and state funds on Intensive Home-Based Treatment services (Virginia Department of Medical Assistance Services 2018), also referred to as Intensive In-Home services. One challenge that has received recent attention, is that these Intensive Home-Based Treatment services generally lack quality standards or consistent components, and therefore can vary greatly in quality and outcomes (Bruns et al., 2021).

As an alternative to unspecified home and/or community-based services, a focus has shifted towards evidence-based programs (Robbins et al., 2016; Zajac, Randall, & Swenson, 2015). Over the last several decades, a number of evidence-based programs for externalizing youth have been developed and shown to be effective (Goense et al, 2014; Kemp, Boxer, & Frick, 2020). While it is acknowledged that there are multiple, and often inter-related factors associated with the onset and maintenance of externalizing behaviors, at the level of the individual, family, community (e.g., school and peer groups) and society (Liu, 2004), the role of the family is considered to be central (Price, Chiapa, & Walsh, 2013). Given evidence on the important role of family, a range of interventions, both preventative and treatment-focused, have been developed that focus on the family system to affect positive outcomes for adolescents (McCart and Sheidow, 2017; Sheidow, McCart, & Drazdowski, 2022).

Some evidence-based family programs for externalizing behavior are focused closely on family member interactions while others are more broadly focused on the family within a larger ecosystem of influences (Baldwin et al., 2012; Hogue et al., 2019). For example, Brief Strategic Family Therapy (BSFT; Szapocznik, Scopetta, & King, 1978) draws on the work of Minuchin (Minuchin, 1974) and focuses on diagnosing family interactional patterns and restructuring (i.e., changing) the family interactions associated with the adolescent's problem behaviors. Another treatment program, Multisystemic Therapy (Henggeler, 1998), provides 24/7 access to therapists

and focuses on cognitive and/or behavior change, communication skills, parenting skills, family relations, peer relations, school performance, and social networks. As such, MST takes a more social ecological (Bronfenbrenner, 1979) perspective, incorporating assessment and intervention at multiple levels.

Functional Family Therapy (FFT)

One widely used family therapy treatment for youth externalizing behavior problems is Functional Family Therapy (FFT) (Alexander, Waldron, Robbins, & Neeb, 2013). FFT is a family-focused treatment intervention for adolescents with externalizing behavioral problems (Alexander & Parsons, 1973) first developed in the US in the 1960s. The model is underpinned by a theoretical assumption that problem behaviors exhibited by adolescents are rooted in dysfunctional family relations (Robbins, Alexander, Turner, & Hollimon, 2016). FFT is a brief, solution focused family therapy intervention that draws upon the family dynamic as a key mechanism for change in reducing adolescent problem behaviors. Goals of FFT include improving family communication, expression of feelings, presentation of demands, and discussion of alternative solutions. FFT usually consists of 12 to 14 therapy sessions delivered over 3 to 5 months by a trained therapist. FFT includes a target population of youth ages 11 to 18 presenting with behavioral and emotional problems. Clients are often referred by juvenile justice, child welfare, and other public service systems, with a focus on preventing out-of-home placements. FFT uses a team structure in which teams are expected to have a minimum of 3 therapists and one supervisor. Each therapist is expected to carry an average of 10-12 cases, with an expected range of 5-15 cases.

Efficacy and Effectiveness of FFT

The evidence for FFT's effects on youth behavioral health has been established by more than 40 years of research, including 10 randomized efficacy trials, five randomized effectiveness studies, and 12 quasi-experimental evaluations (Robbins et al., 2016). One meta-analysis found that FFT has a small aggregate effect on delinquency outcomes (Lee et al., 2012). In a more recent review, Hartnett et al. (2017) found that FFT significantly outperformed randomly assigned control conditions, and both randomly-assigned and non-randomly assigned alternative treatment conditions. FFT is also among the ecological family-based treatments deemed well established for substance abuse (Hogue, et al., 2014).

FFT Treatment Model

Within the FFT model, FFT therapists provide, at a minimum, weekly one-hour, home-based family therapy sessions to clients. The FFT model has five specific phases of treatment that are delivered in order: engagement, motivation, relational assessment, behavior change, and generalization. The *Engagement* phase emphasizes within youth and family factors that protect youth and families from early program dropout. The *Motivation* phase is designed to change maladaptive emotional reactions and beliefs, and increase alliance, trust, hope, and motivation for lasting change. The *Assessment* phase is used to clarify individual, family system, and larger system relationships, especially the interpersonal functions of behavior and how they relate to change techniques. The *Behavior Change* phase consists of communication training, specific tasks and technical aids, parenting skills, contracting and response-cost techniques, and youth compliance and skill building. Finally, during the *Generalization* phase, family case management is guided by individualized family functional needs.

Fidelity

Fidelity, also called treatment integrity, has been defined as delivering a practice as intended by design (Perepletchikova & Kazdin, 2005; Schoenwald et al., 2011). For manualized treatments, this usually means adhering to the treatment manual. Fidelity is often measured by a score or multiple scores, and many evidence-based practices have their own individual fidelity measures, tailored specifically to that practice. Fidelity data can be used in supervision for continuous training and outcome monitoring (McLeod et al., 2013) with the goal of scaling up and sustaining the practice. Additionally, and of growing importance, is the role of fidelity in providing accountability in public mental health systems, including the use of differential reimbursement rates (e.g., pay for performance, incentivized payment for use of evidence-based treatments) (Cox et al., 2019). Maintaining adequate fidelity over time can be particularly challenging, as many purveyor supports are withdrawn and programs are expected to be more self-sufficient after an initial training period (Wiltsey-Stirman, 2012; Birken et al., 2020).

FFT Fidelity

Studies of FFT have demonstrated that therapist fidelity to the FFT model is positively related to outcomes. Two FFT studies have looked at the effects of fidelity on recidivism. Both studies found that higher fidelity scores were associated with lower recidivism. Barnowski (2002) categorized therapists as highly competent, competent, borderline competent, and not competent in their delivery of FFT. They found that youth served by the competent and highly competent FFT therapists had lower felony recidivism rates than those in a randomly assigned control group, while those served by therapists categorized as borderline competent or not competent had higher recidivism than the control group. Sexton and Turner (2010) confirmed these findings with additional follow-up analyses assessing effect size and identifying small effects. Among FFT recipients, fidelity effects were dependent on a measure of negative peer

risk at baseline, with adherence only having an effect in cases with high negative peer risk at baseline (Sexton & Turner, 2010).

A more recent study (Turner et al., 2018) expands on these findings, examining risk-fidelity interactions. They found that among sites with low-risk caseloads, therapists had higher fidelity scores when working with lower-risk clients. The authors also examined therapist experience delivering FFT. In general, therapist fidelity improved as they had more experience with more cases. When therapists had experience with fewer cases, fidelity was lower among high-risk caseloads than low-risk caseloads, while among therapists with more experience with FFT, fidelity was higher among high-risk caseloads. Overall, increases in therapist fidelity and experience were associated with lower recidivism among low-risk caseloads, although effect sizes were minimal.

As FFT has evolved, different measures and methods have been used for assessing fidelity to the FFT model as it is implemented in real world settings. Early studies used the Treatment Adherence Measure (TAM) measure of adherence, rated 0-6 (e.g., Barnowski, 2002; Sexton and Turner, 2010). A newer measure, the Weekly Supervision Checklist fidelity measure, which includes subscale ratings for both adherence and competence has been integrated into an online FFT practice tracking system, the Clinical Services System (CSS). Although this measure, and the use of the CSS overall, does not yet have substantial research on validity and reliability, purveyor-set benchmarks categorize fidelity as adequate (3 or greater out of 6) or below adequate fidelity, and this benchmark is used as part of training and quality improvement. This benchmark has been used to signify adequate fidelity in recent publications (e.g., Olseth et al., 2024; Robbins & Midouhas, 2021). Fidelity rating is expected to occur weekly during group supervision, with each therapist presenting on at least one case, for rating by supervisor. Fidelity

and frequency of fidelity rating will be used as quality assurance targets (discussed in the next section) in the current study, although trends in fidelity scores will not be interpreted, given the absence of adequate validity and reliability support for this measure.

Program Evaluation

Given the growing emphasis on accountability within the behavioral health care field (Berenson, Pronovost, & Krumholz, 2013), evaluation has become a routine aspect of delivering care. As part of program evaluation, it is useful to examine quality of services across teams and at the system level to ensure that the program is implemented as intended. This is particularly important when evaluating evidence-based practices that use a team service delivery format, such as FFT. Inherent to the model, FFT therapists do not operate independently, therapists and supervisors together make up a team/site, which is certified by the purveyor, FFT, LLC. As such, it is of interest to both the purveyor and service system administrators (e.g., DJJ administrators, DSS administrators) to evaluate targets for quality assurance by team, when considering implementation and sustainment of an EBP like FFT. Many funding initiatives also require program evaluation and reporting at the service system level (e.g., state level), such as the Family First Prevention Services Act (U.S. Department of Health and Human Services, 2019).

Maintaining fidelity and other program supports tends to be increasingly difficult as time from initial training increases (Wiltsey-Stirman et al., 2012). EBP purveyors and system administrators alike may be interested in monitoring *quality assurance targets*, beyond fidelity, to assess the continued implementation of an EBP as intended by the developers. The present study will assess FFT quality assurance targets that have been outlined by FFT as standard or benchmarked components of implementation and FFT services, including team size, average caseload, treatment duration, fidelity scores, and frequency of fidelity monitoring.

Implementation Frameworks

Many models and frameworks have been developed to organize and guide implementation science research (e.g., Tabak, Khoong, Chambers, & Brownson, 2012). Some models describe the implementation process in stages, while other models focus on the aspects of stakeholder groups that might be involved, context, characteristics, action steps, or outcomes to be measured.

One such model, The Exploration Preparation Implementation Sustainment (EPIS) model has been proposed as one framework to examine different stages of the implementation process within service systems (Aarons, Hurlburt, & Horwitz, 2011). The framework is divided into four main phases: *Exploration* (assessing needs, considering new practices), *Preparation* (planning and preparing for implementation of a new practice), *Implementation* (delivering the new practice), and *Sustainment* (maintaining this practice with fidelity after initial implementation period ends).

FFT Training includes three phases. Phases I and II fall under the *Implementation* stage of the EPIS framework. Phase I is completed during the first 12-18 months of implementation and involves putting in place an infrastructure at the local FFT site and initial clinical training (FFT LLC, n.d; Robbins & Alexander, 2019; Robbins & Midouhas, 2021). Training includes five onsite trainings and weekly consultation led by an FFT LLC expert via videoconferencing. Therapists are also expected to meet once a week as a group on their own to continue to develop their model skills. Therapists are also trained in how to use FFT's Clinical Services System (CSS) to gather data. The objective for Phase 1 of training is for clinicians to demonstrate strong adherence and high competence in the FFT model by the end of this phase (FFT LLC, n.d.).

Phase II lasts approximately one year and the training model shifts to focus on training the site supervisor, identified at the end of Phase 1. Training for the supervisor in Phase II includes two offsite supervisor trainings, one onsite training, and bi-weekly consultations with the site supervisor (Robbins & Midouhas, 2021). The focus of this phase is on FFT clinical supervision techniques (FFT LLC, n.d). The role of FFT consultants shifts to primarily supporting the supervisor rather than supporting individual clinicians, as the latter becomes the responsibility of the supervisor (FFT LLC, n.d).

Phase III falls under the *Sustainment* stage of the EPIS framework. Phase III, which is ongoing on an annual basis, is focused on the maintenance and/or expansion of FFT after the site has completed the first two phases (FFT LLC, n.d., Robbins & Alexander, 2019). This phase involves a monthly meeting with the site supervisor and an annual site visit (Robbins & Midouhas, 2021). In this stage, the FFT consultants support the site with activities related to fidelity, inter-agency linking, leadership, and expansion of FFT (Robbins & Midouhas, 2021; FFT LLC, n.d.).

In addition to considering *stages* of implementation, it can be helpful to consider frameworks that guide evaluation of specific aspects of implementation. One such framework is the Reach, Effectiveness, Adoption, Implementation, and Maintenance (RE-AIM) framework (Glasgow, Vogt & Boles, 1999; Glasgow & Estabrooks, 2018). In this model, *Reach* refers to the absolute number and proportion of eligible individuals (e.g., clients, students, patients) who receive an intervention, as well as their representativeness compared to the total target population who could receive it (often compared to those who are invited but decline participation). *Effectiveness* is whether an intervention "works" and leads to targeted outcomes, produces generalization effects (e.g., broader outcomes including quality of life), unintended

consequences, variation in outcomes across subgroups, and multilevel effects (e.g., individual-level changes catalyzing systems-level changes). *Adoption* focuses on the proportion of settings or providers that adopt the intervention. *Implementation* refers to the extent to which the intervention is implemented as intended. This often includes a focus on fidelity to the model and adaptation or modification to the model. *Maintenance*, similar to Sustainment (discussed earlier), is a focus on the long-term effects of a program and/or the extent to which a program is sustained over time.

Modification

The RE-AIM framework specifies that within the implementation dimension, it may be important to identify modifications or adaptations to a program. Modification is an umbrella term that refers to changes made to interventions in reaction to unanticipated challenges to the content or the context of a practice (Wiltsey Stirman et al., 2019). One form of modification is adaptations, which are changes that are planned deliberately and proactively. Alternatively, modifications can occur naturally as part of implementation or sustainment in response (Wiltsey Stirman et al., 2019). Within the field of dissemination and implementation science, some have viewed modification or adaptation as a negative outcome that is inconsistent with fidelity, while others have highlighted modification or adaptation as a necessary aspect of implementation for successfully transporting EBPs into real world settings (Baumann et al., 2017).

Wiltsey Stirman and colleagues (2013; 2019) have developed and updated the Framework for Reporting Adaptations and Modifications-Enhanced (FRAME) in order to characterize modifications to interventions. Overall, FRAME includes the following eight aspects: (1) when and how in the implementation process the modification was made, (2) whether the modification was planned/proactive or unplanned/reactive, (3) who determined that

the modification should be made, (4) what is modified, (5) at what level of delivery the modification is made, (6) type or nature of context or content-level modifications, (7) the extent to which the modification is fidelity-consistent, and (8) the reasons for the modification, including (a) the intent or goal of the modification (e.g., improve fit, adapt to a different culture, reduce costs, etc.) and (b) contextual factors that influenced the decision.

The framework includes four primary types of modifications (e.g., what is modified): Content, Contextual, Training and Evaluation, and Implementation and Scale-up Activities. Content modification refers to changes to the content itself or that impact how aspects of the treatment are delivered. Contextual modifications refer to modifications made to the way the overall treatment is delivered. Training and Evaluation modifications refer to changes in how staff are trained or how the intervention is evaluated. Modifications to Implementation and Scale-up Activities are modifications to strategies used to implement or spread the intervention. These types of modifications can be made at various levels of delivery, for example, the individual client level, organization level, or system level. Contextual modifications can be made to various aspects of the intervention context, including format, setting, personnel and population.

Present Study

The present study is a program evaluation and observational study of the implementation and sustainment of FFT in Virginia from 2017-2022. The study is grounded in the EPIS framework of implementation (Aarons, Hurlburt, & Horwitz, 2011), evaluating FFT quality assurance targets over 4.75 years of FFT implementation in Virginia, as well as the RE-AIM framework (Glasgow et al., 1999; Glasgow & Estabrooks, 2018), examining the participant population served by FFT, as well as comparing outcomes across participants, and identifying

modifications to FFT that occurred during the study time period. FFT national standards for quality assurance (Robbins & Midouhas, 2021) were used as a benchmark to assess implementation and sustainment in Virginia. The study provides quality improvement recommendations based on findings and situates findings within the context of FFT and the implementation science literature.

Rationale and Significance of the Current Study

FFT has been implemented in Virginia for over five years, but there have not been statewide consistent quality assurance efforts to examine implementation stability over time. With opportunities to continue to fund FFT in Virginia, via initiatives such as the Family First Prevention Services Act, it is important to assess the quality of implementation over time, understand the population who has received FFT so far, assess outcomes, and identify and characterize any modifications to FFT that have occurred in Virginia.

As statewide initiatives to roll out evidence-based practices systemwide become more common, there is a shift away from singular focus on effectiveness, and more need for focus on program evaluation and quality assurance (Sedlar et al., 2017). The implementation of an evidence-based practice within a statewide service system can be evaluated to examine the impact on clients, as well as evaluating the "health" of the program within the system (e.g., sustainment of teams providing the service after initial training, cost effectiveness; Lau & Brookman-Frazee, 2015). Program evaluation often assesses process and outcomes of programs at a system level to inform quality assurance and improvement efforts. One common method for quality assurance and improvement is to compare quality assurance components to standards or benchmarks put forth by the program purveyor (e.g., national benchmarking; Hogue et al., 2013). For example, a recently published evaluation of FFT assessed the modification of the delivery

format of FFT to teletherapy format during the Covid-19 pandemic (Robbins & Midouhas, 2021). The authors compared quality assurance targets during a teletherapy modification period to the FFT national standard benchmarks and found similar completion rates and therapist-reported outcomes.

The present study will focus on several constructs grounded in the RE-AIM framework. First, the study will examine *implementation* by evaluating quality assurance targets against FFT benchmark standards and identify modifications to FFT as implemented in Virginia. Although fidelity is often a focus of the *implementation* dimension of RE-AIM, the present study will be limited in examination of fidelity, due to the fidelity measure used and the frequency of measurement for the present study. The study will report on fidelity scores but not interpret trends in results. The study will also report on frequency of fidelity ratings and compare against FFT program expectations for fidelity ratings to occur weekly. In addition to fidelity scores and frequency of fidelity rating, the present study will assess several quality assurance targets that compose standard implementation of FFT.

Evidence-based practice operates under the theory that maintaining fidelity to an evidence-based intervention is important to producing intended outcomes for the program. While this often focuses on clinical fidelity to the model (e.g., what is delivered in session), this paradigm also can be applied to aspects of service delivery and program structure for the program. For example, for FFT, to achieve intended outcomes, it is likely important that service delivery and program structure are implemented as intended by the program. There are many aspects of service delivery and program structure that make up FFT. The present study will focus on *quality assurance targets*, or aspects of FFT that have been considered standard aspects of program structure and service delivery (e.g., team size, average caseload, duration of services).

As discussed earlier, FFT requires time and financial resources to implement. By assessing quality assurance targets and developing any applicable quality improvement recommendations, the present study can aid in ensuring that FFT is implemented as intended, supporting the longevity of the program and protecting investments, and supporting intended outcomes for participants.

As part of assessing *implementation*, the present study will also identify and characterize *modifications* to FFT in Virginia. Although there has been an increased focus on modifications and adaptations in the last decade, they are often not documented and therefore not evaluated (Wiltsey Stirman et al., 2019). Research on modifications generally has tended to include *planned* adaptations, with less research focused on highlighting modifications as they naturally occur during real world implementation of EBPs (Wiltsey Stirman et al., 2017). Tracking and understanding modifications is important to implementation science, as implementation is a dynamic process and may differ across settings. In the present study, modifications to FFT will be characterized by the eight components of FRAME (Wiltsey Stirman et al., 2019) discussed earlier, as data available to the current study allows. It is important to characterize modifications to FFT as implemented in Virginia because modifications may occur for a variety of purposes and may impact implementation outcomes (e.g., accessibility, feasibility) and/or client outcomes. Documenting and understanding modifications can shine light on systemic issues that might be impacting FFT in Virginia and can set the groundwork for future research to empirically test modifications as intentional adaptations.

Second, the present study will describe one component of *Reach* (of several components included in the RE-AIM framework): the absolute number of participants who received FFT. Demographic characteristics will be examined descriptively to understand the participant

composition in Virginia (e.g., *describe* who participated). FFT is an intervention that may be applicable to a wide range of participants across multiple referral sources, reasons for referral, genders, race/ethnicities, and ages. FFT implementation in Virginia was not rolled out as a research study (e.g., recruiting specific participants) and as such, program evaluation will help program administrators understand who is actually participating in FFT. This information can inform opportunities to expand the participant population (e.g., through referral efforts), ascertain if FFT is being delivered to any participants outside of the intended participant population (e.g., unintended reach), provide information for potential agencies that might want to be trained as providers of FFT, and help support any efforts to acquire additional funding. Describing the composition of participants also will also allow for the present study to assess outcomes among subgroups, which will be described in the next section.

