

Fidelity Management: A Low-Cost Alternative to Proprietary Evidence-based Programs¹

Pamela Meadowcroft, Ph.D., Maria Townsend, Ph.D., Art Maxwell, Ph.D., and Katy Collins²

FINAL REPORT TO WESLEY SPECTRUM

July 5, 2011

Introduction

Although the movement toward evidence-based practice in the human services field has been underway for years, progress has been limited and the path toward widespread implementation of evidence-based practices is unclear. Many service providers, funders and public agencies are unsure how to proceed.

The motivation for the field's move toward evidence-based practice (EBP) is a combination of fiscal pressures and a quest for more accountability toward the intended beneficiaries of human services. State and Federal policies increasingly require the use of evidence-based practices and in some instances adoption of whole evidence-based packaged programs. The Obama administration has said that only those services that can prove their effectiveness will be funded in the future. Given these pressures, the need to provide credible evidence as to the effectiveness of human services programs is clear. EBPs hold the promise of enabling providers to make the best use of scarce funds while providing the most effective available services to those in need.

So it is no surprise that the fastest growing service providers are those that can demonstrate measurable results. In a recent national study of youth serving agencies (Campbell and Menezes, 2005) a full 32 percent of the 100 fastest-growers reported using evidence-based practices proven to deliver positive outcomes in a randomized control trial, and many others employ promising interventions tested with less rigorous forms of evaluation. These financially healthy youth-serving nonprofits report that "their organizations were deeply engaged in tracking results, with more than half employing full-time performance assessment staff." Measuring results – being able to provide evidence of effectiveness - is essential to positive participant outcomes, continuous quality improvement and, ultimately, successful human service operations.

¹ We wish to thank the leadership and staff of Wesley Spectrum Services who have been strong partners in building this new process for creating low cost evidence-based program models. Additionally we thank the Allegheny County Juvenile Justice and Department of Human Services for funding the initial research on the Wesley Spectrum In Home Services "model."

² Pamela Meadowcroft, President, Maria Townsend and Art Maxwell, Senior Associates, and Katy Collins, Associate with Meadowcroft and Associates. Meadowcroft is also a faculty associate with the Graduate School of Public Health Center for Evaluation Science, University of Pittsburgh, Townsend is Adjunct Faculty, Maxwell is a Visiting Assistant Professor, and Katy Collins is a Doctoral Candidate at the Graduate School of Public and International Affairs, University of Pittsburgh.

But in some circles, 'Evidence-Based Practice' has become conflated with one particular approach. Pre-packaged, brand-name evidence-based programs have emerged that are available for purchase and implementation by states, local communities, and/or providers on a turn-key basis. The effectiveness of these brand-name programs is typically supported by rigorous research using randomized clinical trials (RCTs) conducted at one or more research sites and subsequently "certified" by third-party national groups (e.g., Blueprints, OJJDP, 2010). In that sense, such brand-name programs are truly 'evidence-based.'

Evidence-Based Practice Defined:
Approaches to prevention or treatment that are based in theory and have undergone scientific evaluation. "Evidence-based" stands in contrast to approaches that are based on tradition, convention, belief, or anecdotal evidence. [Substance Abuse and Mental Health Services Administration (SAMHSA) National Registry of Evidence-Based Programs and Practices]

However, it is clear from the implementation research literature that such programs are not a panacea for the industry as it seeks to make EBPs more widespread.

Benefits - and limitations - of brand-name programs

It is important for providers and funders to recognize not only the benefits, but also the limitations, that are inherent in brand-name programs. And given those limitations, it is also important to understand the alternatives available for achieving evidence-based programs in our communities.

Benefits

Brand-name programs, supported by the highest standards of rigorous science for determining cause-effect relationships - randomized clinical trials or RCTs - show clear return on investment. The outcomes for these brand-name programs prove that, for particular clients under particular circumstances, they are more cost-effective than 'traditional' services or no service. Recidivism is lower, rates of high school graduation higher, criminal activities are lower or curtailed, and family cohesion is increased. If one of these programs is implemented properly, it is likely that the positive outcomes the program is designed to produce will occur. In this way, brand-name programs enable limited public resources to be used wisely. Another benefit of adopting a brand-name program is efficiency. Instead of having to "reinvent the wheel," agencies can select from the growing number of programs that are known to be well designed and have undergone rigorous evaluation.

Examples of brand-name programs include Multi-systemic Therapy, Functional Family Therapy, and Multidimensional Treatment Foster Care. Brand-name programs typically offer for purchase or license a comprehensive package, including treatment protocols, training for staff practitioners, data collection and service infrastructure specifications, and ongoing support. In instances where no existing program is available, brand-name programs represent an effective option for implementing evidence-based programs to address community needs.

Limitations

The benefits of brand-name programs described above are only part of the picture. Along with these benefits come several important limitations.

- **Initial and ongoing costs:** First, there are the initial costs of implementing the program: start-up fees, staff and supervisor training, travel expense, and implementation of whatever service infrastructure and data collection system is required by the program vendor. Second, providers adopting brand-name programs typically pay ongoing costs; annual program support fees per site and annual team and site license fees paid to the program vendor, retraining costs required to maintain staff competency, and training expenses associated with staff turnover. Providers considering implementing brand-name programs must consider both the initial and the ongoing costs. Funders and policy-makers should do a full cost-comparison considering this limitation as well as the “scalability” one below.
- **Limited external validity** (generalization to new settings and varied populations): The research design – randomized clinical trials (RCTs) - used to develop the evidence in support of many brand-name programs has limited external validity; in other words, this particular research design is excellent for determining if a highly specified set of practices provided under highly specified circumstances to highly specified consumers can be causally connected to the changes the consumers experience. It is in the very nature of the research method used – RCTs – that while they can establish very strong proof that a treatment is effective, that strong proof is generalizable only to the specific circumstances of the study design. In fact, many experts in the field of evaluation claim that the emphasis on RCTs as the gold-standard for evidence-based programs is unwarranted and inappropriate: “Today more and more practitioners, decision makers, and consumers find that traditional scientific evaluation results tend not to be useful to the everyday issues about which [real-world practice] is concerned; i.e., evidence-based interventions are difficult to implement in the real world (Chen and Garbe, 2011)” largely because of a lack of attention to generalizability to different settings, circumstances, and persons.
- **Difficulty maintaining fidelity to the model:** Related to problems with external validity (difficulties in generalizing to new locations/contexts/times) is the difficulty in implementing model programs with high levels of fidelity. Fidelity refers to how faithfully a real-world implementation adheres to the program model – its specifications for the intervention to succeed. Research suggests that brand-name model programs are often not implemented with high model fidelity; thus the positive outcomes demonstrated during the randomized clinical-trial may not be achievable in alternate community settings (Fagan, 2008; Gottfredson, 2002; Hallfors, 2002; Lispey, 2010).
- **Scalability (economies of scale):** The ongoing costs, limited service populations and the challenges of model fidelity are combining to pose serious challenges to providers

attempting to scale up brand-name programs. Attempts to manage the above problems frequently entail limiting the volume of consumers served. Achieving target costs for operating a brand-name program appears to be increasingly difficult when the anticipated volume of consumers to be served is limited by operational constraints. Such limitations can include brand-name program issues such as limiting the types of consumers that can be served in order to maintain “model fidelity,” certification criteria that limit growth until certain benchmarks of staff competence are achieved (often taking years), and even requiring pre-approval of referred consumers by owners of the brand-name (to assure model fidelity). Delays and limitations to serving consumers when the service is needed lead to reduced confidence by referral sources, thus leading to fewer and fewer referrals. Providers can find themselves with an expensive program to operate in a diminished referral environment despite the growing need of that community’s children and families for a set of services very much like the evidence-based one.

Although ‘evidence-based’ is sometimes conflated with ‘randomized clinical trials,’ it is important that the human services field bears in mind both the limitations associated with RCTs and the fact that recognized authorities see randomized trials as but one research method for demonstrating effectiveness. The Substance Abuse and Mental Health Services Administration registry of EBPs contains programs that are based on not only experimental (i.e. RCT) research designs, but other methods for producing credible evidence for program models.

Where do we go from here?

Despite the hopes that the brand-name models would produce higher quality services for more participants and an improved return-on-investment, implementation of these models has not met the expectations. In fact, the new field of “implementation science” emerged in large part because of the difficulties of implementing proven programs and interventions in many fields of health and human services. Increasingly, implementation research is teaching us ways of closing “the gap between science and service by improving the science and practice of implementation in relation to evidence-based programs and practices (Fixsen and Blasé, 2011).” We are now learning that high quality (i.e., positive outcomes) for the most number of consumers requires a different strategy than only replicating brand-named models validated via randomized clinical trials (RCT).