Third, the study will examine *two* aspects of the RE-AIM dimension, *Effectiveness*: whether the program achieves targeted outcomes and if there is variation in outcomes across subgroups. Recent work in dissemination and implementation has called for a greater focus on "equitable implementation" to address health inequities and promote health equity (Shelton et al., 2024). Health equity can be defined as "the principle underlying a commitment to reduce, and ultimately, eliminate disparities in health and in its determinants, including social determinants" (Braveman, 2014). Efforts have been made in recent years to update RE-AIM to include a focus on equity (Gomes et al., n.d.). Recent guidance specifically calls for assessment across subgroups of participants, such as social determinants, rural or racial/ethnic populations, healthcare setting resources (high or low resourced), or literacy, to demonstrate *who the program benefits* and *where inequities may continue to exist* (Shelton et al., 2020). In the present study, to learn more about who FFT benefits and where inequities may exist, differences in outcomes

among subgroups will be examined. This can help determine if an increased focus on a specific outcome or group is needed for quality improvement in Virginia (for example, developing potential modifications to address below expected performance or disparities in performance among subgroups or increasing the focus on systemic issues that may impact FFT outcomes).

FFT Implementation Context

FFT was initially introduced in the state of Virginia in 2017, as part of Virginia's Juvenile Justice Transformation Plan. A core goal of this plan was to reduce the overuse of juvenile correctional centers and develop a statewide continuum of evidence-based services and alternatives to incarceration as part efforts to replace large correctional settings. As part of this initiative, the Department of Juvenile Justice launched the Regional Service Coordination (RSC) model. Per this model, two large service coordination agencies sub-contract with a network of public and private provider agencies to provide services. FFT and Multisystemic Therapy (MST) were selected as evidence-based practices to implement within this initiative due to their effectiveness in reducing recidivism (Virginia Department of Juvenile Justice, 2017). Prior to the initiative, there were two extant MST teams in Virginia and no FFT teams. In 2017 the DJJ funded the start-up costs to launch four FFT teams, which cost roughly $65,000 per team.

The DJJ and the RSCs have since partnered with the Department of Social Services, Department of Behavioral Health and Developmental Services, Department of Medical Assistance Services (DMAS), and the Office of Children's Services (OCS) to expand the availability of evidence-based services, like FFT, as part of the Family First Prevention Services Act, Project Bravo (Behavioral Health Redesign for Access, Value and Outcomes), and Medicaid Redesign. As a result of these efforts, FFT can now be funded through Children's

Services Act and Title IV-E funds. Since January 1, 2022, these programs are also eligible for Medicaid reimbursement. Teams have been staggered in implementation over time.

FFT implementation in Virginia began in 2017 and persisted throughout the Covid-19 pandemic beginning in 2020. Prior to the pandemic, FFT LLC had permitted the use of teletherapy services (e.g., video conferencing) in exceptional circumstances (e.g., extreme weather). Beginning on March 15, 2020, FFT, LLC, shifted to allow for teletherapy in any session. A 2021 paper describing this shift to teletherapy reported that 62.7% of FFT sessions completed globally between March 15, 2020 and September 1, 2020, were conducted via teletherapy (Robbins & Midouhas, 2021).

In summary, this study will assess FFT implementation at the system (state) and team level and compare quality assurance targets and client outcomes to purveyor benchmarks of national standards. This study will also describe FFT participants descriptively and assess subgroup differences on outcomes. The study places implementation within the context of the Covid-19 pandemic and identifies and categorizes modifications over the 5 years of implementation. As such, the study will address the following aims:

Specific Aims and Research Questions

Aim 1a: Describe system level and team level trends over time on quality assurance targets for FFT, including average team size, average caseload, treatment duration, and frequency of fidelity rating.

Research Questions: How did the quality assurance targets compare to national benchmarks for FFT implementation? Did quality of implementation decline following the onset of the Covid-19 pandemic?

Aim 1b. Assess client outcomes.

Research Question: How did the rate of treatment completion and rates of client status disposition at the time of termination (i.e., client in community, client in school, client without new law violations) compare to national benchmarks for FFT outcomes?

Aim 1c. Determine implementation/sustainment stage, quality of implementation/sustainment, and client outcomes, by team, as of May 2022.

Research Question: At the end of the study period, were teams implementing or sustaining FFT in accordance with FFT national benchmarks? What are potential areas for quality improvement?

Aim 2a: Describe the participant population served by FFT in Virginia.

Research Questions: Who did FFT serve during its first 5 years of implementation (e.g., race/ethnicity, gender, referral reason, referral source, and age).

Aim 2b: Examine differences in outcomes among subgroups of participants.

Research Questions: Did participants differ on completion of FFT and status outcomes (remain in community, engage in school, no new law violations) based on demographics or time of participation? Aim 2 will examine group differences based on (1) race/ethnicity, (2) gender, (3) referral reason, (4) referral source, and (5) whether or not the participant was within or outside the intended age range for FFT. Additionally, were there group differences between participants served one year prior to and participants served during the first year of Covid-19?

Aim 3: Identify and characterize *modifications* to FFT that occurred as part of implementation of FFT in Virginia.

Research Questions: Based on the FRAME (Wiltsey Stirman et al., 2019), how can we characterize modifications to FFT in Virginia. What implications do these modifications have for FFT in Virginia?

Method

Overview

The present study was conducted as part of a statewide evaluation of FFT in Virginia, using secondary analysis of data collected by FFT, LLC through their data system. For Aim 1, statewide trends over time were examined, by team, on adherence to program components (e.g., team size, caseload, duration, frequency of fidelity rating). Teams were compared to FFT national standard benchmarks and teams were categorized based on phase (implementation or sustainment).

For Aim 2, the participant population is described and group differences in completion rate and status outcomes were examined based on demographics and other participant factors. Additionally, for a subset of participants, these variables, as well as duration and number of sessions were examined based on participant time of participation (i.e., pre- or post- Covid-19). For Aim 3, modifications to FFT are identified and are characterized using the FRAME.

Study Procedure

Description of FFT Implementation

Functional Family Therapy was first implemented in Virginia in 2017. Teams have been staggered in implementation over time, and some FFT teams are no longer active. In Virginia, FFT supervisors must have a master's degree; FFT therapists must have at least a bachelor's degree and be registered as a Qualified Mental Health Professional (QMHP). As noted earlier, FFT standards require a minimum of three therapists and one supervisor to compose an FFT team.

Data Collection

FFT therapists entered case information into the Clinical Services System (CSS), a web-based application developed by FFT LLC to support therapists' delivery of the model. Following every clinical staffing, the clinical supervisor completes a fidelity rating for each case reviewed for each therapist. This fidelity rating reflects the degree of clinical adherence and competence for that therapist's work in that case in a specific session presented. Data were acquired from FFT, LLC as part of a statewide fidelity monitoring project conducted by the Center for Evidence-Based Partnerships in Virginia.

Sample

Two teams in Virginia declined the request for deidentified data to be released to researchers by FFT, LLC. The sample included all youth who *participated* in FFT in Virginia provided by nine teams between November 2017 and June 2022, and their therapists. Client outcome data were collected at the close of treatment. There were 56 therapists included over the course of data collection, across nine teams. For each team, team size ranged from one to seven therapists.

Demographics

Client participants were 1,719 youth. Fully 71.6% of the participants were male, 27.6% were female, and 0.8% indicated a gender identity other than male or female. Age of participants ranged from 2 to 22 years old (0.6% participants were younger than the age range intended for FFT, 11 years old, and 11.6% were older than 18), with a mean age of 16 (SD =2). Participant race included 38.6% Black, 36.6% White, 11.9% Latinx, Hispanic, or Chicano, 6.6% Biracial, .3% Middle Eastern, .3% Native Hawaiian or Pacific Islander, 3.5% race unknown, and 1.6% race indicated as other. It is important to note that data was collected by FFT in a "choose one" option manner, and "Latinx, Hispanic, or Chicano" was presented as one choice, although these

are distinct identities. Referral sources included 60.7% Probation, 9.2% Mental Health, 7.9% Child Welfare, 6.6% Parole, 4.0% referral source unknown, 3.9% other referral source, 3.4% Internal Agency Referral, 3.0% School Referral, 0.9% Self-Referral, and 0.5% Diversion. Primary referral reasons included 58.1% Delinquent behavior, 8.1% Youth/Caregiver Conflict, 8.0% Family Violence, 7.2% Mental health issue, 4.8% Family Reunification, 4.0% reason unknown, 3.7% School Truancy/Behavior, 2.3% Family Substance Abuse/use, 1.7% Other reason, 1.2% Runaway Behavior, 0.7% Family Separation, and 0.1% Gang involvement. Demographics are displayed in *Table 1*.

Measures

Aim 1 Measures

Team Size

For each quarter, team size was calculated by adding up the total number of therapists employed by each team. The standard team is comprised of 3-8 therapists.

Team Average Caseload

For each quarter, first a mean caseload was computed for each therapist on the team. Then, a team average caseload was computed by computing the mean of the therapists' mean caseload. The FFT standard average caseload size is 10-12 cases, with the expectation that the range is between 5-15 cases.

Treatment Duration

For each quarter, average treatment duration was computed for cases that closed that quarter. Treatment duration was calculated by number of days between first and last treatment session. The national standard is 60 to 180 days for FFT (Robbins & Midouhas, 2021).

Team Average Fidelity

Team-level average fidelity scores were calculated by averaging weekly therapist ratings, per team, per quarter. Teams are expected to meet weekly for group supervision. Therapist fidelity was assessed by clinical supervisors during weekly group supervision. In a group format, each therapist is expected to present at least one active case, and the supervisor uses a structured fidelity rating tool to assess the extent to which FFT core principles and manual-specific practices were applied when serving the family (Sexton et al., 2004). These ratings are entered into the CSS. As part of supervisor training, supervisors receive training in rating fidelity, which is specific to the family's current treatment phase (e.g., Engagement/Motivation Phase, Behavior Change Phase, or Generalization Phase) and the relevant techniques (interventions) for that phase (Robbins et al., 2011; Robbins & Alexander, 2019). The Weekly Supervision Checklist captures two domains of FFT fidelity, *Adherence* and *Competence*. Adherence refers to the extent of the therapist's adherence to goals, strategies, and theoretical principles of the FFT model, as they fit with the family and treatment phase, as outlined in the supervision manual (Robbins et al., 2011). Supervision guidelines are provided for rating *Adherence* as the extent to which therapist-reported interventions were *appropriate for the case/session and used to achieve phase-based goals*, rated with the following anchors: 0 *"None/Minimal,"* 1" *Occasional/Infrequent,"* 2 *"Regular/Frequent,"* or 3 *"Extensive/Consistent"* (Robbins et al., 2011; Weekly Supervision Checklist, unpublished). *Competence* refers to the quality or skill level (e.g., depth, sophistication, tailored to family) with which therapists deliver FFT interventions. Supervision guidelines are provided for rating *Competence* as the *depth and sophistication* of interventions delivered, rated with the following anchors: 0 *"None,"* 1 *"Low,"* 2 *"Moderate," and* 3 *"High"* (Robbins et al., 2011; Weekly Supervision Checklist, unpublished). The Adherence and Competence scores are then summed to compute a fidelity

rating score which can range 0-6. Scores are entered into the CSS. As stated earlier, although this measure, and the use of the CSS overall, does not have substantial research on validity and reliability, purveyors have set benchmarks to categorize fidelity scores as adequate (3 or greater out of 6) or below adequate fidelity, and this benchmark is used as part of training and quality improvement. For the present study, if a therapist presented more than one case and therefore received more than one fidelity rating for that week, the ratings were averaged to create one score per therapist for the week. Then, for each quarter, a *team-level* average fidelity score was computed for each team by averaging the therapists' weekly fidelity scores across all team members. Only team-level average fidelity scores will be reported in this study.

Frequency of Fidelity Rating

Frequency of fidelity rating was examined to assess adherence to fidelity monitoring procedures as part of weekly supervision. For each team, frequency was a count of the number of weeks per quarter for which the team had at least one fidelity rating for at least one therapist (e.g., did any fidelity rating occur that week). Then this count was divided by the total number of weeks in the quarter to obtain the *percentage of weeks in which the team had completed fidelity rating*, for each quarter.

Completion Rate

For each quarter, completion rate was calculated by the percentage of cases closed that quarter that completed all phases of FFT before closure. FFT national standards range from 70 to 80% completion rate (Robbins & Midouhas, 2021).

Status at Completion

Therapists indicate youth status at case termination on the following: (1) Youth remain in the community and with family, (2) youth are in school or working, and (3) youth have

committed no new law violations at time of case closure. FFT standard is maintaining over 85% success in these categories (i.e., yes responses to each of the three). For each quarter, percentages were calculated for cases closed by team, for these status outcomes.

Aim 2 Measures

Race/Ethnicity

Race/Ethnicity were collapsed into 3 primary groups for group comparison analyses: Black, Latinx/Hispanic/Chicano, and White.

Gender

Gender groups "Male" and "Female" were included in group comparison analyses. The group "Other Gender" was not included in comparison analyses due to low frequency (n=13).

Referral Source

Referral source was collapsed into four groups for analyses: *Child Welfare, Juvenile Justice* (Probation, Parole, and Diversion), *Mental Health*, and *Other Source* (School Referral, Other Internal Agency, Self-Referral, Other, Unknown).

Primary Referral Reason

Primary referral reason was collapsed into four groups for group comparison analyses: *Youth Behavior/Delinquency* (Delinquent Behavior, School Truancy/Behavior, Runaway Behavior, Gang Involvement*), Family Reason* (Family Violence, Family Separation, Family Reunification, Youth/Caregiver Conflict, Family Substance Use), *Mental Health*, and *Other Reason* (Other, Unknown).

Age Group

Age groups included in comparison analyses were "*11-18*" (FFT intended age range) and "*Older than 18*." The group "Younger than 11" was not included in analyses due to low frequency (n=11).

Served Prior or During Covid

Beginning on March 15, 2020, in response to the Covid-19 pandemic, FFT LLC, shifted to allow for teletherapy in any FFT session. Analyses were also completed using a subset of cases to examine group differences between clients who *terminated within one year prior to March 15, 2020* or *within one year following March 15, 2020*.

Completion

As described earlier, completion was a yes/no variable based on whether the participant completed FFT or discontinued before completing FFT.

Status Outcomes

As described earlier, the following status outcomes were recorded as yes/no by therapist at the end of treatment, when known: client remained in the community, client was attending school or work, and client had no new law violations.

Analytic Plan

Aim 1a: Assessing trends for quality assurance targets over time for FFT teams in Virginia

Line graphs were created for each quality assurance target to permit visual inspection to describe general trend over time. The proportion of teams meeting each program standard or benchmark was also calculated for each quarter when relevant.

Aim 1b: Assess Client Outcomes

Completion and rates of client status disposition at the time of termination (i.e., client in community, client in school, client without new law violations) were compared to national benchmarks for FFT outcomes.

Aim 1c. Summarize implementation and sustainment for most recent quarter

Determine implementation/sustainment stage and summarize quality assurance target performance and client outcomes, by team, as of June 2022.

Categorizing Teams Based on EPIS Phase (Aarons et al., 2012)

Teams were categorized into 3 stages as of the FY quarter beginning April 2022: (1) Teams less than 2-years post-training; (2) Teams 2 or more years post-training; and (3) Teams no longer active. Teams less than 2-years post-training were considered in the *Implementation phase*, as they still received significant consultation support from the purveyor. Teams two or more years post training were considered as in the *Sustainment phase*.

Current Implementation and Sustainment

To provide a summary of FFT in the state of Virginia, implementation and sustainment were assessed based on quality assurance target performance and outcomes computed for the latest FY quarter included in this project (April-June 2022).

Aim 2a: Describing FFT Participants

One component of *Reach* (of several components included in the RE-AIM framework) is used to describe FFT participants in the current study: the absolute number of participants who received FFT and their demographic characteristics. Demographic characteristics will be examined descriptively to understand the participant composition in Virginia (e.g., *describe* who participated). Descriptives for participant demographics and characteristics were examined and

participants were collapsed into broader categories to provide a summary of who has participated in FFT in Virginia.

Aim 2b: Examine differences in outcomes among subgroups of participants.

A series of chi-square tests were used to test group differences on completion and status outcomes, based on demographic variables. Additional analyses were completed using a subsample of participants who terminated FFT within one year prior to, and one year following, the start of the Covid-19 pandemic. For these analyses, chi-square tests were again used to test group differences on completion and status outcomes, and additionally, t-tests were used to test group differences in treatment duration and session number.

Aim 3: Identifying and Characterizing Modifications

Modifications to FFT implementation as intended were identified, described, and characterized by components of FRAME (Wiltsey Stirman et al., 2019), as available data allowed. These components include: (1) when and how in the implementation process the modification was made, (2) whether the modification was planned/proactive or unplanned/reactive, (3) who determined that the modification should be made, (4) what is modified, (5) at what level of delivery the modification is made, (6) type or nature of context or content-level modifications, (7) the extent to which the modification is fidelity-consistent, and (8) the reasons for the modification, including (a) the intent or goal of the modification (e.g., improve fit, adapt to a different culture, reduce costs, etc.) and (b) contextual factors that influenced the decision. The 2023 version of the FRAME Codebook that accompanies the updated framework paper (Wiltsey Stirman et al., 2019 supplemental materials) is used to determine these components for each modification identified by the present study.

Results

Aim 1a: Assessing Trends in Quality Assurance Targets Over Time

Overview

Five quality assurance targets with standards or benchmarks set by the FFT model were examined over the 4.75-year period for each FFT team in Virginia. These included: Team size, average caseload, treatment duration, fidelity scores, and frequency of fidelity rating during supervision.

Number of teams

FFT began in 2017 with two teams in Virginia. From 2017-2022, teams grew to a total of 9 teams.

Team Size

The number of therapists per team across quarters is displayed in Figure 1. Across quarters, team size ranged from 1 -7 therapists. The standard FFT team is comprised of 3-8 therapists. At the system level, all teams had at least 3 therapists for the first 14 quarters, but in the latest 4 quarters there was a downward trend, with the percent of teams with at least 3 therapists ranging from 44%-71%. This is likely overestimated as well by calculating team members across quarter (e.g., there may have been times within the quarter where staffing was lower).

Team Average Caseload

The average caseload per team across quarters is displayed in Figure 2. Average caseload for each team's therapists ranged from 1.3 – 14.5 cases. The average for the system across quarters was 2.3-11.8 cases. The FFT standard caseload size is an average of 10-12 families, with an expected range of 5 - 15. At the system level, the average was within the benchmark of expected average for only 3 of 19 quarters, otherwise the average was lower than 10 cases per

therapist all other quarters, with only one team ever having an average caseload above 15. Figure 3 displays the percentage of teams that met the benchmark of expected range of cases (5 – 15 cases).

Treatment Duration

Treatment duration was calculated by number of days between first and last treatment session. For those that completed treatment, treatment duration ranged from 25-571 days, with a mean of 113.74 days (SD = 43.35). The national standard is 60 to 180 days for FFT. At the system level, only two teams had one quarter where the mean treatment duration was above 180 days, and no teams had an average duration less than 60 days, for cases that completed. Cases that were closed before reaching completion of FFT had duration that ranged from 1 day to 416 days, with a mean of 47.14 (SD=43.35). 14% of cases that closed before completion only attended one session and 49.5% of cases that closed before completion attended 4 sessions or fewer.

Fidelity

Team-level average fidelity scores across the study time period are displayed in Figure 4. Team-level average fidelity scores, by quarter, for each team ranged from 1 – 6.

Frequency of Fidelity Rating

On average, teams reported fidelity ratings for 55% of weeks, and teams varied in their fidelity rating frequency, ranging from 13%-87% of weeks overall with fidelity ratings reported. Within individual quarters, among teams, the lowest percentage of weeks with fidelity ratings was 8% and the highest percentage of weeks with fidelity ratings was 100%.