Providers Need to Deliver Quality with Attention to Cost

The experience to date has shown us how to use evidence-based practices to achieve better quality and effectiveness by delivering better outcomes for program participants. But an approach to EBP that cannot serve sufficient numbers of those in need with acceptable ongoing costs is simply unsustainable.

Consider how prevailing practices across the human services field have evolved in recent decades in terms of scale and quality/effectiveness. The lower left-hand quadrant in Figure 1 represents human services of the past: providers served small numbers of families and

children with minimal attention to quality or measuring effectiveness. Over time, greater numbers of consumers had access to services, but there was still little attention to measured effectiveness. That state of affairs (upper left quadrant) represents most of human services today. The right lower quadrant represents the recent emphasis on measuring effectiveness, with evidence-based programs as the Gold Standard. But unless evidence-based programs are readily accessible to the numbers of families and youth needing these services, they remain only a small part of the future: the upper right-hand quadrant represents the future in which community based programs are all able to achieve a higher level of “evidence-based practice” credibility.

Figure 1: Evolution of Human Services Programs: What is “value?”

| | | | |
|--|--|---------------------------|---|
| NUMBERS OF KIDS AND FAMILIES SERVED and AFFORDABILITY  | QUALITY  | | |
| | | Lower Quality/no outcomes | Higher Quality/proven effectiveness |
| | Greater access, higher volume | CURRENT SYSTEM | THE FUTURE: Evidence-based locally-developed models within context of CQI |
| | Lower access, smaller volume | THE PAST | THE CATALYST: Evidence-based (RCT) brand-models |

Programs provide the greatest value when they demonstrate effectiveness while at the same time serving the numbers of participants required to sustain provider operations. The RCT brand-name programs may provide the best “evidence-based” service, but despite ten years of research and effort, the juvenile justice field estimates that only 5% of eligible youth are being served in such programs.

To increase access to proven-programs for many more consumers requires that we move existing programs into the upper right-hand quadrant. We need to convert well-established programs that are serving large numbers of consumers in a community to evidence-based programs without having to conduct expensive RCTs.

An Alternative Path to Evidence-Based Practice

There is a lower-cost/high value alternative to brand-name programs. Providers of existing programs can embed in their programs a continuous improvement process that demonstrates the link between program structure and practices, costs and outcomes.

Research into the effectiveness of human services programs shows that both locally-developed and brand-name programs exhibit variation in the outcomes they produce in the real world. One large-scale “meta” study of juvenile justice programs found that whether a particular program was locally-developed or one of the brand-name programs was not itself a predictor of effectiveness. Rather, three factors- the risk profile of program participants, the particular types of interventions implemented in a given program, and the fidelity with which a given program is implemented – are the most important explanatory factors in program effectiveness (Lipsey, 2009). The finding that fidelity – the faithfulness with which a research-based intervention is implemented in the real-world – is an important factor has practical implications that should be considered by providers and funders in search of EBPs. Mounting evidence suggests that the next evolution in the increasingly accountable youth services industry will be evidence-based guidance for increasing the effectiveness of locally-developed programs without replacing them with brand-name, EBP model programs (Lipsey, Howell, Kelly, Chapman, & Carver, 2010).

Strengthening existing programs by focusing on model fidelity in the context of continuous quality improvement has several important advantages:

- Builds on existing community values,
- Adapts researched practices to existing community populations and needs,
- Provides the mechanism for serving more children and families effectively,
- Gives programs the low-cost, program-owned tools and metrics for continually monitoring their effectiveness.
- Gives individual providers and the youth service “system” as a whole the means by which innovative new practices can be pursued and incorporated into ongoing practice.

A new tool for providers – Evidence-Based Practices through Fidelity Management

In a recent study of a locally-developed youth-serving program (Wesley Spectrum In Home Services) Meadowcroft & Associates was able to demonstrate the role of model fidelity in influencing participant outcomes. Based on current research into what works in youth services, Meadowcroft, with collaborators at Wesley Spectrum, developed a set of steps that can be applied within many human service agencies as a recipe for demonstrating that their program has high fidelity to evidence-based practices, produces highly effective results, and accomplishes both at a lower cost than the brand-name models. The result is a management tool that can enable providers of existing programs to measure model fidelity on an ongoing basis as part of a continuous quality improvement program focused on not only documenting but also improving participant outcomes.