Aim 1b: Assessing Client Outcomes

Completion Rate

62.7% of participants completed FFT at case close overall. Completion rate ranged from *0 -100% per quarter,* across teams. Completion rates were quite variable by team and quarter. FFT national standards range from 70 to 80% completion rate, 75% was used as a benchmark in this study. At the system level, 75% of teams were at or above this standard for one quarter, otherwise the total percent of teams at or above this standard ranged from 11-33% across quarters. Results are displayed in Figure 5.

Outcomes at Closure – Remain in Community, Attend School/Work, No New Law Violations

For participants with available outcome data, 77.5% of participants remained in community at case closure, 79.2% participants were engaged in school or work at case closure, and 75.8% of participants had no new law violations at case closure. Status outcome results are displayed in *Table 2*. For each quarter, percentages were calculated for completed cases, for status indicators. FFT standard is maintaining over 85% success in these categories. Across quarters, the percentage of youth remaining in the community at the end of treatment ranged from 81%-100%. Figure 6. Across quarters, the percentage of youth in school or working at the end of treatment ranged from 90%-100% per team. Results are displayed in Figure 7. Across quarters, the percentage of youth with no new law violations at the end of treatment ranged from 66%-100% per team. Results are displayed in Figure 8.

Aim 1c: Summarize Implementation and Sustainment

At the end of the study period, team implementation time ranged from two quarters to nineteen quarters, with two of nine teams in the Implementation phase (implementing for less than two years) and seven teams in the Sustainment phase. No teams had closed by the end of the study period, but one team is known to have closed in the quarter after.

In the final quarter, team size was the most common benchmark that was not met by teams. In the final quarter, 6 of 9 teams had 2 or fewer therapists. The next most common benchmark that teams were challenged by was average caseload. In the final quarter, 4 of 9 teams had below benchmark average caseloads.

Completion rate was the most common outcome for which teams did not achieve the FFT benchmark standard, with 6 of 9 teams having completion rates below 75% in the most recent quarter. In the final quarter, most teams had very high rates for status outcomes, with only one team falling below benchmark for both rate of youth remaining in community and rate of youth having no new law violations. The newest team had no cases completed yet and therefore no outcomes at completion available for report. Overall, there was variability across teams in meeting standards and benchmarks for quality assurance targets and outcomes and no clear pattern based on time since implementation.

Aim 2a: Describe FFT Participants in Virginia

To describe the FFT participant population in Virginia and make group comparisons across participants, participants were collapsed into smaller descriptive categories. Overall descriptives, along with the collapsed categories are indicated on *Table 1*. Participants were collapsed into race/ethnicity categories that included: Black (38.6%), White (36.6%), and Latinx/Hispanic/Chicano (11.9%). Participants with other races/ethnicity identities made up 12.86% of the participants. The majority of FFT participants in Virginia identified as male (71.6%), with 27.6% identifying as female, and 0.8% identifying as other genders. Most participants (86.9%) were categorized as "within the age range intended for FFT (11-18 years old)," with 0.6% categorized as "younger than 11", and 11.6% of participants in Virginia categorized as "older than 18 years old." Participants were grouped into four categories based on

referral sources related to: Juvenile Justice (67.8%), Child Welfare (7.9%), Mental Health (9.2%), and Other Sources (15.1%). Participants were grouped into four categories based on primary referral reasons related to: Youth Behavior Reason (*62.8%),* Family Reasons (*24.0%*), Mental Health Reason (*7.2%*), and Other Reasons (*6.0%*). A subset of participants received FFT during the first year of Covid-19 (n=385) and within one-year preceding the start of Covid-19 (n=426).

Aim 2b: Assessing Outcomes Across Subgroups

Race/Ethnicity

Group differences based on Race/Ethnicity were examined. There was not a significant difference in completion rate based on race/ethnicity, X^2 (df = 3, N = 1719) = 6.53, $p = 0.09$. For status outcomes, there was not a significant difference in participants remaining in community X^2 (df = 3, N = 1618) = 5.55, $p = 0.14$, or engaging in school/work X^2 (df = 3, N = 1554) = 3.40, $p = 0.34$, based on race/ethnicity. There was a significant difference in participants having no new law violations based on race, X^2 (df = 3, N = 1589) = 30.43, $p < 0.001$. Post-hoc analyses indicated that a significantly smaller proportion of Black participants (68.6%) had no new law violations than White participants (81.5%; Z = 05.08, p < 0.001) and Latinx/Hispanic/Chicano participants (77%; Z = -2.22, p < 0.05).

Table 3 displays race/ethnicity groups by referral source. Of note, Black youth were referred more frequently from juvenile justice related sources than participants from other race/ethnicity groups.

Gender

Next, group differences based on gender were examined. There was not a significant difference in completion rate based on gender, X^2 (df = 1, N = 1705) = 0.18, $p = 0.67$. For status

outcomes, there was a significant difference in all three status outcomes based on gender. A significantly larger proportion of girls remained in community (85.8%), as compared to boys (80.8%), X^2 (df = 1, N = 1705) = 5.44, $p < 0.05$). A significantly larger proportion of girls (90.5%) engaged in school or work as compared to boys (86.6%) X^2 (df = 1, N = 1705) = 4.31, $p < 0.05$). Finally, a significantly larger proportion of girls (82%) had no new law violations as compared to boys (73.1%), X^2 (df = 1, N = 1705) = 13.58, $p < 0.001$).

Referral Source

Next, group differences based on referral source were examined. Groups based on referral source did significantly differ on completion rate, X^2 (df = 3, N = 1719) = 9.8, $p < 0.05$). Post-hoc analyses indicated that a significantly larger proportion of participants referred from Juvenile Justice related sources (65.2%) completed FFT, as compared to participants referred from Mental Health related sources (54.5%, Z = -2.6, p < 0.01). Other groups did not differ. Groups based on referral source did significantly differ on rate of remaining in community, X^2 (df = 3, N = 1618) = 10.18, $p < 0.05$). Post-hoc analyses indicated that a significantly smaller proportion of participants referred from Juvenile Justice related sources (80.5%) remained in community, as compared to participants referred from Mental Health sources (88.4%, Z = 2.26, p < 0.05) or Other sources (87.2%, Z = 2.47, p < 0.05). There was not a significant difference in rates of engaging in school or work based on referral source. Groups based on referral source did significantly differ on rate of no new law violations, X^2 (df = 3, N = 1589) = 36.21, $p < 0.01$). Post-hoc analyses indicated that a significantly smaller proportion of participants referred from Juvenile Justice related sources (71.5%) had no new law violations, as compared to participants referred from Child Welfare sources (84.7%, Z = 3.13, p < 0.01), Mental Health sources (87.5%, Z = 3.98, p < 0.001) or Other sources (84.5%, Z = 4.09, p < 0.001).

Referral Reason

Next, group differences based on primary referral reason were examined. There was not a significant difference in completion rates, X^2 (df = 3, N = 1719) = 2.53, p = 0.47), rates of remaining in community, X^2 (df = 3, N = 1618) = 7.28, p = 0.06), and rates of engaging in school or work, X^2 (df = 3, N = 1554) = 3.60, p = 0.31), based on primary referral reason. Groups based on primary referral reason did significantly differ on rate of No New Law Violations, X^2 (df = 3, N = 1589) = 17.07, $p < 0.001$). Post-hoc analyses indicated that a significantly smaller proportion of participants referred for primary reason related to behavior/delinquency (73.3%) had no new law violations, as compared to participants referred for primary reason related to mental health (90.6%, Z = -3.90, p < 0.001). Proportions for primary reasons related to behavior/delinquency (73.3%) as compared to primary reasons related to family (78.2%, Z = -1.84, p = 0.07) and Other reasons (75.5%, Z = -0.47, p = 0.64) were not significantly different.

Age

Next, group differences based on age group were examined. For completion rate, participants within the FFT intended age range (62.7%) and older than 18 years old (66.3%) did not significantly differ, X^2 (df = 1, N = 1693) = 0.99, $p = 0.32$). A significantly larger proportion of youth older than 18 remained in the community (89.5%), as compared to youth within the intended FFT age range (81.4%), X^2 (df = 1, N = 1601) = 7.57, $p < 0.01$). There was not a significant difference in proportion of participants of engaging in school or work, X^2 (df = 1, N = 1538) = 2.02, $p = 0.09$) based on age group. There was also not a significant difference in proportion of No New law violations, X^2 (df = 1, N = 1573) = 1.09, $p = 0.17$) by age group.

It was not expected that 11.7% of the sample would be older than the intended age range for FFT. To learn more about the referral composition of this group, additional analyses were

performed to examine group differences in referral source, based on age group. Results indicated that the groups did significantly differ based on referral source, X^2 (df = 3, N = 1693) = 14.91, $p < 0.01$). Post-hoc analyses indicated that the group of participants who were older than the intended age range for FFT had a significantly higher proportion of participants referred by Juvenile Justice related sources (79.9%), as compared to the group of participants within the intended age range for FFT (66.5%), Z = -3.8, p < 0.001), as well as significantly lower proportion of participants referred by Child Welfare related sources (4.0%), as compared to participants within the intended age range for FFT (8.4%), Z=2.166, p <0.05.

Pre- and Post-COVID

Next, group differences for a subset of cases (n=811), including participants with a termination date within one-year prior to the start of Covid-19 (n=426) or within the first year of Covid-19 (n=385), were examined. For completion rate, participants pre- (65.5%) and post Covid-19 (67.3%) did not significantly differ, X^2 (df = 1, N = 811) = 0.29, $p = 0.32$). A significantly greater proportion of participants who participated during the first year of Covid-19 remained in community (88.7%), as compared to those that participated in one year prior to Covid-19 (79.3%), X^2 (df = 1, N = 773) = 12.57, $p < 0.001$). There was not a significant difference in proportion of participants engaging in school or work pre- and post-Covid-19. A significantly larger proportion of participants during the first year of Covid-19 had no new law violations (83.3%), as compared to those that participated in one year prior to Covid-19 (66.9%), X^2 (df = 1, N = 765) = 27.15, $p < 0.001$). However, within this overall improvement in rate of clients having no new law violations during the first year of Covid-19, racial disproportionality in this outcome remained, with a smaller proportions of Latinx/Hispanic/Chicano (71.7%) and Black (79.2%) clients having no new law violations, as compared to White (89.3%) participants.

Additionally, although there were overall improvements on this outcome for Black and White participants, when comparing one year prior to Covid-19 and the first year of Covid-19, in contrast, the percentage of Latinx/Hispanic/Chicano participants with no new law violations actually decreased slightly in the first year of Covid-19 (75.4% vs 71.1%), but sample size for this subset of the total sample is small (n=57 total Latinx/Hispanic/Chicano youth for year prior to Covid and n=46 total Latinx/Hispanic/Chicano youth for first the year of Covid-19).

Finally, duration of treatment and number of sessions were examined to assess differences between cases that closed within one year prior to Covid-19 and cases that closed within the first year of Covid-19. There was not a significant difference for duration of participation for cases that closed during the year prior to Covid-19 (M = 87.9, SD =50.4) and cases that closed during the first year of Covid-19 (M = 80.8, SD = 55.1), $t(809) = 1.91$, $p = 0.06$. Similarly, there was not a significant difference in the number of sessions, for cases that closed one year prior to Covid (M=9.77, SD=4.7) and closes that closed during the first year of Covid-19 (M=9.51, SD=5.2), $t(809) = 0.74$, $p = 0.46$).

Aim 3: Modifications

Two major modifications were noted in this evaluation of FFT implementation and sustainment in Virginia. In order to characterize each modification, FRAME components are reported, based on data available to the present study.

First, FFT was provided via telehealth for some sessions following the start of the Covid-19 pandemic. (1) *When/how:* This modification was made during implementation. Prior to the pandemic, FFT LLC had permitted the use of teletherapy services (e.g., video conferencing) in exceptional circumstances (e.g., extreme weather). Beginning on March 15, 2020, FFT, LLC, shifted to allow for teletherapy in any session. A 2021 paper describing this shift to teletherapy

reported that 62.7% of FFT sessions completed globally between March 15, 2020 and September 1, 2020, were conducted via teletherapy (Robbins & Midouhas, 2021), although the proportion of sessions delivered via teletherapy in Virginia is not known based on data included in the present study. (2) *Planned/proactive or unplanned/reactive:* This modification was unplanned and reactive. (3) *Who determined that the modification should be made:* FFT Purveyors determined this modification. (4) *What is modified*: The modification is *contextual*, as modification was made to the overall way treatment was delivered. (5) *At what level of delivery the modification is made:* The modification was made across the entire service system. (6) *Type or nature of context or content-level modifications:* The format was changed from in-person to virtually delivered via video. (7) *The extent to which the modification is fidelity-consistent:* This is not known by the present study, as the present study does not capture if content modifications were made in Virginia as a result of context modification (delivery format). (8) *The reasons for the modification:* In response to health concerns and regulations (e.g., lock down) related to the COVID-19 pandemic, FFT LLC allowed for sessions to be conducted via teletherapy to maintain continuity of care and continued accessibility of FFT services.

A second modification identified by this study was that participants outside of the intended age range for FFT were provided FFT in the state of Virginia (0.6% of participants were younger than 11 and 11.6% were older than 18 years old.). (1) *When/how:* This modification was made during implementation. Participants outside of the intended age range participated in FFT throughout the study time period. (2) *Planned/proactive or unplanned/reactive:* This modification was unplanned and may be reactive, although this is unknown. (3) *Who determined that the modification should be made:* It is not known by the present study who determined this modification. (4) *What is modified*: The modification is

contextual, as modification was made to the overall way treatment was delivered. (5) *At what level of delivery the modification is made:* The modification was made across the service system. (6) *Type or nature of context or content-level modifications:* Contextual modification was made to extend the population served, outside of the intended age range. (7) *The extent to which the modification is fidelity-consistent:* This is not known by the present study, as the present study does not capture if content modifications were made depending on age. (8) *The reasons for the modification:* Reason is not captured by data available to the present study.

Discussion

The present study sought to address three aims. First, the study evaluated the implementation and sustainment of FFT in Virginia based on performance on program components quality assurance targets and client outcomes across the first five years of FFT's implementation in the state. Second, the study examined FFT participants in Virginia by describing the participant population and assessed group differences in completion rate and status outcomes across participant subgroups. Third, the study identified and characterized modifications to FFT that occurred in Virginia during the study time period.

To summarize, at the end of the study period, two teams were in the implementation stage and seven teams were in the sustainment phase. Highlights for Virginia teams included largely positive status outcomes for youth (e.g., remain in community, engage in school/work, and have no new law violations). Some areas of concern for Virginia teams included significant drop in the number of the therapists per team from 2020 forward, a notable decrease in caseloads for teams starting in 2019 and hitting a plateau in 2020, and overall low and below national benchmark case completion rates.

Subgroup differences in completion rates were found based on gender. Subgroup differences in rate of remaining in community were found based on gender, referral source, age, and participation during the first year of the Covid-19 pandemic. No group differences were found for rate of engaging in school/work. Group differences were found for rate of participants having no new law violations, based on race/ethnicity, referral source, referral reason, and participation during Covid-19. There was not a significant difference in duration or number of sessions for cases closed in the year prior to Covid and those that closed during the first year of Covid. Each of these findings is discussed in turn.

Assessing Trends in Quality Assurance Targets Over Time

Research Questions: How does the implementation of FFT in Virginia compare to national benchmark standards of implementation of FFT and what, if any, components would benefit from focused quality improvement efforts? Given the organizational and individual impacts of the Covid-19 pandemic, how was the quality of implementation of FFT in Virginia impacted by the Covid-19 pandemic?

Analyses aimed to describe the stability of FFT quality in Virginia across the first five years of implementation produced a number of notable results. Virginia started with two FFT teams and ended the study time period with nine FFT teams, with no teams closing during this time. Two teams were in the implementation phase (providing services for two years or less) and seven teams were in the sustainment phase (providing services for more than two years). No teams met all FFT benchmarks at the final quarter included in the study period, which is somewhat expected, as it is unlikely that any program implemented in community settings would achieve all benchmarks. These results are in line with findings from a systematic review of 125 empirical studies focused on the sustainment of public health and clinical interventions, which

indicated that partial sustainment is much more common than full sustainment of an intervention (Wiltsey Stirman et al., 2012).

FFT requires a team size of three to six therapists. In Virginia, teams met this benchmark for the first three years of implementation, with a notable decrease in team size beginning in the first quarter of 2020 and continuing from that point. The decrease coincides with the COVID-19 pandemic and also with changes in the payment rate for FFT via Medicaid. In the last three quarters of data available for this analysis, the average Virginia team was below the national benchmark for team size.

Average caseload is another quality assurance target for standard FFT. The model purports that quality of delivery for the program is improved by ongoing therapist experience with the model (Turner et al., 2018). Data on implementation in Virginia demonstrated a steady rise in caseloads through the first two years of implementation, followed by a steady decrease beginning in 2019, hitting a nadir in the second quarter of 2020 and then plateauing at around 7 cases, also coinciding with the onset of the Covid-19 pandemic. While most quarters teams did meet the general benchmark for range of cases (5-15) overall the mean caseload was below the benchmark mean of 10-12 cases, throughout the study period. It is possible that overall, teams were only able to maintain smaller caseloads due to turnover and hiring of new staff, and this overall delayed building and maintaining large caseloads.

Supervisors reported different scores for the teams on the FFT fidelity measure. A major caveat for interpreting trends in fidelity in this study is that fidelity averages were based only on available ratings, which were variable in frequency of rating by team, and there is not yet adequate evidence for the validity and reliability of this fidelity measure. Despite the FFT model guidance of maintaining weekly fidelity monitoring during group supervision, in the present

study reports of fidelity ratings were less than weekly overall (on average teams had at least one fidelity rating for one therapist, 55% of weeks). There was variability among teams, with some teams following protocol 100% of weeks some quarters, and others as low as 8% of weeks for some quarters.

The frequency of rating fidelity during weekly supervision was much lower than intended by the FFT model. Although not captured by the current study based on available data, potential explanations for drift in practice might be that (1) supervision in general was not occurring as frequently as intended, (2) supervision was occurring, but other topics took precedence over case discussion/fidelity rating during supervision (e.g., discussion of administrative issues, referrals), or (3) any number of other reasons, such as burnout, that fidelity rating would not be conducted (or recorded). Qualitative data collection would be important to elucidating more information about the reason for, and nature or this drift from protocol. This kind of drift may occur naturally as part of implementation and can be viewed negatively as decreasing fidelity to the intervention as designed, but alternatively, can also be viewed pragmatically as attempts to increase the feasibility of a practice and therefore contribute to long-term sustainability of an intervention (Shelton et al., 2018; Chambers et al., 2013). From a research perspective, if weekly fidelity rating is not feasible, it may be helpful to empirically test various frequencies of fidelity monitoring as an intentional adaptation, to see if there is a lower "threshold" of fidelity rating frequency, and/or frequency of supervision in general, that is more feasible and still maintains an acceptable level of program effectiveness (Shelton et al., 2018).

There has been limited research on supervision, EBP fidelity rating, and outcomes. Some studies have looked at supervision content or techniques, while other studies have looked at supervision format (e.g., observational - including review of recorded sessions). In a study of

Multisystemic Therapy (MST), Schoenwald et al. (2009) found that supervision that focused on adherence to MST principles, predicted greater therapist adherence and subsequent client outcomes. A study of Trauma-Focused Cognitive Behavioral Therapy (TF-CBT) found that supervision that included more of a directive approach (e.g., didactic instruction and clinical suggestions), as compared to supervision that was primarily supportive supervision, resulted in greater likelihood therapist's completing an important component of TF-CBT with clients, the "trauma-narrative" (Meza et al., 2023). Similarly, a study of Cognitive Behavioral Therapy indicated that supervision utilizing experiential learning (e.g., modeling and role-plays) and feedback on session recordings, resulted in greater increases in provider EBP fidelity, as compared to supervision as usual (Bearman et al., 2017). In a study of Motivational Interviewing, Martino et al. (2016) found clinicians supervised with an enhanced supervision including review of recorded sessions, significantly increased their MI competency as compared to those that received supervision as usual, but that this enhanced supervision did not result in improved client treatment retention or outcomes and was very costly.

Robbins et al. (2018) conducted a study to test an enhanced version of FFT supervision, observation-based supervision (e.g., review of audio recordings of sessions), as compared to FFT supervision as usual (therapist reports on a case). Findings indicated that clients with externalizing behavior above clinical thresholds had significantly greater reductions in problem behaviors in the observation-based supervision versus the supervision as usual conditions and clients below thresholds did not respond differentially to conditions. The authors discuss that the modest impact of observation-based supervision, although helpful for more acute clients, may not be worth the investments of time and resources needed. This discussion emphasizes a

common focus on compromise between benefit and overall feasibility in real-world service settings.

There has not been research on the impact of frequency of FFT supervision, or frequency of fidelity rating as part of supervision. Nevertheless, weekly group supervision is considered an integral part of FFT, and supervision sessions are typically divided into two key activities: fidelity review and case planning. Robbins and Alexander (2019) discuss the importance of the group supervision format for collaborative learning, exposing therapists to many clinical issues and allowing for therapists to think about and provide feedback on their colleague's cases. The supervision model also follows the assumption that "the more cases that are reviewed in supervision, the broader the impact of supervision on therapist fidelity and in turn, on clinical outcomes" and "therapists learn as much from their successes as their failures; therefore, supervision (and fidelity review) cannot solely focus on difficult cases" (Robbins & Alexander, 2019). The authors emphasize that fidelity monitoring, as part of group supervision, should occur weekly. The authors state, "Every week, supervisors assess both adherence and competence and provide feedback to therapists about their performance. This ongoing quality assurance and improvement process is as essential to ensuring successful outcomes as the therapist interventions themselves. Without this review and feedback, FFT is not being implemented according to recommended practice guidelines." More work is needed to assess the impact of less frequent fidelity rating on fidelity and outcomes. Additionally, fidelity monitoring, while usually studied with regards to impact on client outcomes, has also been shown to be beneficial to staff retention. For example, Aarons et al. (2009) tested the effect of fidelity monitoring as part of the implementation of SafeCare, a EBP for child neglect. They found that fidelity monitoring as part of SafeCare implementation resulted in greater staff retention, as compared to

SafeCare implementation without fidelity monitoring and services as usual with fidelity monitoring. They discussed fidelity monitoring as likely offering opportunity for a sense of mastery, competency, and self-efficacy within the context of implementation a new practice. Future work could examine the impact of frequency of fidelity monitoring on FFT therapist retention.

Maintaining positive outcomes is considered an important indicator of sustainment at the population level, and particularly for "partially sustained" programs (e.g., Even if not fully sustained according to model, is the population served continuing to see positive impacts?) (Shelton et al., 2018). First, the present study looked at completion rate. Virginia teams demonstrated unsteady performance on this metric. In general, the completion rate for cases was very low in Virginia as compared to national benchmarks. For cases that did not complete FFT, about half of those cases received fewer than 5 sessions before case closure, ending treatment relatively soon after starting. Analyses specifically comparing the year prior to Covid-19 and the first year of Covid-19 indicated that completion rate was not significantly different across those two years.

The present study looked at the following status outcomes: youth remain in the community, youth are attending school or work, and youth have no new law violations at the time of completing FFT, over time. Virginia's performance for these three outcomes areas was overall very promising, with average of achieving this benchmark on two outcomes, youth remain in the community and youth are attending school and work, consistently at or above this benchmark. Although still promising, fewer clients overall met the benchmark for the status outcome of having no new law violations, across the study time period.

Recommendations for Quality Improvement

Taken together, these findings suggest a few areas for quality improvement for FFT in Virginia. The first is related to maintaining FFT team size. Study findings related to team size (e.g., team size decreased over the study time period, at times leaving teams with 1 therapist) point to staff leaving their positions, although this study does not capture reasons for decreasing team size. More qualitative research is needed, specific to FFT teams and retention and hiring, in order to confirm this interpretation. *If* decreasing team size is due to staff leaving their positions and difficulty with rehiring for these positions, workforce changes might be a promising focus. Historically, the field of mental health, particularly public mental health services, has experienced difficulties with maintaining a robust workforce (U.S. Department of Health and Human Services [HHS], 2013). Community-based mental health is a particularly challenging setting to work in and community-based programs experience significant turnover of employees (Woltmann et al., 2008). Prior to the COVID-19 pandemic, there was an existing shortage in the behavioral health care workforce, and this shortage has been exacerbated by increased demand for mental health care (National Institute for Health Care Management, 2023). Additionally, during the Covid-19 pandemic, similar to healthcare workers across disciplines, the behavioral health workforce experienced exacerbated burnout (Fish & Metal, 2021). Behavioral healthcare workers faced work challenges during the pandemic such as increased workload, changing roles, decreased job satisfaction, telehealth challenges, difficulties with work-life balance, and vicarious trauma (Crocker et al., 2023). A recent study conducted during the Covid-19 pandemic found that community mental health providers who were experiencing more work changes in terms of their tasks, settings, and teams were more likely to report higher levels of burnout and were subsequently more likely to consider leaving their job (Sklar, Ehrhart, & Aarons, 2021).

Recruitment and retention of therapists, particularly for challenging home-based approaches like FFT, has become a national crisis (Crocker et al., 2023, Rossi et al., 2023) and recruiting and training new staff to replace staff who leave is costly to agencies (Brabson et al., 2020). In a study of Oregon's public behavioral health workforce, low wages, documentation burden, poor infrastructure, lack of career development opportunities, and a chronically traumatic work environment were found to be key factors influencing turnover and attrition (Hallett et al., 2024). Similarly, in Virginia, the behavioral health workforce shortage has been reported as a major barrier to services, with common influences including lack of funding to offer competitive pay, increased practitioner credentialing requirements and burdensome licensure processes, and lack of consistent, sufficient, and affordable resources, training, and education for behavioral health professionals (Sale et al., 2023). While behavioral workforce shortages have impacted access to care nationally, Virginia ranks 40th among US states for mental health worker availability (including psychiatrists, psychologists, Licensed Clinical Social Workers (LCSW), Licensed Professional Counselors (LPC), Marriage and Family Therapists (MFT) and advanced nurse practitioners) according to a 2021 report on mental health care access (Reinert et al., 2021).

Again, *if* decreasing team sizes are due to staff leaving their positions and difficulty rehiring, the sustainment of FFT in Virginia might be strengthened with efforts to bolster staff recruitment and retention efforts. Although a challenging task, there have been a number of recommendations for strengthening staff recruitment and retention of behavioral health workforces. These include incentivizing employees by offering things such as opportunities for professional development, student loan repayments, retention bonuses, or more flexible time (Yang & Hayes, 2020; Covino, 2019). Other recommendations, based on previous research,

include efforts to improve the organizational climate (Ehrhart & Kuenzi, 2017) and culture (Powell et al., 2021; Glisson et al., 2008) of organizations implementing evidence-based practices. For example, this could look like providing training to leaders to improve leadership skills that are specific to supporting staff through the implementation evidence-based practices (Aarons et al., 2017) or supporting staff through initiatives focused on preventing or addressing burnout (Awa et al., 2010). Organizational interventions could be particularly important for a model like FFT, where therapists make up a team (rather than operating fully independently), and therefore may be particularly influenced by organizational changes.

A second, and related issue recognized in this study was average therapist caseload, which was on lower than FFT national benchmark for most quarters in the study period. Caseload may tie closely with the workforce issues FFT in Virginia experienced during its first five years of implementation, as turnover impacts stability of implementation of EBPs (Woltmann et al., 2008). With an inconsistent workforce, it is difficult to expect that FFT teams would be able to consistently maintain adequate caseloads. Additionally, caseload may also be tied to funding available for FFT. Virginia has a complex funding context in which FFT can be paid for by multiple different types of funding sources (i.e., Medicaid, Children's Services Act, Department of Juvenile Justice, or Department of Social Services). In Virginia, payment via these multiple sources is not equalized (e.g., different rates across funders). Further, each payment source presents its own technical and time barriers. Increasing cases across FFT teams is an important goal for Virginia, but it is important to acknowledge these barriers when providing such a recommendation.

Another challenge noted in the Virginia data was relatively low case completion rate (62.7%). Drop out in community-based youth mental health services is a widespread challenge in

general, and rates of dropout from children's mental health services have been found to vary from 28 to 88% (e.g., De Haan et al. 2013). For FFT, as a family-based treatment, there may be additional challenges to retention, including caregiver expectations, differing engagement from caregiver and youth, balancing alliance across family members, and perceived need for FFT (Collyer et al., 2020; Robbins et al., 2008; McPherson et al., 2017). More work is needed to assess factors that impact completion or drop out for FFT.

To address below-average completion rates in Virginia, it may be helpful for FFT training and supervision to include an increased focus on youth and family engagement at the beginning of treatment (Becker et al., 2018; Haine-Schlagel & Walsh, 2015). For example, establishing credibility with the family, disrupting negativity and blame among family members, and reframing presenting problems to foster a sense of hope for change are all critical program activities delineated in the FFT treatment manual, particularly during the initial sessions (Alexander et al., 2013). In the present study, about half of non-completing cases received fewer than 5 sessions before case closure. Presumably increasing a focus on the engagement phase of FFT would likely decrease drop out and will take better advantage of resources invested into FFT in Virginia. Of course, case completion challenges may be driven by the aforementioned challenges with team sizes and caseloads.

In summary, Virginia's implementation of FFT started off strongly for the first two to two and half years by most quality assurance targets. However, around the time of the Covid-19 pandemic, implementation faltered across several targets, especially with regards to team size and caseload. Outcomes for Virginia's teams, though variable, were also mostly strong, if somewhat below national benchmark standards.

FFT Participants

Research Question: Who participated in FFT during the first 4.75 years of implementation, with regards to race/ethnicity, gender, referral source, referral reason, and age.

FFT grew from two teams to nine teams of service providers over the study time period. The overall participant population of FFT included 1719 youth (and their families) over 4.75 years of implementation. Of note, there was overrepresentation of Black youth in the overall sample (38.6%), as compared to Virginia's 2020 census data which indicates Black youth represented only 19% of youth in Virginia (Children's Bureau, 2020). Further, within the sample, Black youth were disproportionately referred from DJJ related referral sources (83.6% of Black youth in the study). On one hand, this could be considered a promising aspect of participation for FFT in Virginia, as historically, youth from minoritized backgrounds are often disproportionately *underserved* by mental health services (Cook et al., 2019; Hoffman et al., 2022). On the other hand, this overrepresentation likely reflects larger issues of systemic racism inherent in juvenile justice systems (Abrams et al., 2021) and is in line with racial disproportionality specifically in Virginia's juvenile justice system (Joint Legislative Audit and Review Commission, 2021). With regards to gender, FFT in Virginia served a higher proportion of boys than girls, which is generally in line with higher rates of externalizing behavior in boys (Eme et al., 2016). Boys also have a higher rate of DJJ involvement (U.S. Department of Justice, 2020), and in the current study sample, DJJ related sources were the largest proportion of referral sources.

Limited referrals or changes in referral frequency can sometimes pose a challenge to the implementation and sustainment of an evidence-based practice (Zazzali et al., 2008; Wright et al., 2024; Cooper et al., 2015). Although DJJ related sources represented the largest category of referral sources in the present study, overall, Virginia had variety in referral sources and variety in referral reasons for FFT. When considering sustainment of an evidence-based practice within

a state, this variety can illuminate challenges related to differing funding and reimbursement. However, overall, this variety can be viewed as a strength for the longevity of FFT as a practice in Virginia, with multiple referral sources increasing the number of participants overall, and general awareness of FFT as a service established in the state. Variety in referral sources can also be useful in ensuring longevity, as different referral sources may face different barriers to providing referrals, as context and policy landscapes change, and variety may act as a safeguard. Additional work should continue to assess the impact of funding and referral processes on this variety of referral sources and referral reasons, as well as opportunities for cross-collaboration between referring agencies. This is particularly important as Virginia continues to roll out initiatives related to the Family First Prevention Services Act, with a focus on funding for child welfare prevention services, which presents opportunity to further expand the reach of FFT for child welfare referred youth. Within the present study, child welfare related referrals represented 7.9% of FFT recipients.

The present study found unexpected reach of FFT in Virginia – FFT provided to youth outside of the intended age range of FFT (mostly older than the age range, although a small number youth were also younger than the intended age range). This "extended" reach can be considered a modification to FFT in Virginia. Of note, the older age group was most commonly referred by juvenile justice related sources. This modification will be discussed in more detail in later sections of the discussion.

Assessing Outcomes Across Subgroups

Research Questions: Did completion rates and status outcomes differ across these groups? Did completion rates, status outcomes, duration of treatment, and number of sessions completed

differ between youth who received FFT one year prior to Covid-19 and during the first year of Covid-19?

The second aim of the study was to assess for subgroup differences in completion of FFT and status outcomes at the end of participation, based on race/ethnicity, gender, referral source, primary referral reason and age. Differences were also assessed for cases that closed during the year prior to Covid and cases that closed during the first year of Covid.

Overall, differences in case completion and status outcomes did not differ by race, with exception of one outcome -- the proportion youth with no new law violations. A significantly lower proportion of Black youth had no new law violations at the end of their participation in FFT. Although we cannot rule out that FFT may not work as well for Black families as it does for non-Black families on this outcome, it seems more likely that longstanding racial discrimination in law enforcement and the juvenile justice system (Abrams et al., 2021; Joint Legislative Audit and Review Commission, 2021) are a driving factor for this difference. Additionally, within a smaller subset of youth who participated one year prior to Covid-19 or during the first year of Covid-19 racial disproportionality remained within the outcome of youth who had no new law violations, with Black and Latinx/Hispanic/Chicano participants performing worse on this metric. Overall, a larger proportion of participants had no new law violations during the first year of Covid-19, as compared to the year prior, however, in contrast, this was not true for Latinx/Hispanic/Chicano youth, although sample size is small for this subset of participants.

A larger proportion of girls remained in community, engaged in school and work, and had no new law violations as compared to boys in the sample. This is in contrast to previous research that has found non-significant differences in FFT outcomes based on gender (Celinska

& Cheng, 2017; Robbins et al., 2000). Again, the present study is not able to account for factors prior to FFT involvement that might impact these outcomes and as such results should be interpreted with caution. With regards to referral source, a significantly larger proportion of participants referred from Juvenile Justice related sources completed FFT, as compared to participants referred from Mental Health related sources. One possible explanation for this is that completion might be encouraged or mandated by a DJJ parole or probation plan (this data was not captured consistently in the current study). Although more participants referred from Juvenile Justice related sources completed FFT, a significantly smaller proportion of these participants remained in community, as compared to participants referred from Mental Health sources or other sources. Additionally, a significantly smaller proportion of participants referred from Juvenile Justice related sources had no new law violations, as compared to participants referred from Child Welfare sources, Mental Health sources or Other sources. While this is somewhat in contrast to results related to case completion, it is possible that, Juvenile Justice related referrals may enter FFT at higher risk for out of home placement and law violation. Relatedly, a significantly smaller proportion of participants referred for primary reason related to behavior/delinquency had no new law violations, as compared to participants referred for primary reason related to mental health.

As there was unexpected reach of FFT outside of the intended age range (primarily for ages 19-22), differences between this group and those within the intended age group were examined. A small number of youth (n=11) were younger than the intended age group, although this group was not included in comparison analyses due to small sample size. Results indicated that a significantly larger proportion of youth older than 18 remained in the community, as compared to youth within the intended FFT age range. This may be due to the development stage

of this age group and ability for greater independence (e.g., living outside of the family home, in community). Alternatively, it is possible that there were fewer out of home placement options available for this older age group. There were no differences in completion rate, school/work, or new law violation for this older age group which is a promising indicator for use of FFT for this "extended" age group.

It is important to note that as compared to those within the intended age range for FFT, participants who were older than the intended age range for FFT were a significantly higher proportion of participants referred by Juvenile Justice related sources and significantly lower proportion of participants referred by Child Welfare related sources. This may point to FFT as potentially "filling a gap" in services for "transition age youth" (e.g., ages 19-22) specifically in the juvenile justice system. This is discussed in more detail in the following modifications section of the discussion section.

Differences were also examined for participants that completed FFT in the year prior to the COVID-19 pandemic, and those that participated during the first year of the pandemic. A significantly greater proportion of participants who participated during the first year of Covid-19 remained in community, as compared to the year prior to Covid-19. This could be a product of fewer out of home placement options available in response to Covid-19 (e.g., higher need, facilities overloaded, occupancy restrictions (Pinals et al., 2020). This could also be a product of "lock down" and social distancing in general, with youth and families remaining where they are. A significantly larger proportion of participants during the first year of Covid-19 had no new law violations, as compared to the year prior. Similarly, these results might also be indicative of lock-down and social distancing (e.g., less peer interaction, less opportunity for law violation, overwhelmed systems) (Buchanan et al., 2020) and are in line with Virginia DJJ reports of lower

recidivism during this time period (Virginia Department of Juvenile Justice, 2021). It is important to highlight that although the rate of no new law violations improved in the first year of the COVID-19 pandemic overall, racial/ethnic disproportionality in this outcome remained, with a smaller proportion of Black and Latinx/Hispanic/Chicano participants having no new law violations than the proportion of White participants having no new law violations, during the first year of Covid.

There was not a significant difference in treatment duration or number of sessions pre-covid and during first year of COVID-19. Additionally, the first year of the COVID-19 pandemic (including FFT shift to more sessions via telehealth) did not impact length or dose of FFT. These are promising results for use of teletherapy in FFT and are discussed more in the following section of discussion.

Modifications

The first modifications that occurred during the study time period was a global modification to FFT – permission to conduct FFT sessions via teletherapy following the start of the Covid-19 pandemic. Prior to the pandemic, FFT LLC had only permitted the use of teletherapy (e.g., video conferencing) in exceptional circumstances (e.g., extreme weather). Beginning on March 15, 2020, in response to the Covid-19 pandemic, FFT LLC, shifted to allow for teletherapy in any FFT session.

Although individual session data was not included in the present study, it is known that Virginia shifted to incorporate teletherapy, following this global shift initiated by FFT LLC (and globally, 62.7% of FFT sessions between March 15, 2020 and September 1, 2020, were conducted via teletherapy according to Robbins & Midouhas (2021)). Therefore, the present study was interested in generally comparing FFT services one year prior to this modification, and

one year following this modification. Previous research on FFT's modification to allow for teletherapy following the Covid-19 pandemic did not find any difference regarding treatment length and number of sessions between March 15, 2019 to September 1, 2019, and March 15, 2020 to September 1, 2020 (Robbins & Midouhas, 2021). In line with these findings, the present study also did not find a significant difference in treatment length and number of sessions in Virginia between March 16, 2019 to March 15, 2020 and March 16, 2020 to March 15, 2021, generally indicating that the allowance for sessions to be conducted via teletherapy did not speed up or slow down services on average for the year, nor did it increase or decrease the number of sessions participants completed, on average for the year, in Virginia. As discussed earlier, in the present study, there were also no significant differences in completion rate and status outcomes in the first year of Covid, as compared to the year prior to Covid.

Additional qualitative data collection would help to qualitatively assess components of FRAME that were not captured in the present study, in order to characterize this modification in greater detail. For example, it would be very useful to know if providers in Virginia made any specific delivery *process* or *content* modifications to FFT, while delivering FFT in a teletherapy format (e.g., more need for in-session engagement, modifying delivery of relevant strategies for teletherapy), and what impact this has on fidelity. It would also be interesting to collect qualitative data on how decisions about the use of teletherapy are *currently* considered, as the public health impact of the Covid-19 pandemic has declined considerably since 2020, but flexibility with use of teletherapy for certain families may remain useful or increase feasibility of FFT in Virginia.

The second modification identified by this evaluation was the use of FFT for participants outside of the intended age group for FFT. In Virginia, 0.6% of participants who received FFT

were younger than 11 years old, and a much larger group, 11.6% of participants, were older than 18 years old. This "extension" of FFT to youth outside of the intended age range is also considered a *context* modification, to the population served by the intervention. In particular, the percentage of older youth (19-22 years old) served in Virginia might point to FFT "filling a need" for this population that is not met by other services in Virginia.