We call this approach to evidence-based practice ‘Fidelity Management.’ Fidelity Management enables locally-developed programs, which are already operating within the community, to demonstrate comparable outcomes to the certified, highly researched and

more expensive brand-name programs.

Essential steps for demonstrating Evidence-Based Practices through Fidelity Management

1. **Define the Program:** determine the extent to which the existing, essential program practices (based on a two-level detailed Logic Model) have research support and adjust program practices if warranted;
2. **Track Model Fidelity:** develop and use program-specific tools for tracking fidelity to the program practices;
3. **Monitor Outcomes:** measure relevant (defined by the existing research literature) program and client outcomes;
4. **Validate the Locally-Developed Program Model:** correlate the relationship between fidelity to the program model and its outcomes to validate the model as “evidence-based;”
5. **Build-in CQI:** continuously use the fidelity scores of direct service providers and outcome monitoring to improve the program results.

The bottom line for providers: it is very likely that existing programs already contain research-proven practices that, when consistently used, will lead to positive and significant outcomes. Our work suggests that it is possible to build on the mission-advancing services that are the life-changing work of an organization and elevate the public confidence in the program’s effectiveness. There is a cost effective methodology to better document, maintain fidelity and measure the impact of locally-developed programs.

Case Study: Wesley Spectrum Services

Wesley Spectrum Services, a nonprofit located in Southwestern PA, with experts from the fields of evidence-based practice, spearheaded the development of steps for demonstrating their program uses Evidence-Based Practices and has the capacity for Fidelity Management. Partnering with Pennsylvania’s Allegheny County Juvenile Probation and Children and Youth Services, this initiative enabled the Wesley Spectrum In-Home Program to:

- Identify a set of essential program practices that were supported by credible research (e.g., “Concentrate on changing behavior and improving pro-social skills by providing goal oriented treatment...”);
- Identify a set of process measures (key activities, dosages for service, criteria for staff hiring and training) that were then incorporated into ongoing supervision tracking tools;
- Show that the clients of the family-workers with the higher levels of “fidelity” to the In-Home Program “model” produced statistically better youth/family outcomes than workers with lower fidelity scores;
- Show that the outcomes for the families of the “higher-fidelity” workers were

statistically comparable to (or exceeded that of) the brand-name programs;

- Produce comparable or better results than brand-name programs for about 1/3 the cost of a comparable brand-name program;
- Use the fidelity scores for ongoing staff training and supervision as an essential component to their CQI process.

For less cost than a typical randomized-clinical-trial, the Wesley Spectrum In-Home Program was able to show that when their model is implemented with high levels of fidelity, the program produces statistically comparable or better results than brand-name models, including...

- Positive outcomes: low recidivism, higher family cohesion, and improved school performance;
- Evidence that the practices of their staff are supported by existing rigorous research;
- Lower program costs that don't require any high start-up training fees common among the brand-name models, nor the ongoing annual costs of recertification, supervision, online data services;
- Staff with a higher degree of pride and "ownership" of the model they've created and, therefore, higher morale and less turnover (a cost-reducer);
- The data methods and mechanics to use results for ongoing continuous program improvement including adapting to changing community needs;
- Services for more and varied populations while sustaining positive outcomes.

In other words, Wesley Spectrum's Evidence-Based Practices with Fidelity Management produces higher service volume, higher quality services, more highly committed staff, the ability to continuously evolve as the needs of youths and family changes, and confidence in setting program priorities that will enhance positive outcomes for those they serve.

The Wesley Spectrum Study

Study Summary

During the first half of 2011, Wesley Spectrum In-Home Services (WSIH) implemented their program-specific Fidelity Management tools (Therapist Checklist and Supervisor Checklist), which monitored the level of fidelity to the program model as well as tracking client outcomes. Eighty-nine cases were used to examine the degree to which:

- the program services showed fidelity to the WSIH model for each case,
- high fidelity or low fidelity to the model affected outcomes for youth and families,
- the WSIH model compares in terms of service outcomes to the brand-name, evidence-based model MST.