Additional analyses showed that this group of older participants (19-22 years old) were disproportionately referred from juvenile justice related sources, as compared to those that were within the intended age range. In their article on adaptation in implementation, Baumann et al. (2017) state, "Often interventions have not been evaluated for a given population or context but there may be a pressing need to improve access to quality care. In such cases, interventions that are considered to be evidence based may represent the best option available." It is possible that this modification in Virginia to extend the age range for FFT reflects a dearth in other evidence-based services for "transition age youth," particularly those who are involved in the juvenile justice system. The term transition age youth often refers to individuals age 16-25, and this group is a unique subgroup in public systems, such as the juvenile justice system, because of the changes in educational, vocational, and relational roles, including reduced family influence and changing social networks, that occur during this developmental time period, as well as increased need for navigating transitions between child and adult services during this developmental time period (Zajac et al., 2015). In Virginia, youth may remain in the juvenile justice system through the age of 21 years old, and therefore the later end of the "transition age youth" time period may represent a significant gap in relevant evidence-based services for juvenile justice involved individuals, over the age of 18 in Virginia.

Although more work is needed to study effectiveness of FFT for this age group, results from this study are promising in showing that FFT completion rates for this older age group were similar to completion rate for participants who were within the intended age range for FFT, and rate of remaining in community was significantly higher for the older age group, as compared to the intended FFT age range group. Identifying this modification is an important first step in understanding this modification, although more work is needed. Additional qualitative data collection would help to answer questions related to FRAME components including, *who determined that the modification should be made, the extent to which the modification is fidelity-consistent, and the reasons for the modification*. For example, were any changes made to the expected format of having both the youth and family member(s) attend every session? Was there any additional developmentally appropriate content added to sessions, for example, content related to navigating adult systems of care or increasing independence from family?

Although a much smaller group (n=11), it is important to note that FFT in Virginia also served youth younger than the intended age group. There has been initial evidence in support of FFT-CW, an adapted version of FFT for younger youth (ages 0-18), who are child welfare clients. This version was developed with a greater for on "parent-driven" intervention strategies for younger clients (FFT-CW; Turner et al., 2017), although this adapted version of FFT is still building its evidence-base. It is *possible* that therapists in Virginia were drawing from components of this version when working with youth younger than the intended age range for FFT, although this is not known by the present study. Additional qualitative data collection would be helpful to elucidate how decisions were made to include participants that were outside of the intended age range for FFT (both older and younger), what influenced this modification,

and what, if any, modifications to the format or content of FFT were made to improve developmental or system-relevant fit for these extended populations.

Strengths and Limitations of Present Study

The present study included a number of strengths. This study evaluated the implementation and sustainment of FFT across a large service system (statewide) and included data from almost five years. The study also included a wide range of quality assurance targets to give a broad picture of implementation and sustainment quality in the state. Despite these and other strengths, there were also several limitations that warrant consideration when interpreting these findings. The present study was limited, in that it did not include a control group or comparison group, and it also did not include data from before the study, such as client's prior delinquency record or child welfare placement history, therefore differences in outcomes by group should be considered with this in mind. The study also only contains participant outcome data assessed by the therapist at case closure (e.g., no follow-up). Additionally, status outcome data were not available for some participants (between 23% - 30% of participants, for at least one status outcome). The majority of participants without outcome data did not complete FFT, which limits what we are able to know about non-completers overall and non-completers in comparison to completers in the study.

This study had a number of limitations related to fidelity measurement. First, fidelity was assessed only from the supervisor's perspective based on the therapist's recollection of their own practices, an approach that, although consistent with how FFT has assessed fidelity in past studies (Alexander et al., 2013; Robbins et al., 2016), has weaknesses. Additionally, although rating fidelity is part of supervisor training, it is not known if supervisors provide consistent fidelity ratings when presented with the same case material (reliability) or if the scores on the

measure translate into different levels of actual FFT fidelity (validity). Additionally, fidelity was not rated on a weekly basis, which would be expected, especially in a statewide implementation like this one. However, any interpretation of findings related to average team-level fidelity over time must bear in mind that fidelity scores were the average of available data quarterly, with ratings provided inconsistently among teams across quarters. As such, the present study does not interpret trends over time.

The scope of the current study was also limited in the operationalization of *Implementation* and operationalization of *Reach* as fully described by the RE-AIM framework. Future work could include a much more expansive evaluation of implementation components not captured in the current study, such as adherence to specific training components and timelines. Additionally, the present study was limited in its evaluation of who participated in FFT. In future work, it would be informative to evaluate reach as the total number of youth in the state that meet eligibility criteria for FFT (e.g., potentially using juvenile justice and/or child welfare system referral information), in order to assess the proportion of potentially eligible youth that actually are referred to FFT. This operationalization would be more comprehensively in line with the RE-AIM framework.

Additionally, the present study is limited in that the quality assurance targets assessed (e.g., team size, caseload, treatment duration, frequency of fidelity rating) are standards set by FFT for implementation but have not been tested individually and the direct impact of each of these components on client outcomes is not known. Finally, as noted earlier, two teams declined to release their data for use in the present study, so the present study cannot provide a full picture of FFT in Virginia and is limited to data for the clients served by the nine teams that provided data.

Future Directions

There are several future research and policy directions to consider from these findings. Due to the present study being a secondary data analysis of data collected as part of a statewide implementation of FFT, there was not the opportunity to include measures and methods prospectively that could have improved the potential impact of the findings. Changes that could have benefitted the present study include integrating relevant outcome data from existing juvenile justice and child welfare databases, to examine variables such as client's prior juvenile justice involvement and child welfare placement history. Additionally, the current study outcome measures were gathered by FFT at case closure. To look at these outcomes more long-term, it would be beneficial to collect additional follow up data, one year after participating in FFT, for example. The current study was limited even with data available at case closure, as sometimes therapists were not able to obtain status outcome data for some cases that closed before completion of FFT.

Research attention has been given to the factors influencing implementation and sustainment phases. Examples of research foci include training, supervision, leadership, therapist attitudes, and organizational climate (Fixsen, Naoom, Blase, Friedman, & Wallace, 2005; Glisson et al., 2008). Researchers in the field of implementation science have developed implementation and sustainment-specific measures of leadership, climate, and citizenship behavior (Aarons et al., 2014: Weiner et al., 2011; Ehrhart et al., 2015), which could be used to assess these concepts and their impact on FFT implementation and sustainment.

The policy landscape for EBPs continues to transform in Virginia, especially with introduction of the Family First Prevention Services Act. Future work could examine the impact of such policy changes on aspects of FFT such as referral flow, team size, turnover, and case

completion. It is notable that Virginia's Family First plan was approved July 1, 2021, in the midst of significant national workforce challenges (Crocker et al., 2023, Rossi et al., 2023, Brabson et al., 2020) also reflected in the results of this study, as indicated by the significant drop in team sizes beginning in 2020. The present program evaluation can serve as a baseline assessment for the implementation of an evidence-based program like FFT. In addition, it may be useful to consider collection of qualitative data from FFT teams across Virginia to understand how regional differences may influence implementation of the program (Virginia has five regions and 133 localities). Of note, while most states have state administered social services systems, Virginia is one of nine states to have a locally administered child welfare system, (Child Welfare Information Gateway, 2018). This means that the state can set guidelines and some policies, but individual counties also make key decisions, such as funding of services. This leads to the potential for varying procedures and preferences, and more variability in performance of the child welfare system across counties, as compared to state-administered states.

One way to integrate qualitative and quantitative approaches is to conduct a "landscape analysis" to provide a rich picture of the system context in which an evidence-based practice is being implemented and capture local knowledge (Drainoni et al., 2022). Hoagwood et al. (2012) suggest that behavioral health services research is more impactful when it assesses a wide range of treatment outcomes, which broadens its relevance to more stakeholders. Future research should examine additional indicators of quality to stakeholders (such as funders and policy makers), including cost effectiveness of FFT in Virginia as part of program evaluation of FFT over time.

Future research would also benefit from a focus on physical geographic differences in Virginia. For example, it would be useful to look at differences between regions and teams in

different parts of the state. Geographic information such as differences in spread of clients (and therefore provider time spent traveling between clients), available referral sources in different geographic areas, as well as local knowledge of FFT would all be informative. Previous research has shown that collaboration across teams within the same geographic regions can support implementation and sustainment (Aarons et al., 2014). Although there have been some opportunities for collaboration across the state on topics related to FFT, there is not a formal mechanism for FFT team leads and administrators to collaborate on an ongoing basis. Developing a more formal format for cross-team communication and collaboration could be a helpful future direction for sustaining FFT across the state.

As noted earlier, although this study found decreasing team sizes over the study period, it did not capture reasons for this decrease. Future work would benefit from collection of qualitative data (for example, exit interviews with staff, interviews with hiring managers) to elucidate reasons for staff no longer holding their positions and these positions not being refilled. Additionally, qualitative data collection would be particularly helpful in elaborating on and elucidating modifications to FFT related to use of the teletherapy format and the use of FFT with individuals outside of the intended age range, as well as the implementation of FFT with less frequent fidelity monitoring. Quantitative or mixed-methods studies should include these modifications and drift to further assess impact on implementation and participant outcomes. As noted throughout, research and policy must also continue to address racial disparity in the juvenile justice system and continue to examine racial differences in outcomes of FFT participation. Although the number of youth who come into formal contact with the court system has declined in recent years, youth of color are disproportionately represented at every decision point of the juvenile delinquency court process (Children's Bureau, 2016) and ongoing calls to

reform policies and practices that drive bias and structural racism persist (National Council of Juvenile and Family Court Judges, 2019).

This study conducted a program evaluation and observational study of the implementation and sustainment of FFT during the first five years of implementation in the state of Virginia. The study highlighted both strengths and weaknesses across quality assurance targets for FFT implementation. It is particularly notable that challenges noted in the data have an onset coincident with the Covid-19 pandemic and with a significant change in Medicaid funding for FFT. Recommendations were developed to support the ongoing implementation and sustainment of FFT across the state and modifications to FFT were identified and discussed. Implementing and sustaining evidence-based practices in statewide systems is a complex and challenging task, especially for a complex community-delivered intervention like FFT. This program evaluation can serve as an initial base for ongoing quality improvement and evaluation efforts in Virginia as implementation and sustainment progresses.

References

Aarons, G. A., Ehrhart, M. G., & Farahnak, L. R. (2014). The implementation leadership scale (ILS): development of a brief measure of unit level implementation leadership. *Implementation Science*, 9, 1-10.

Aarons, G. A., Ehrhart, M. G., Moullin, J. C., Torres, E. M., & Green, A. E. (2017). Testing the leadership and organizational change for implementation (LOCI) intervention in substance abuse treatment: a cluster randomized trial study protocol. *Implementation Science*, *12*, 1-11.

Aarons, G. A., Fettes, D. L., Hurlburt, M. S., Palinkas, L. A., Gunderson, L., Willging, C. E., & Chaffin, M. J. (2014). Collaboration, negotiation, and coalescence for interagency-collaborative teams to scale-up evidence-based practice. *Journal of Clinical Child & Adolescent Psychology*, 43(6), 915-928.

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

Aarons, G. A., Hurlburt, M., & Horwitz, S. M. (2011). Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health and Mental Health Services Research*, *38*(1), 4-23.

Aarons, G. A., Sommerfeld, D. H., Hecht, D. B., Silovsky, J. F., & Chaffin, M. J. (2009). The impact of evidence-based practice implementation and fidelity monitoring on staff turnover: evidence for a protective effect. *Journal of Consulting and Clinical Psychology*, *77*(2), 270.

Abrams, L. S., Mizel, M. L., & Barnert, E. S. (2021). The criminalization of young children and overrepresentation of black youth in the juvenile justice system. *Race and Social Problems*, *13*, 73-84.

Alexander, J. F., & Robbins, M. S. (2011). Functional family therapy. *Clinical handbook of assessing and treating conduct problems in youth*, 245-271.

Robbins, M., & Alexander, J. F. (2019). Training Functional Family Therapists. In *Encyclopedia of Couple and Family Therapy* (pp. 2981-2985). Cham: Springer International Publishing.

Alexander, J. F., & Parsons, B. V. (1973). Short-term behavioral intervention with delinquent families: impact on family process and recidivism. *Journal of Abnormal Psychology*, *81*(3), 219.

Alexander, J. F., Waldron, H. B., Robbins, M. S., & Neeb, A. A. (2013). *Functional family therapy for adolescent behavior problems*. American Psychological Association.

Annie E. Casey Foundation. (2012). Reducing congregate care: Worth the fight. Retrieved from http://www.aecf.org/blog/reducing-congregate-care-worth-the-fight

Annie E. Casey Foundation. (2015). Too Many Teens: Preventing Out-of-Home Placements. Retrieved from https://www.aecf.org/resources/too-many-teens

Awa, W. L., Plaumann, M., & Walter, U. (2010). Burnout prevention: A review of intervention programs. *Patient Education and Counseling*, *78*(2), 184-190.

Baldwin, S. A., Christian, S., Berkeljon, A., & Shadish, W. R. (2012). The effects of family therapies for adolescent delinquency and substance abuse: A meta-analysis. *Journal of Marital and Family Therapy*, 38, 281–30 4.

Barnowski, R. (2002). Preliminary Findings for the Juvenile Rehabilitation Administration's Dialectic Behavior Therapy Program. Retrieved from http://www.wsipp.wa.gov/ReportFile/803/Wsipp_WashingtonStatesImplementation-of-Functional-Family-Therapy-for-Juvenile-OffendersPreliminaryFindings_Full-Report.pdf

Bartollas, C. (2010). *Juvenile justice in America*. Pearson.

Baumann, A., Cabassa, L. J., Stirman, S. W., & Brownson, R. (2017). Adaptation in dissemination and implementation science. *Dissemination and implementation research in health: translating science to practice*, *2*, 286-300.

Bearman, S. K., Schneiderman, R. L., & Zoloth, E. (2017). Building an evidence base for effective supervision practices: An analogue experiment of supervision to increase EBT fidelity. *Administration and Policy in Mental Health and Mental Health Services Research*, *44*, 293-307.

Becker, K. D., Boustani, M., Gellatly, R., & Chorpita, B. F. (2018). Forty years of engagement research in children's mental health services: Multidimensional measurement and practice elements. *Journal of Clinical Child & Adolescent Psychology*, *47*(1), 1-23.

Becker, K. D., & Chorpita, B. F. (2023). Future directions in youth and family treatment engagement: finishing the bridge between science and service. *Journal of Clinical Child & Adolescent Psychology*, *52*(2), 284-309.

Beidas, R. S., Marcus, S., Wolk, C. B., Powell, B., Aarons, G. A., Evans, A. C., ... & Mandell, D. S. (2016). A prospective examination of clinician and supervisor turnover within the context of implementation of evidence-based practices in a publicly-funded mental health system. *Administration and Policy in Mental Health and Mental Health Services Research*, *43*(5), 640-649.

Beidas, R. S., Stewart, R. E., Adams, D. R., Fernandez, T., Lustbader, S., Powell, B. J., & Barg, F. K. (2016). A multi-level examination of stakeholder perspectives of implementation of evidence-based practices in a large urban publicly-funded mental health system. *Administration and Policy in Mental Health and Mental Health Services Research*, *43*(6), 893-908.

Birken, S. A., Haines, E. R., Hwang, S., Chambers, D. A., Bunger, A. C., & Nilsen, P. (2020). Advancing understanding and identifying strategies for sustaining evidence-based practices: a review of reviews. *Implementation Science*, *15*(1), 1-13.

Bond, M., Bowling, A., Abery, A., McClay, M., & Dickinson, E. (2000). Evaluation of outreach clinics held by specialists in general practice in England. *Journal of Epidemiology & Community Health*, *54*(2), 149-156.

Bond, G. R., & Drake, R. E. (2020). Assessing the fidelity of evidence-based practices: History and current status of a standardized measurement methodology. *Administration and Policy in Mental Health and Mental Health Services Research, 47*, 874–884.

Brabson, L. A., Harris, J. L., Lindhiem, O., & Herschell, A. D. (2020). Workforce turnover in community behavioral health agencies in the USA: A systematic review with recommendations. *Clinical Child and Family Psychology Review*, *23*(3), 297-315.

Braveman, P. (2014). What are health disparities and health equity? We need to be clear. *Public Health Reports*, 129, 5–8.

Bronfenbrenner, U. (1979). Contexts of child rearing: Problems and prospects. *American Psychologist*, *34*(10), 844.

Bronsard, G., Alessandrini, M., Fond, G., Loundou, A., Auquier, P., Tordjman, S., & Boyer, L. (2016). The prevalence of mental disorders among children and adolescents in the child welfare system: a systematic review and meta-analysis. *Medicine*, *95*(7).

Bruns, E. J., Benjamin, P. H., Shepler, R. N., Kellogg, M., Pluckebaum, H., Woolston, J. L., ... & Zabel, M. D. (2021). Manuscript: Defining Quality Standards for Intensive Home Based Treatment Programs for Youth with Serious Emotional Disorders. *Administration and Policy in Mental Health and Mental Health Services Research*, *48*(6), 1065-1088.

Buchanan, M., Castro, E. D., Kushner, M., & Krohn, M. D. (2020). It's F** ing Chaos: COVID-19's impact on juvenile delinquency and juvenile justice. *American Journal of Criminal Justice*, *45*, 578-600.

Burkhardt, J. T., Schröter, D. C., Magura, S., Means, S. N., & Coryn, C. L. (2015). An overview of evidence-based program registers (EBPRs) for behavioral health. *Evaluation and Program Planning*, *48*, 92-99.

Bryson, S. A., Akin, B. A., Blase, K. A., McDonald, T., & Walker, S. (2014). Selecting an EBP to reduce long-term foster care: Lessons from a university–child welfare agency partnership. *Journal of Evidence-Based Social Work*, *11*(1-2), 208-221.

Celinska, K., & Cheng, C. C. (2017). Gender Comparisons in the Processes and Outcomes of Functional Family Therapy. *Journal of Juvenile Justice*, *6*(1).

Chambers, D. A., Glasgow, R. E., & Stange, K. C. (2013). The dynamic sustainability framework: addressing the paradox of sustainment amid ongoing change. *Implementation Science*, *8*, 1-11.

Children's Bureau (2020). Virginia General Child Population. Retrieved from https://cwoutcomes.acf.hhs.gov/cwodatasite/pdf/virginia.html.

Children's Bureau, U.S. Department of Health & Human Services (2016). Racial Disproportionality and Disparity in Child Welfare. Retrieved from https://www.childwelfare.gov/pubPDFs/racial_disproportionality.pdf.

Child Welfare Information Gateway. (2018). *State vs. county administration of child welfare services*. Washington, DC: U.S. Department of Health and Human Services, Children's Bureau.

Chorpita, B. F., Daleiden, E. L., & Weisz, J. R. (2005). Identifying and selecting the common elements of evidence based interventions: A distillation and matching model. *Mental Health Services Research*, 7(1), 5-20.

Chorpita, B. F., Yim, L. M., Donkervoet, J. C., Arensdorf, A., Amundsen, M. J., McGee, C., Serrano, A., Yates, A., Burns, J. A., & Morelli, P. (2002). Toward large-scale implementation of empirically supported treatments for children: A review and observations by the Hawaii Empirical Basis to Services Task Force. *Clinical Psychology: Science and Practice*, *9*(2), 165-190.

Collyer, H., Eisler, I., & Woolgar, M. (2021). Parent and youth perspectives and retention in functional family therapy. *Family Process*, *60*(2), 316-330.

Collyer, H., Eisler, I., & Woolgar, M. (2020). Systematic literature review and meta-analysis of the relationship between adherence, competence and outcome in psychotherapy for children and adolescents. *European Child & Adolescent Psychiatry*, 29, 417-431.

Cook, B. L., Hou, S. S. Y., Lee-Tauler, S. Y., Progovac, A. M., Samson, F., & Sanchez, M. J. (2019). A review of mental health and mental health care disparities research: 2011-2014. *Medical Care Research and Review*, *76*(6), 683-710.

Cooper, B. R., Bumbarger, B. K., & Moore, J. E. (2015). Sustaining evidence-based prevention programs: Correlates in a large-scale dissemination initiative. *Prevention Science*, *16*, 145-157.

Covino, N. A. (2019). Developing the behavioral health workforce: Lessons from the states. *Administration and Policy in Mental Health and Mental Health Services Research*, *46*(6), 689-695.

Cox, J. R., Martinez, R. G., & Southam-Gerow, M. A. (2019). Treatment integrity in psychotherapy research and implications for the delivery of quality mental health services. *Journal of Consulting and Clinical Psychology*, *87*(3), 221.

Crocker, K. M., Gnatt, I., Haywood, D., Butterfield, I., Bhat, R., Lalitha, A. R. N., ... & Castle, D. J. (2023). The impact of COVID-19 on the mental health workforce: A rapid review. *International Journal of Mental Health Nursing*, *32*(2), 420-445.

Cuddeback, G. S., Morrissey, J. P., Domino, M. E., Monroe-DeVita, M., Teague, G. B., & Moser, L. L. (2013). Fidelity to recovery-oriented ACT practices and consumer outcomes. *Psychiatric Services*, *64*(4), 318-323.

De Haan, A. M., Boon, A. E., de Jong, J. T., Hoeve, M., & Vermeiren, R. R. (2013). A meta-analytic review on treatment dropout in child and adolescent outpatient mental health care. *Clinical Psychology Review,* 33(5), 698–711.

Drainoni, M. L., Knudsen, H. K., Adams, K., Andrews-Higgins, S. A., Auritt, V., Back, S., ... & McAlearney, A. S. (2022). Community coalition and key stakeholder perceptions of the community opioid epidemic before an intensive community-level intervention. *Journal of Substance Abuse Treatment*, *138*, 108731.

DuMont, K., Metz, A., & Woo, B. (2019). Five recommendations for how implementation science can better advance equity. *Washington, DC: Academy Health*, 2019-04.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*(3), 327-350.

Eccles, M. P., & Mittman, B. S. (2006). Welcome to *Implementation Science*. *Implementation Science, 1,* 1.

Ehrhart, M. G., Aarons, G. A., & Farahnak, L. R. (2015). Going above and beyond for implementation: the development and validity testing of the Implementation Citizenship Behavior Scale (ICBS). *Implementation Science,* 10, 1-9.

Ehrhart, M. G., & Kuenzi, M. (2017). The impact of organizational climate and culture on employee turnover. *The Wiley Blackwell handbook of the psychology of recruitment, selection and employee retention*, 494-512.

Eme, R., Beauchaine, T. P., & Hinshaw, S. P. (2016). Sex differences in the prevalence and expression of externalizing behavior. *The Oxford Handbook of Externalizing Spectrum Disorders*, 239-263.

Epstein, R. A., Fonnesbeck, C., Potter, S., Rizzone, K. H., & McPheeters, M. (2015). Psychosocial interventions for child disruptive behaviors: A meta-analysis. *Pediatrics*, *136*(5), 947-960.

FFT LLC Phases of FFT Training (n.d.). Retrieved September 19, 2024, from https://www.fftllc.com/fft-training

Fish, J. N., & Mittal, M. (2021). Mental health providers during COVID-19: essential to the US public health workforce and in need of support. *Public Health Reports*, *136*(1), 14-17.

Fixsen, D., Blase, K, Naoom, S., & Duda, M. (2015). Implementation drivers: Assessing best practice. National Implementation Science Network, University of North Carolina Chapel Hill. Retrieved from https://implementation.fpg.unc.edu/sites/implementation.fpg.unc.edu/files/NIRNImplementationDriversAssessingBestPractices.pdf

Fukui, S., Wu, W., & Salyers, M. P. (2019). Mediational paths from supervisor support to turnover intention and actual turnover among community mental health providers. *Psychiatric Rehabilitation Journal*, *42*(4), 350.

Garland, A. F., Bickman, L., & Chorpita, B. F. (2010). Change what? Identifying quality improvement targets by investigating usual mental health care. *Administration and Policy in Mental Health and Mental Health Services Research*, *37*(1-2), 15-26.

Garland, A. F., Haine-Schlagel, R., Brookman-Frazee, L., Baker-Ericzen, M., Trask, E., & Fawley-King, K. (2013). Improving community-based mental health care for children: Translating knowledge into action. *Administration and Policy in Mental Health and Mental Health Services Research*, *40*, 6-22.

Glisson, C., Dukes, D., & Green, P. (2006). The effects of the ARC organizational intervention on caseworker turnover, climate, and culture in children's service systems. *Child Abuse & Neglect*, *30*(8), 855-880.

Glisson, C., Landsverk, J., Schoenwald, S., Kelleher, K., Hoagwood, K. E., Mayberg, S., ... & Research Network on Youth Mental Health. (2008). Assessing the organizational social context (OSC) of mental health services: Implications for research and practice. *Administration and Policy in Mental Health and Mental Health Services Research*, *35*(1-2), 98.

Glisson, C., Landsverk, J., Schoenwald, S., Kelleher, K., Hoagwood, K. E., Mayberg, S., ... & Research Network on Youth Mental Health. (2008). Assessing the organizational social context (OSC) of mental health services: Implications for research and practice. *Administration and Policy in Mental Health and Mental Health Services Research*, *35*, 98-113.

Glisson, C., Schoenwald, S. K., Kelleher, K., Landsverk, J., Hoagwood, K. E., Mayberg, S., ... & Research Network on Youth Mental Health. (2008). Therapist turnover and new program sustainability in mental health clinics as a function of organizational culture, climate, and service structure. *Administration and Policy in Mental Health and Mental Health Services Research*, *35*, 124-133.

Gomes, R., Battaglia, C., Fort, M., Maw, M., McCreight, M., Rabin, B., Robertson, E., Studts, C., Trinkley, K., & Glasgow, R. (n.d.). A guidebook to the pragmatic and iterative use of the Practical, Robust Implementation and Sustainability Model (PRISM) and Reach, Effectiveness, Adoption, Implementation, Maintenance framework (RE-AIM) for planning, implementation, and sustainment. Retrieved on September 24, 2024 from https://medschool.cuanschutz.edu/docs/librariesprovider94/di-docs/guides-and-tools/iprism-and-reaim-guidebook_wip.pdf?sfvrsn=adea27bb_0

Gottfredson, D. C., Kearley, B., Thornberry, T. P., Slothower, M., Devlin, D., & Fader, J. J. (2018). Scaling-up evidence-based programs using a public funding stream: A randomized trial of Functional Family Therapy for court-involved youth. *Prevention Science*, *19*(7), 939-953.

Green, A. E., Albanese, B. J., Shapiro, N. M., & Aarons, G. A. (2014). The roles of individual and organizational factors in burnout among community-based mental health service providers. *Psychological Services*, *11*(1), 41.

Green, A. E., Miller, E. A., & Aarons, G. A. (2013). Transformational leadership moderates the relationship between emotional exhaustion and turnover intention among community mental health providers. *Community Mental Health Journal*, *49*, 373-379.

Haine-Schlagel, R., & Walsh, N. E. (2015). A review of parent participation engagement in child and family mental health treatment. *Clinical Child and Family Psychology Review*, *18*, 133-150.

Hallett, E., Simeon, E., Amba, V., Howington, D., McConnell, K. J., & Zhu, J. M. (2024). Factors influencing turnover and attrition in the public behavioral health system workforce: Qualitative study. *Psychiatric Services*, *75*(1), 55-63.

Hallfors, D., & Cho, H. (2007). Moving behavioral science from efficacy to effectiveness. *International Journal of Behavioral Consultation and Therapy*, *3*(2), 236.

Hartnett, D., Carr, A., Hamilton, E., & O'Reilly, G. (2017). The effectiveness of functional family therapy for adolescent behavioral and substance misuse problems: A meta-analysis. *Family Process*, *56*(3), 607-619.

Hasson, H. (2010). Systematic evaluation of implementation fidelity of complex interventions in health and social care. *Implementation Science*, 5(67).

Henggeler, S. W., Melton, G. B., Brondino, M. J., Scherer, D. G., & Hanley, J. H. (1997). Multisystemic therapy with violent and chronic juvenile offenders and their families: The role of treatment fidelity in successful dissemination. *Journal of Consulting and Clinical Psychology*, 65(5), 821–833.

Higa-McMillan, C. K., Powell, C., Daleiden, E. L., & Mueller, C. W. (2011). Pursuing an evidence-based culture through contextualized feedback: Aligning youth outcomes and practices. *Professional Psychology: Research and Practice*, *42*(2), 137.

Hoagwood, K. E., Jensen, P. S., Acri, M. C., Olin, S. S., Lewandowski, R. E., & Herman, R. J. (2012). Outcome domains in child mental health research since 1996: Havethey changed and why does it matter?. Journal of the *American Academy of Child and Adolescent Psychiatry,* 51(12), 1241–1260.

Hoagwood, K. E., Olin, S. S., Horwitz, S., McKay, M., Cleek, A., Gleacher, A., Lewandowski, E., Nadeem, E., Acri, M., Chor, K. H., Kuppinger, A., Burton, G., Weiss, D., Frank, S., Finnerty, M., Bradbury, D. M., Woodlock, K. M., & Hogan, M. (2014). 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 & Adolescent Psychology,* 43(2), 145.

Hodgkinson, S., Godoy, L., Beers, L. S., & Lewin, A. (2017). Improving mental health access for low-income children and families in the primary care setting. *Pediatrics,* 139(1), e20151175.

Henggeler, S.W., Pickrel, S.G., & Brondino, M. J. (1999). Multisystemic treatment of substance-abusing and -dependent delinquents: Outcomes, treatment fidelity, and transportability. *Mental Health Services Research*. 1(3), 171–184.

Hoagwood, K., Burns, B. J., Kiser, L., Ringeisen, H., & Schoenwald, S. K. (2001). Evidence-based practice in child and adolescent mental health services. *Psychiatric Services*, *52*(9), 1179-1189.

Hoffmann, J. A., Alegría, M., Alvarez, K., Anosike, A., Shah, P. P., Simon, K. M., & Lee, L. K. (2022). Disparities in pediatric mental and behavioral health conditions. *Pediatrics*, *150*(4).

Hogue, A., Bobek, M., Dauber, S., Henderson, C. E., McLeod, B. D., & Southam-Gerow, M. A. (2019). Core elements of family therapy for adolescent behavior problems: Empirical distillation of three manualized treatments. *Journal of Clinical Child & Adolescent Psychology*, *48*(1), 29-41.

Hogue, A., Bobek, M., Dauber, S., Henderson, C. E., McLeod, B. D., & Southam-Gerow, M. A. (2017). Distilling the core elements of family therapy for adolescent substance use: conceptual and empirical solutions. *Journal of Child & Adolescent Substance Abuse*, *26*(6), 437-453.

Hogue, A., Henderson, C. E., Ozechowski, T. J., & Robbins, M. S. (2014). Evidence base on outpatient behavioral treatments for adolescent substance use: Updates and recommendations 2007–2013. *Journal of Clinical Child & Adolescent Psychology*, *43*(5), 695-720.

Hogue, A., Henderson, C. E., Dauber, S., Barajas, P. C., Fried, A., & Liddle, H. A. (2008). Treatment adherence, competence, and outcome in individual and family therapy for adolescent behavior problems. *Journal of Consulting and Clinical Psychology*, *76*(4), 544.

Hogue, A., Henderson, C. E., Ozechowski, T. J., & Robbins, M. S. (2014). Evidence base on outpatient behavioral treatments for adolescent substance use: Updates and recommendations 2007–2013. *Journal of Clinical Child & Adolescent Psychology*, *43*(5), 695-720.

Hogue, A., Ozechowski, T. J., Robbins, M. S., & Waldron, H. B. (2013). Making fidelity an intramural game: Localizing quality assurance procedures to promote sustainability of

evidence-based practices in usual care. *Clinical Psychology: Science and Practice*, *20*(1), 60.

Huey Jr, S. J., Henggeler, S. W., Brondino, M. J., & Pickrel, S. G. (2000). Mechanisms of change in multisystemic therapy: reducing delinquent behavior through therapist adherence and improved family and peer functioning. *Journal of Consulting and Clinical Psychology*, *68*(3), 451.

Hurlburt, M. S., Garland, A. F., Nguyen, K., & Brookman-Frazee, L. (2010). Child and family therapy process: Concordance of therapist and observational perspectives. *Administration and Policy in Mental Health and Mental Health Services Research, 37*, 230–244.

Jensen, P. S., & Foster, M. (2010). Closing the research to practice gap in children's mental health: Structures, solutions, and strategies. *Administration and Policy in Mental Health and Mental Health Services Research*, *37*(1-2), 111-119.

Joint Legislative Audit and Review Commission (2021). Virginia's Juvenile Justice System. Retrieved from https://jlarc.virginia.gov/pdfs/reports/Rpt558-2.pdf

Kazdin, A. E. (2003). Psychotherapy for children and adolescents. *Annual Review of Psychology*, *54*, 253.

Karver, M. S., Handelsman, J. B., Fields, S., & Bickman, L. (2006). Meta-analysis of therapeutic relationship variables in youth and family therapy: The evidence for different relationship variables in the child and adolescent treatment outcome literature. *Clinical Psychology Review*, *26*(1), 50-65.

Knudsen, H. K., Ducharme, L. J., & Roman, P. M. (2008). Clinical supervision, emotional exhaustion, and turnover intention: A study of substance abuse treatment counselors in

the Clinical Trials Network of the National Institute on Drug Abuse. *Journal of Substance Abuse Treatment*, *35*(4), 387-395.

Lau, A. S. (2006). Making the case for selective and directed cultural adaptations of evidence-based treatments: examples from parent training. *Clinical Psychology: Science and Practice*, *13*(4), 295-310.

Lau, A. S., & Brookman-Frazee, L. (2015). The 4KEEPS study: identifying predictors of sustainment of multiple practices fiscally mandated in children's mental health services. *Implementation Science*, *11*, 31.

Leijten, P., Raaijmakers, M. A., de Castro, B. O., & Matthys, W. (2013). Does socioeconomic status matter? A meta-analysis on parent training effectiveness for disruptive child behavior. *Journal of Clinical Child & Adolescent Psychology*, *42*(3), 384-392.

Liddle, H. A., Dakof, G. A., Parker, K., Diamond, G. S., Barrett, K., & Tejeda, M. (2001). Multidimensional family therapy for adolescent drug abuse: Results of a randomized clinical trial. *The American Journal of Drug and Alcohol Abuse*, *27*(4), 651-688.

Liu, J. (2004). Childhood externalizing behavior: Theory and implications. *Journal of Child and Adolescent Psychiatric Nursing*, *17*(3), 93-103.

Martino, S., Paris Jr, M., Añez, L., Nich, C., Canning-Ball, M., Hunkele, K., ... & Carroll, K. M. (2016). The effectiveness and cost of clinical supervision for motivational interviewing: a randomized controlled trial. *Journal of Substance Abuse Treatment*, *68*, 11-23.

McCart, M. R., & Sheidow, A. J. (2016). Evidence-based psychosocial treatments for adolescents with disruptive behavior. *Journal of Clinical Child & Adolescent Psychology*, *45*(5), 529-563.

McLeod, B. D., Southam-Gerow, M. A., Tully, C. B., Rodríguez, A., & Smith, M. M. (2013). Making a case for treatment integrity as a psychosocial treatment quality indicator for youth mental health care. *Clinical Psychology: Science and Practice*, *20*(1), 14.

McPherson, K. E., Kerr, S., Casey, B., & Marshall, J. (2017). Barriers and Facilitators to Implementing Functional Family Therapy in a Community Setting: Client and Practitioner Perspectives. *Journal of marital and family therapy*, *43*(4), 717–732.

Merikangas, K. R., He, J. P., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., ... & Swendsen, J. (2010). Lifetime prevalence of mental disorders in US adolescents: results from the National Comorbidity Survey Replication–Adolescent Supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*(10), 980-989.

Merkel-Holguin, L. (2017). *A history of child welfare*. Routledge.

Meza, R. D., AlRasheed, R., Pullmann, M. D., & Dorsey, S. (2023). Clinical supervision approach predicts evidence-based trauma treatment delivery in children's mental health. *Frontiers in Psychiatry*, *13*, 1072844.

Minuchin, S. (1974). Families & family therapy. Cambridge, MA: Harvard University Press.

National Council of Juvenile and Family Court Judges (2019). *Addressing bias in delinquency and child welfare systems*. Retrieved from https://www.ncjfcj.org/wp-content/uploads/2019/07/Addressing-Bias-Bench-Card-1-1.pdf

National Institute for Health Care Management (2023). *The Behavioral Health Care Workforce*. Retrieved from https://nihcm.org/publications/the-behavioral-health-care-workforce-shortages-solutions

NICE. (2013). Antisocial behaviour and conduct disorders in children and young people: Recognition, intervention and management. London: NICE.

Novins, D. K., Green, A. E., Legha, R. K., & Aarons, G. A. (2013). Dissemination and implementation of evidence-based practices for child and adolescent mental health: A systematic review. *Journal of the American Academy of Child & Adolescent Psychiatry*, *52*(10), 1009-1025.

Olseth, A. R., Hagen, K. A., Keles, S., & Bjørnebekk, G. (2024). Functional family therapy for adolescent disruptive behavior in Norway: Results from a randomized controlled trial. *Journal of Family Psychology*.

Orimoto, T. E., Higa-McMillan, C. K., Mueller, C. W., & Daleiden, E. L. (2012). Assessment of therapy practices in community treatment for children and adolescents. *Psychiatric Services*, *63*(4), 343-350.

Orimoto, T. E., Mueller, C. W., Hayashi, K., & Nakamura, B. J. (2014). Community-based treatment for youth with co-and multimorbid disruptive behavior disorders. *Administration and Policy in Mental Health and Mental Health Services Research*, *41*(2), 262-275.

Pascoe, K. M., Petrescu-Prahova, M., Steinman, L., Bacci, J., Mahorter, S., Belza, B., & Weiner, B. (2021). Exploring the impact of workforce turnover on the sustainability of evidence-based programs: a scoping review. *Implementation Research and Practice*, *2*.

Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science and Practice*, *12*(4), 365.

Pinals, D. A., Hepburn, B., Parks, J., & Stephenson, A. H. (2020). The behavioral health system and its response to COVID-19: a snapshot perspective. *Psychiatric Services*, *71*(10), 1070-1074.

Price, J. M., Chiapa, A., & Walsh, N. E. (2013). Predictors of externalizing behavior problems in early elementary-aged children: The role of family and home environments. *The Journal of Genetic Psychology*, *174*(4), 464-471.

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(2), 65–76.

Powell, B. J., Mettert, K. D., Dorsey, C. N., Weiner, B. J., Stanick, C. F., Lengnick-Hall, R., ... & Lewis, C. C. (2021). Measures of organizational culture, organizational climate, and implementation climate in behavioral health: A systematic review. *Implementation Research and Practice*, *2*.

Reinert, M, Fritze, D. & Nguyen, T. (2021). *The state of mental health in America 2022*. Mental Health America, Alexandria VA.

Robbins, M. S., Alexander, J. F., Turner, C. W., & Hollimon, A. (2016). Evolution of functional family therapy as an evidence-based practice for adolescents with disruptive behavior problems. *Family Process*, *55*(3), 543-557.

Robbins, M. S., Hollimon, A., Alexander, J. F., Mason, K., Armey, K., & Neeb, A. (2011). *The process of supervision in Functional Family Therapy*.

Robbins, M. S., Mayorga, C. C., Mitrani, V. B., Szapocznik, J., Turner, C. W., & Alexander, J. F. (2008). Adolescent and parent alliances with therapists in brief strategic family therapy

with drug-using hispanic adolescents. *Journal of Marital and Family Therapy*, 34(3), 316–328.

Robbins, M. S., & Midouhas, H. (2021). Adapting the delivery of functional family therapy around the world during a global pandemic. *Global Implementation Research and Applications*, *1*(2), 109-121.

Robbins, M. S., Waldron, H. B., Turner, C. W., Brody, J., Hops, H., & Ozechowski, T. (2019). Evaluating supervision models in functional family therapy: Does adding observation enhance outcomes?. *Family Process*, *58*(4), 873-890.