Generally, the higher model fidelity cases produced significantly better outcomes. Cases in which model fidelity was higher showed significantly better improvement (pre-post assessment) compared to those cases with lower fidelity for the following outcomes:

- child well-being
- child behavior
- overall improvement in environment
- supervision and discipline
- overall family interactions
- child's relationship with parents
- school performance
- overall family safety

In comparing the WSIH model with MST, WSIH produced either comparable or statistically better results for the standard MST outcomes: recidivism, in school/working, living at home. Differences in populations of youths served are not likely the reason for these significantly better outcomes: WSIH served significant greater proportion of youths who are African-American and more males. WSIH produced a higher percentage of youths and their families that completed the program; only 49% of the youths and families served in the local Allegheny County program completed the service versus 78% of those in the WSIH program. Finally, cost data shows the WSIH model to be about 1/3 the cost of the MST program in Allegheny County.

Methodology

Eighty-nine closed cases were examined that had both fidelity and outcome data. Two thirds of the cases were juvenile justice cases, 24% were child-welfare cases from Westmoreland County and 10% were child-welfare cases from Allegheny County. Over half of the cases (51%) involved females. Race data were missing for 31 cases (36%), 26 cases were identified as African-American (29%) and 27 were identified as Caucasian (30%). The average length of treatment

was 140.5 days with a range of 35 to 455 days, a median of 104 days and mode of 90 days. The majority of the cases ended:

- In a planned discharge (78%),
- With the child/youth in their family (81%)
- With the child/youth in school (84%)

Results

In order to determine the relationship between model fidelity and child and family outcomes, a score for model fidelity was created for each case based on the sum of fidelity items on the Therapist Checklist from intake and discharge plus the average monthly score of fidelity during treatment.

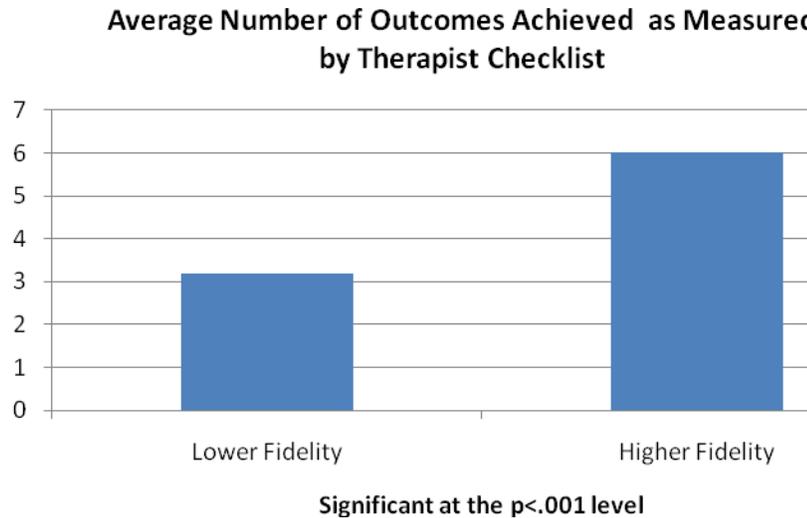
An examination of the Overall Fidelity scores across treatment found 12 cases of the 89 total cases or 13.5% that were one or more standard deviations below the mean representing the lowest fidelity to the model and 12 cases (13.5%) that were one or more standard deviations above the mean representing the highest fidelity to the model. In order to have enough cases to detect significant differences, the cases were split around the Overall Fidelity Scores mean into two groups, referred to as Lower vs. Higher Fidelity. See table below for descriptive statistics for the Overall Fidelity Scores by groups.

| | Mean | Median | Standard Deviation | Minimum | Maximum |
|-----------------------|------|--------|--------------------|---------|---------|
| All cases (n=89) | 23.2 | 24 | 5.9 | 5 | 32.2 |
| Lower Fidelity (n=41) | 18.6 | 21 | 5.4 | 5 | 23.5 |
| Higher Fidelity(n=48) | 27.2 | 26.7 | 2.5 | 24 | 32.2 |

There was no significant difference between these two groups based on gender but there was based on case type. Juvenile Justice cases represented a higher percentage of all the cases in the Lower Fidelity group compared to the Higher Fidelity group (80.5% vs. 54.2%, respectively, $p < .05$).