Rossi, M. F., Gualano, M. R., Magnavita, N., Moscato, U., Santoro, P. E., & Borrelli, I. (2023). Coping with burnout and the impact of the COVID-19 pandemic on workers' mental health: a systematic review. *Frontiers in Psychiatry*, *14*, 1139260.

Sale, R., Wu, J., Robinson, A., Finn, N., Aisenberg, G., Kaur, N., Riso, A., & Southam-Gerow, M. A. (2023). Workforce and Other Factors Impeding Implementation and Sustainment of FFPSA Evidence-Based Programs: A Study of Obstacles and Opportunities.

Scanlan, J. N., Still, M., Stewart, K., & Croaker, J. (2010). Recruitment and retention issues for occupational therapists in mental health: Balancing the pull and the push. *Australian Occupational Therapy Journal*, *57*(2), 102-110.

Sedlar, G., Bruns, E. J., Walker, S. C., Kerns, S. E., & Negrete, A. (2017). Developing a quality assurance system for multiple evidence based practices in a statewide service improvement initiative. *Administration and Policy in Mental Health and Mental Health Services Research*, *44*(1), 29-41.

Schoenwald, S. K., Garland, A. F., Chapman, J. E., Frazier, S. L., Sheidow, A. J., & Southam-Gerow, M. A. (2011). Toward the effective and efficient measurement of implementation

fidelity. *Administration and Policy in Mental Health and Mental Health Services Research*, *38*(1), 32-43.

Schoenwald, S. K., Sheidow, A. J., & Chapman, J. E. (2009). Clinical supervision in treatment transport: effects on adherence and outcomes. *Journal of Consulting and Clinical Psychology*, *77*(3), 410.

Schoenwald, S. K., Sheidow, A. J., Letourneau, E. J., & Liao, J. G. (2003). Transportability of multisystemic therapy: Evidence for multilevel influences. *Mental Health Services Research, 5*(4), 223–239

Sexton, T., & Turner, C. W. (2010). The effectiveness of functional family therapy for youth with behavioral problems in a community practice setting. *Journal of Family Psychology*, *24*(3), 339.

Sheidow, A. J., McCart, M. R., & Drazdowski, T. K. (2022). Family-based treatments for disruptive behavior problems in children and adolescents: An updated review of rigorous studies (2014–April 2020). *Journal of Marital and Family Therapy*, *48*(1), 56-82.

Shelton, R. C., & Brownson, R. C. (2024). Enhancing impact: a call to action for equitable implementation science. *Prevention Science*, *25*, 174-189.

Shelton, R. C., Chambers, D. A., and Glasgow, R. E. (2020). An extension of RE-AIM to enhance sustainability: Addressing dynamic context and promoting health equity over time. *Frontiers in Public Health,* 8, 134.

Shelton, R. C., Cooper, B. R., & Stirman, S. W. (2018). The sustainability of evidence-based interventions and practices in public health and health care. *Annual Review of Public Health*, *39*(1), 55-76.

Sklar, M., Ehrhart, M. G., & Aarons, G. A. (2021). COVID-related work changes, burnout, and turnover intentions in mental health providers: A moderated mediation analysis. *Psychiatric Rehabilitation Journal*, *44*(3), 219.

Southam-Gerow, M. A., Chapman, J. E., Martinez, R. G., McLeod, B. D., Hogue, A., Weisz, J. R., & Kendall, P. C. (2021). Are therapist adherence and competence related to clinical outcomes in cognitive-behavioral treatment for youth anxiety?. *Journal of Consulting and Clinical Psychology*, *89*(3), 188.

Southam-Gerow, M. A., & McLeod, B. D. (2013). Advances in applying treatment integrity research for dissemination and implementation science: Introduction to special issue. *Clinical Psychology: Science and Practice*, *20*(1), 1.

Substance Abuse and Mental Health Services Administration. (2011). *National registry of effective programs and practices*. Retrieved from http://nrepp.samhsa.gov/

Tabak, R. G., Khoong, E. C., Chambers, D. A., & Brownson, R. C. (2012). Bridging research and practice: models for dissemination and implementation research. *American Journal of Preventive Medicine*, *43*(3), 337-350.

Turner, C. W., Robbins, M. S., Rowlands, S., & Weaver, L. R. (2017). Summary of comparison between FFT-CW® and Usual Care sample from Administration for Children's Services. *Child Abuse & Neglect*, *69*, 85-95.

Turner, C.W., Robbins, M. S., Winokur Early, K., Blankenship, J. L., Weaver, L. R. (2019). Juvenile justice risk factors and Functional Family Therapy fidelity on Felony Recidivism. *Criminal Justice and Behavior, 46,* 697–717.

Underwood, L. A., & Washington, A. (2016). Mental illness and juvenile offenders. *International Journal of Environmental Research and Public Health, 13*(2), 228.

U.S. Department of Health and Human Services. (2019). *ACYF-CB-IM-19-04: Evaluation plan development tip sheet*. Retrieved from: https://www.acf.hhs.gov/sites/default/files/cb/im1904.pdf.

U.S. Department of Health and Human Services [HHS]. (2013). Report to Congress on the Nation's substance abuse and mental health workforce issues. Retrieved from http://store.samhsa.gov/ shin/content//PEP13-RTC-BHWORK/PEP13-RTC-BHWORK.pdf.

U.S. Department of Justice. 2020 Juvenile Justice Statistics. Retrieved from https://ojjdp.ojp.gov/publications/delinquency-cases-in-juvenile-court-2020.pdf

Virginia Department of Juvenile Justice (2017). *Transformation Plan 2017 Update*. Retrieved from https://www.djj.virginia.gov/documents/policy/data-research/publications/Transformation%20Plan%20Final.pdf

Virginia Department of Juvenile Justice (2021). *Transformation Plan 2017 Update*. Retrieved from https://www.djj.virginia.gov/documents/policy/data-research/publications/DJJ%20Transformation%20Plan%202021%20Update.pdf

Wandersman, A., Chien, V. H., & Katz, J. (2012). Toward an evidence-based system for innovation support for implementing innovations with quality: Tools, training, technical assistance, and quality assurance/quality improvement. *American Journal of Community Psychology*, *50*(3-4), 445-459.

Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 78*(2), 200–211.

Weiner, B. J., Belden, C. M., Bergmire, D. M., & Johnston, M. (2011). The meaning and measurement of implementation climate. *Implementation Science*, 6, 1-12.

Weekly Supervision Checklist. Unpublished measure. FFT, LLC.

Weersing, V. R., Jeffreys, M., Do, M. C. T., Schwartz, K. T., & Bolano, C. (2017). Evidence base update of psychosocial treatments for child and adolescent depression. *Journal of Clinical Child & Adolescent Psychology*, *46*(1), 11-43.

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

Weisz, J. R., Hawley, K. M., & Jensen Doss, A. (2004). Empirically tested psychotherapies for youth internalizing and externalizing problems and disorders. *Child and Adolescent Psychiatric Clinics*, *13*(4), 729-815.

Weisz, J. R., Kuppens, S., Ng, M. Y., Eckshtain, D., Ugueto, A. M., Vaughn-Coaxum, R., ... & Weersing, V. R. (2017). What five decades of research tells us about the effects of youth psychological therapy: a multilevel meta-analysis and implications for science and practice. *American Psychologist*, *72*(2), 79.

Willging, C. E., Aarons, G. A., Trott, E. M., Green, A. E., Finn, N., Ehrhart, M. G., & Hecht, D. B. (2016). Contracting and procurement for evidence-based interventions in public-sector human services: A case study. *Administration and Policy in Mental Health and Mental Health Services Research*, *43*(5), 675-692.

Wiltsey Stirman, S., Baumann, A. A., & Miller, C. J. (2019). The FRAME: an expanded framework for reporting adaptations and modifications to evidence-based interventions. *Implementation Science*, *14*, 1-10.

Wiltsey Stirman, S., Gutner, C., Crits-Christoph, P., Edmunds, J., Evans, A. C., & Beidas, R. S. (2015). Relationships between clinician-level attributes and fidelity-consistent and fidelity-inconsistent modifications to an evidence-based psychotherapy. *Implementation Science*, *10*(1), 1-10.

Wiltsey Stirman, S., Gamarra, J. M., Bartlett, B. A., Calloway, A., & Gutner, C. A. (2017). Empirical examinations of modifications and adaptations to evidence-based psychotherapies: Methodologies, impact, and future directions. *Clinical Psychology: Science and Practice*, *24*(4), 396.

Wiltsey Stirman, S., Gutner, C., Crits-Christoph, P., Edmunds, J., Evans, A. C., & Beidas, R. S. (2015). Relationships between clinician-level attributes and fidelity-consistent and fidelity-inconsistent modifications to an evidence-based psychotherapy. *Implementation Science*, *10*(1), 1-10.

Wiltsey Stirman, S., Kimberly, J., Cook, N., Calloway, A., Castro, F., & Charns, M. (2012). The sustainability of new programs and innovations: a review of the empirical literature and recommendations for future research. *Implementation Science*, *7*(1), 1-19.

Woltmann, E. M., Whitley, R., McHugo, G. J., Brunette, M., Torrey, W. C., Coots, L., ... & Drake, R. E. (2008). The role of staff turnover in the implementation of evidence-based practices in mental health care. *Psychiatric Services*, 59(7), 732-737.

Wright, B., González, I., Chen, M., Aarons, G. A., Hunter, S. B., Godley, M. D., ... & Dopp, A. R. (2024). Multi-level alignment processes in the sustainment of a youth substance use treatment model following a federal implementation initiative: A mixed method study. *Journal of Substance Use and Addiction Treatment*.

Yang, Y., & Hayes, J. A. (2020). Causes and consequences of burnout among mental health professionals: A practice-oriented review of recent empirical literature. *Psychotherapy, 57*(3), 426–436.

Zajac, K., Sheidow, A. J., & Davis, M. (2015). Juvenile justice, mental health, and the transition to adulthood: A review of service system involvement and unmet needs in the US. *Children and Youth Services Review*, *56*, 139-148.

Zazzali, J. L., Sherbourne, C., Hoagwood, K. E., Greene, D., Bigley, M. F., & Sexton, T. L. (2008). The adoption and implementation of an evidence based practice in child and family mental health services organizations: A pilot study of functional family therapy in New York State. *Administration and Policy in Mental Health and Mental Health Services Research*, *35*, 38-49.

Zima, B. T., Hurlburt, M. S., Knapp, P., Ladd, H., Tang, L., Duan, N., ... & Wells, K. B. (2005). Quality of publicly-funded outpatient specialty mental health care for common childhood psychiatric disorders in California. *Journal of the American Academy of Child & Adolescent Psychiatry*, *44*(2), 130-144.

Zullig, L. L., Deschodt, M., & De Geest, S. (2019). Embracing implementation science: A paradigm shift for nursing research. *Journal of Nursing Scholarship*, *52*(1), 3-5.

Table 1. *Total Sample Descriptives and Completion of FFT*

	Total (*n*)	**% of Total**	**Completed FFT (*n*)**	**Completed FFT (%)**
Race/Ethnicity				
Black	*663*	*38.6%*	*396*	*59.7%*
Latinx, Hispanic, Chicano	*205*	*11.9%*	*142*	*69.3%*
White	*630*	*36.6%*	*401*	*63.7%*
Other Race/Ethnicity	*221*	*12.86%*	*139*	*62.9%*
Biracial	114	6.6%	73	64%
Asian	9	0.5%	6	66.7%
Middle Eastern/North African	5	0.3%	5	100%
Native Hawaiian/Pacific Islander	5	0.3%	3	60%
Other	27	1.6%	17	63%
Unknown	61	3.5%	35	57.4%
Gender				
Male	1230	71.6%	775	63%
Female	475	27.6%	294	61.9%
Other Gender	14	0.8%	9	64.3%
Age				
Younger than 11	11	0.6%	8	72.7%
11-18	1494	86.9%	937	62.7%
Older than 18	199	11.6%	132	66.3%
Referral Reason				
Youth Behavior Reason	*1079*	*62.8%*	*667*	*62.7%*
Delinquent behavior	998	58.1%	636	63.7%
Gang involvement	2	0.1%	1	50%
Runaway Behavior	20	1.2%	9	45%
School Truancy/Behavior	63	3.7%	34	54%

Family Reasons	*413*	*24.0%*	*257*	*62.2%*
Family Reunification	83	4.8%	44	53%
Family Separation	12	0.7%	6	50%
Family Substance Abuse/Use	40	2.3%	25	62.5%
Family Violence	135	8.0%	91	65.9%
Youth/Caregiver Conflict	140	8.1%	92	65%
Mental Health Problem	*124*	*7.2%*	*73*	*58.9%*
Other Reasons	*103*	*6.0%*	*71*	*68.9%*
Other	30	1.7%	22	73.3%
Unknown	69	4.0%	46	66.7%
Referral Source				
Child Welfare	*135*	*7.9%*	*79*	*58.5%*
Mental Health	*158*	*9.2%*	*86*	*54.4%*
Juvenile Justice	*1166*	*67.8%*	*759*	*65.1%*
Parole	114	6.6%	62	54.4%
Probation	1043	60.7%	691	66.3%
Diversion	9	0.5%	6	66.7%
Other Sources	*260*	*15.1*	*154*	*59.2%*
Internal Agency Referral	58	3.4%	31	53.4%
Self-Referral	15	0.9%	4	26.7%
School Referral	51	3.0%	33	64.7%
Other	67	3.9%	40	59.7%
Unknown	69	4.0%	46	66.7%
Total	1719		1078	62.7%

Table 2. *Status Outcomes at Case Closure*

	Remain in community (n)	**Remain in community (%)**	**In school/ work (n)**	**In school/ work (%)**	**No new law violations (n)**	**No new law violations (%)**
Race/Ethnicity						
Black	*499*	*79.6%*	*525*	*85.8%*	*428*	*68.6%*
Latinx, Hispanic, Chicano	*166*	*85.1%*	*161*	*88%*	*147*	*77%*
White	*488*	*83.6%*	*498*	*89.1%*	*460*	*81.4%*
Other Race/Ethnicity	*179*	*84.4%*	*178*	*89.0%*	*169*	*80.9%*
Biracial	92	84.4%	90	88.2%	86	80.4%
Asian	9	100%	8	88.9%	8	88.9%
Middle Eastern/ North African	5	100%	5	100%	5	100%
Native Hawaiian/Pacific Islander	4	75%	4	100%	3	75%
Other	21	80.8%	22	84.6%	16	64%
Unknown	49	83.1%	49	90.7%	51	86.4%
Gender						
Male	938	80.8%	962	86.6%	834	73.1%
Female	308	85.8%	389	90.5%	356	82%
Other Gender	14	100%	11	84.6%	14	100%
Age						
Younger than 11	10	90.9%	10	100%	9	81.8%
11-18	1148	81.4%	1187	87.2%	1042	75.2%
Older than 18	170	89.5%	160	90.9%	148	78.7%
Referral Reason						
Youth Behavior	*830*	*80.9%*	*856*	*86.8%*	*737*	*73.3%*
Delinquent behavior	766	80.5%	796	87.2%	678	73%
Gang involvement	1	50%	2	100%	1	50%

Runaway Behavior	13	81.3%	14	82.4%	10	55.6%
School Truancy/ Behavior	53	86.9%	48	82.8%	49	81.7%
Family Reasons	*320*	*83.6%*	*328*	*89.9%*	*297*	*78.2%*
Family Reunification	62	81.6%	68	91.9%	64	84.2%
Family Separation	9	90%	9	90.0%	10	100%
Family Substance Abuse/use	27	73%	25	75.8%	18	48.6%
Family Violence	107	83.6%	107	90.7%	101	78.9%
Youth/Caregiver Conflict	115	87.1%	119	91.5%	104	80.6%
Mental Health Problem	*99*	*90.8%*	*93*	*90.3%*	*96*	*90.6%*
Other Reasons	*83*	*83.0%*	*85*	*85.0%*	*74*	*75.5%*
Other	28	93.3%	27	96.4%	24	85.7%
Unknown	52	78.8%	54	79.4%	49	74.2%
Referral Source						
Child Welfare	*104*	*82.5%*	*107*	*91.5%*	*150*	*84.7%*
Mental Health	*122*	*88.4%*	*124*	*93.2%*	*119*	*87.5%*
Juvenile Justice	*894*	*80.5%*	*933*	*86.9%*	*784*	*71.5%*
Parole	72	69.2%	79	76%	59	56.7%
Probation	813	81.5%	846	87.9%	717	72.8%
Diversion	9	100%	8	100%	8	100%
Other Sources	*212*	*87.2%*	*198*	*86.1%*	*196*	*84.5%*
Internal Agency Referral	47	94%	42	93.3%	44	89.9%
Self-Referral	11	84.6%	10	83.3%	11	100%
School Referral	46	93.9%	36	78.3%	42	91.3%
Other	56	86.2%	56	94.9%	50	83.3%
Unknown	52	78.8%	54	79.4%	49	74.2%
Total	1332	77.5%	1362	79.2%	1204	75.8%

Table 3. Race/Ethnicity Group by Referral Source

	Juvenile Justice	Child Welfare	Mental Health	Other Referral Source
Black	*83.6%*	*3.5%*	*3.6%*	*9.4%*
Latinx, Hispanic, Chicano	*68.8%*	*2.9%*	*9.8%*	*18.5%*
White	*58.7%*	*11.3%*	*12.9%*	*17.1%*
Other Race/Ethnicity	*45.7%*	*15.8%*	*14.9%*	*23.5%*

Figure 1. *Team Size*

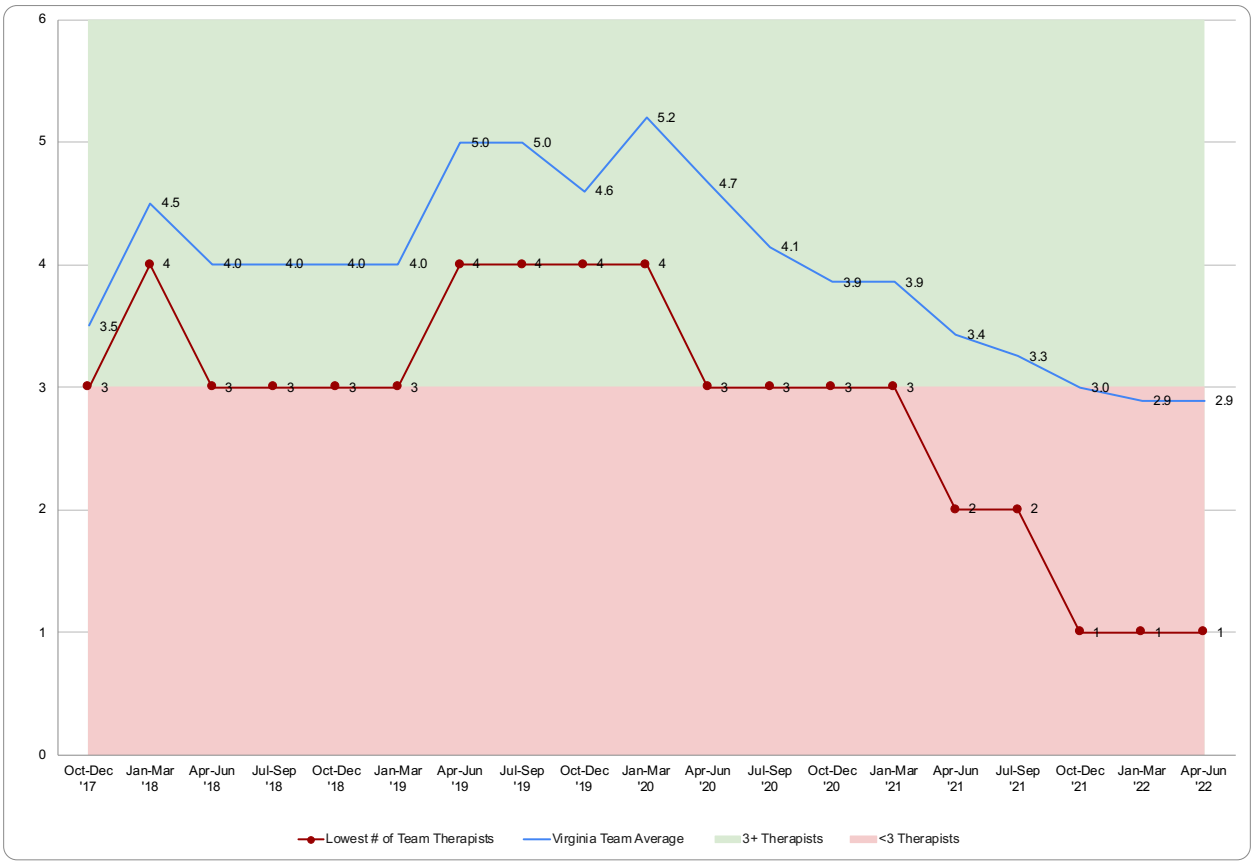


Figure 2. *Average Caseload*

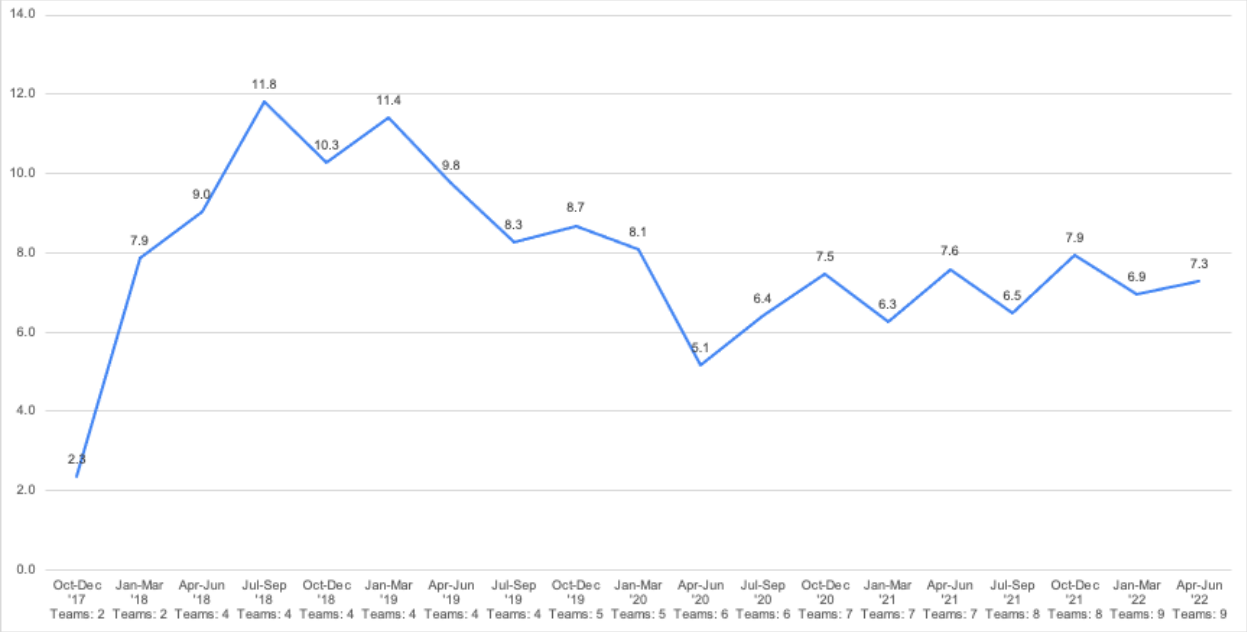


Figure 3. *Percentage of Teams Meeting Caseload Range Per Therapist Benchmark*

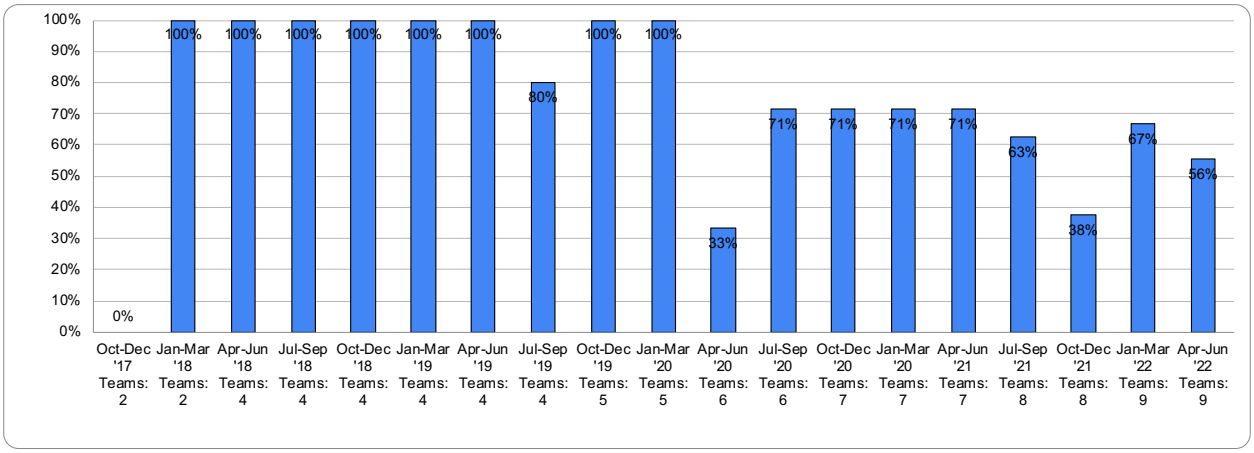


Figure 4. *Team-Level Average Fidelity*

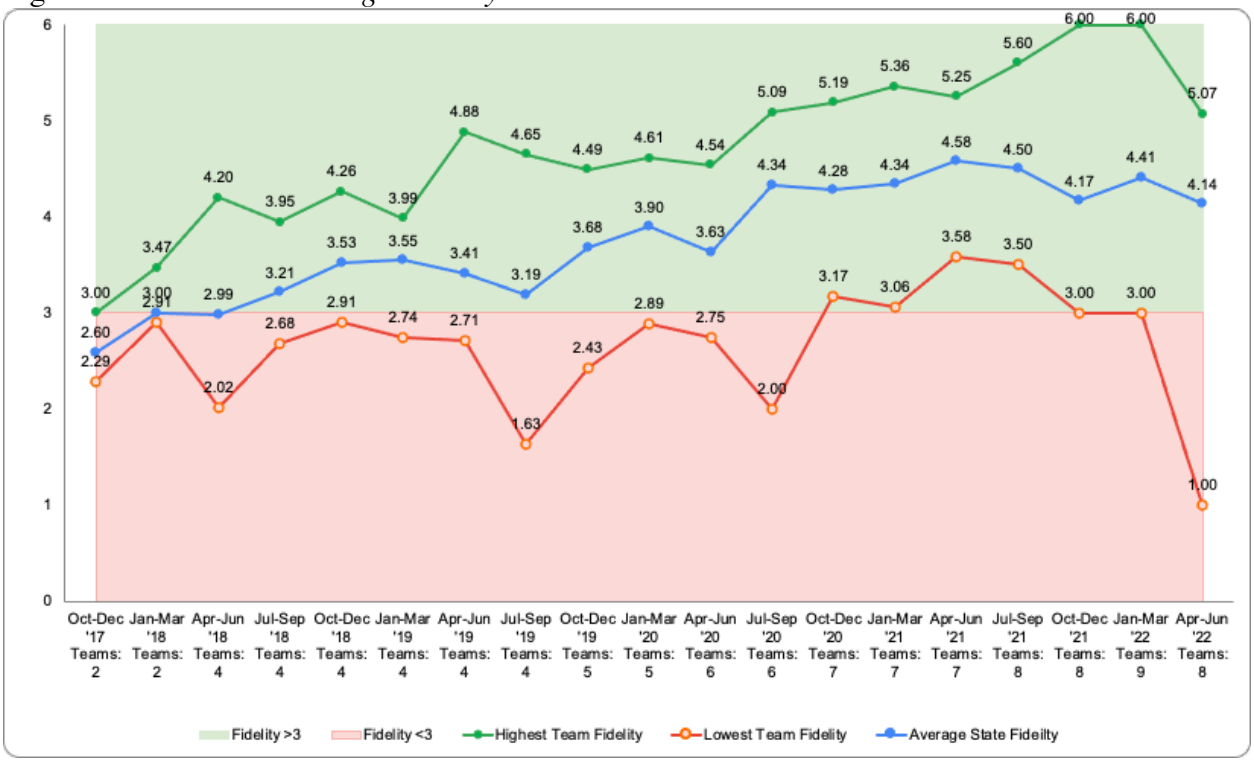


Figure 5. *Case Completion Rate*

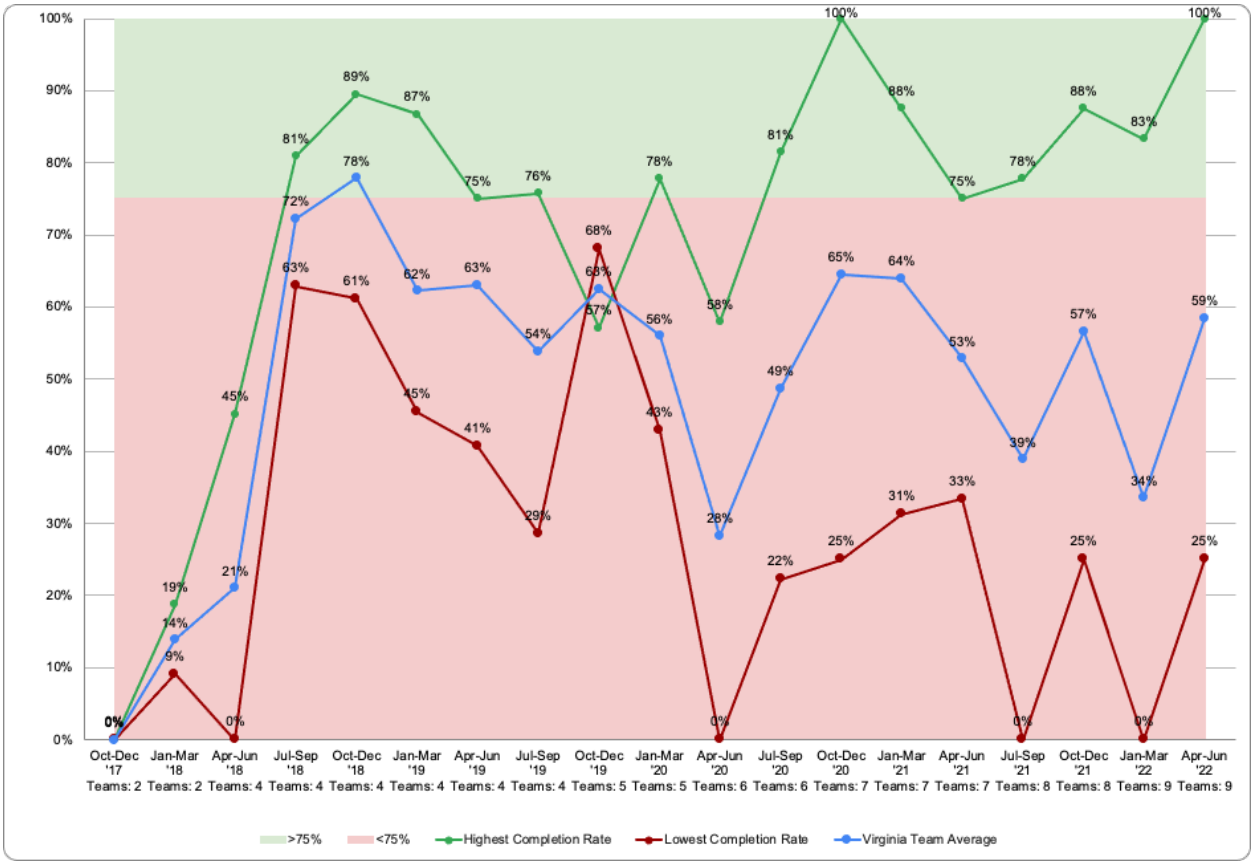


Figure 6. *Youth Remaining in the Community at Completion*

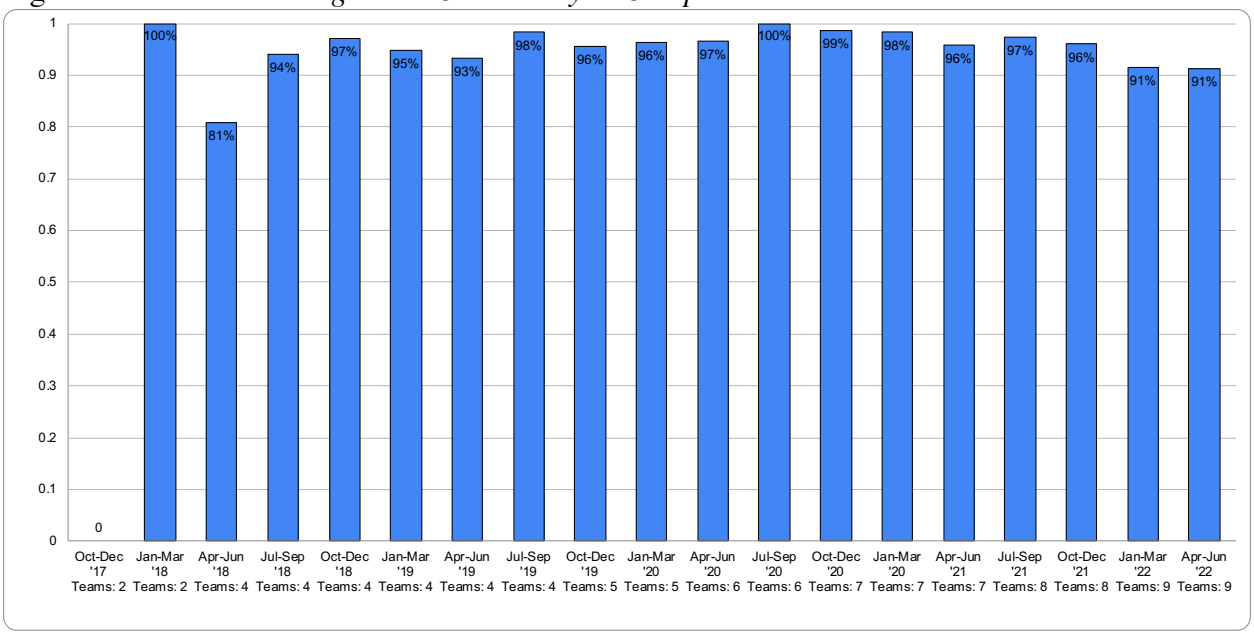


Figure 7. *Youth Attending School or Work at Completion*

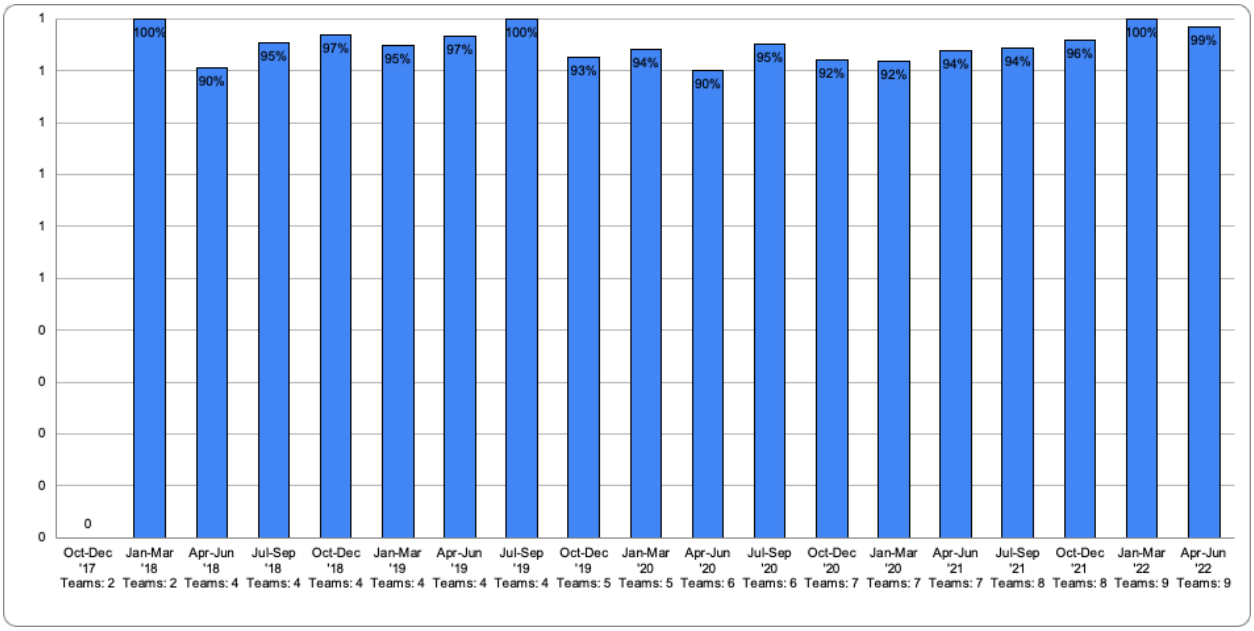
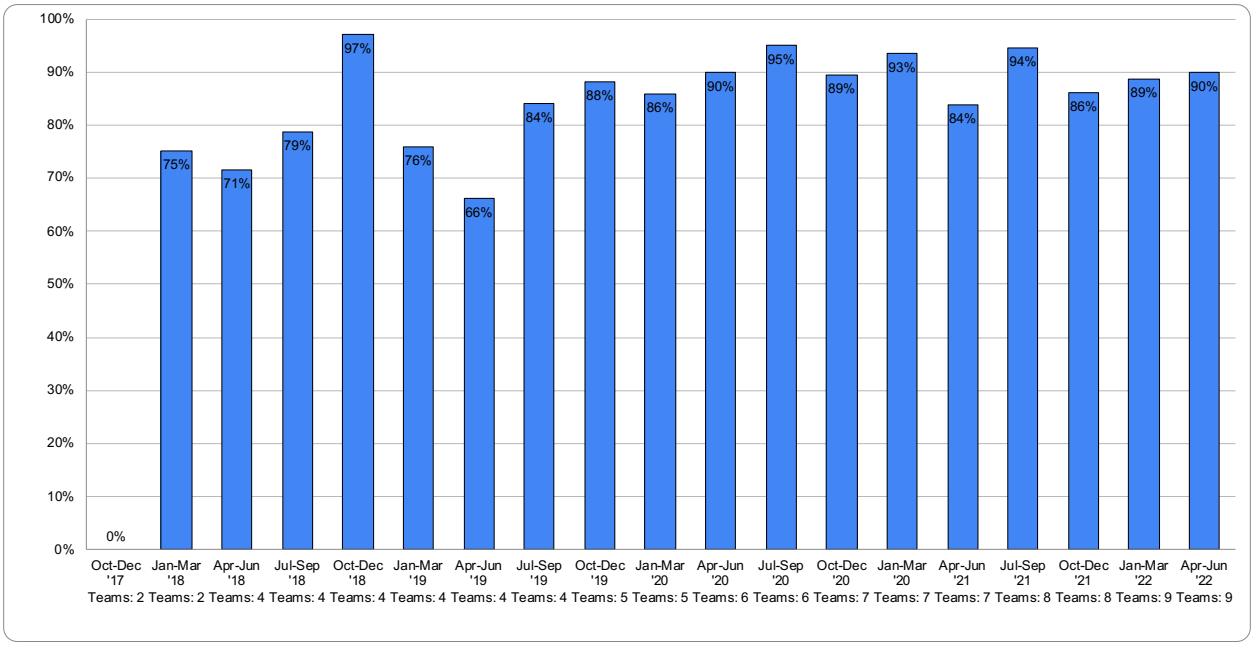


Figure 8. *Youth Having No New Law Violations at Completion*



Vita

Natalie Finn was born on September 30, 1991, in New York. Natalie received her Bachelor of Science in Human Development from Cornell University in 2013 and subsequently worked for the University of California, San Diego before starting her graduate studies in the Clinical Psychology Doctoral Program at Virginia Commonwealth University in 2017. She earned her Master of Science in Psychology from Virginia Commonwealth University in 2019. Natalie began her pre-doctoral clinical internship at Denver Health Medical Center in 2023.