Because there were a number of outcomes possible during treatment and at discharge, we summed all of these to see if the total achieved differed depending on the level of model fidelity. Outcomes monitored on the Therapist Checklist include: planned discharge, child is with family or living independently, development of community supports, number of issues addressed through referrals and for juvenile cases, completion of court orders such as restitutions and skill building classes. The overall measure of model fidelity across intake, monthly treatment and discharge was positively related to number of outcomes achieved as

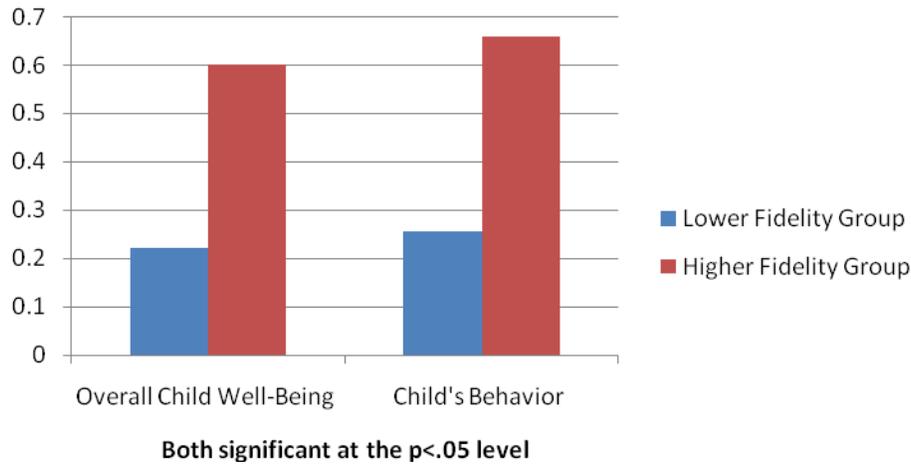
measured by the Therapist Checklist ($R=.444$, $p<.01$). The Higher Fidelity group had on average achieved 6 outcomes compared to 3.2 outcomes for the Lower Fidelity group ($p<.001$).³



All cases had at intake a completed assessment using the North Carolina Family Assessment Scale (NCFAS). This scale is considered one of the most valid tools for assessing family needs and for measuring change (change-scores) on these needs from intake to discharge. Comparison of the lower and higher fidelity groups on the change-scores of the relevant subscales (those that reflected the desired program outcomes) showed the Higher Fidelity group had significantly larger increases in their NCFAS-G scores for child well-being and child behavior from intake to discharge compared to the Lower Fidelity group.

³ Though there were no significant differences between the Lower and Higher Fidelity group for number of closed cases with child living with family or planned discharge, the fact that the Higher Fidelity group achieved more program outcomes (development of community supports, issues addressed through referral and completion of court orders) suggests that this group may be able to sustain their outcomes for longer periods of time. A follow-up survey to track child's living arrangements and recidivism rates will be completed to see if this is indeed the trend.

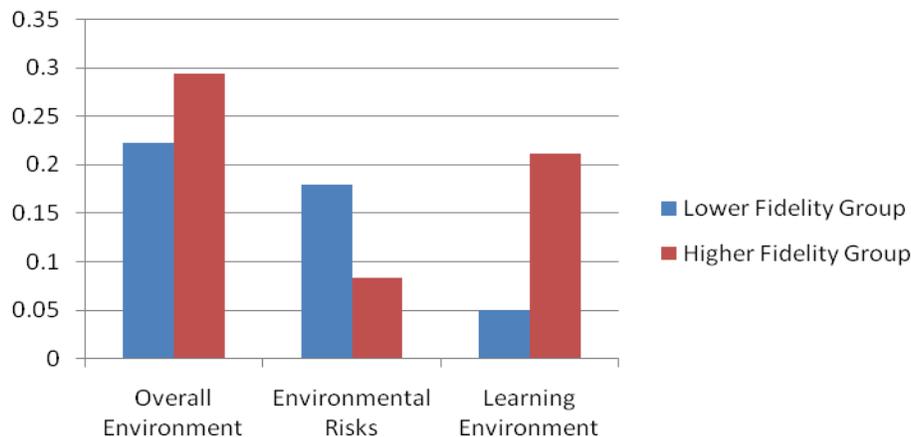
NCFAS-G Child Well-Being Change Scores



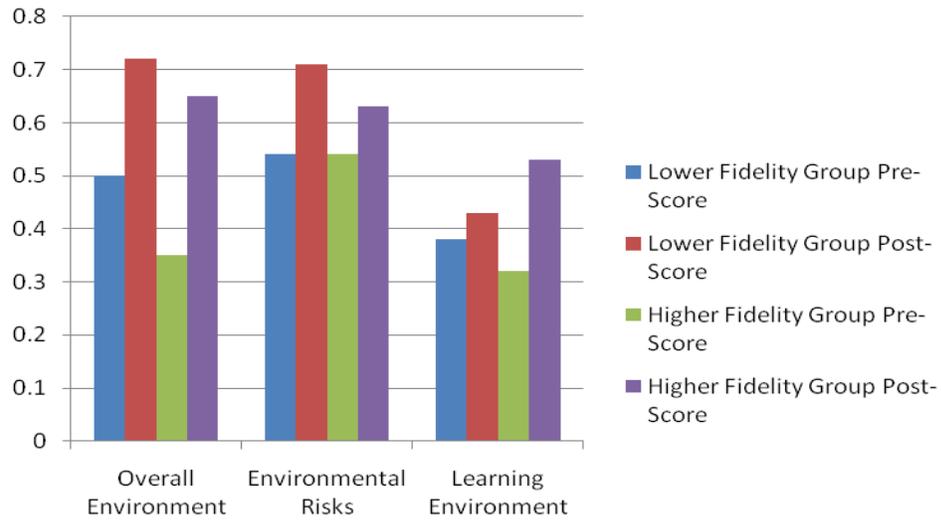
More significant findings were discovered during within group analyses. Comparisons were made between intake and discharge ratings on the NCFAS-G for those children and families in the Higher Fidelity group and for those children and families in the Lower Fidelity Group. The Higher Fidelity group had significantly higher scores at discharge compared to intake for ten NCFAS-G scales and sub-scales compared to the Lower Fidelity group (see attachments for more data detail). Specifically, the Higher Fidelity group showed:

1. Improvement in family environment: Fidelity to the Wesley Spectrum Services In-Home Family Therapy (WSIH) Model lead to larger change-scores in both overall environment and learning environment. However, within group analysis found significant increases in both fidelity groups for the Overall Environment Scale scores from intake to discharge. This means that being in therapy regardless of the therapist's fidelity to the model was enough to make significant improvement in a family's overall environment but larger increases were gained with higher model fidelity.

NCFAS-G Environment Change Scores



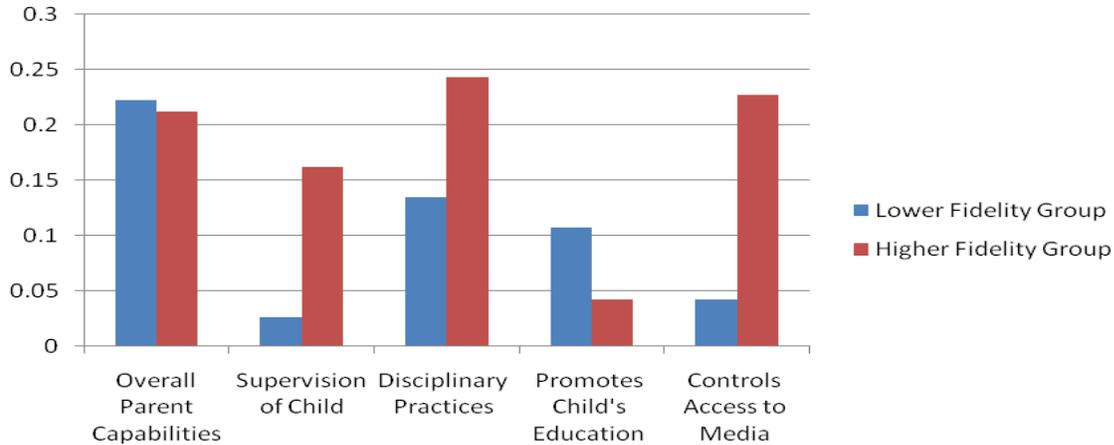
NCFAS-G Environment Pre-Post Scores



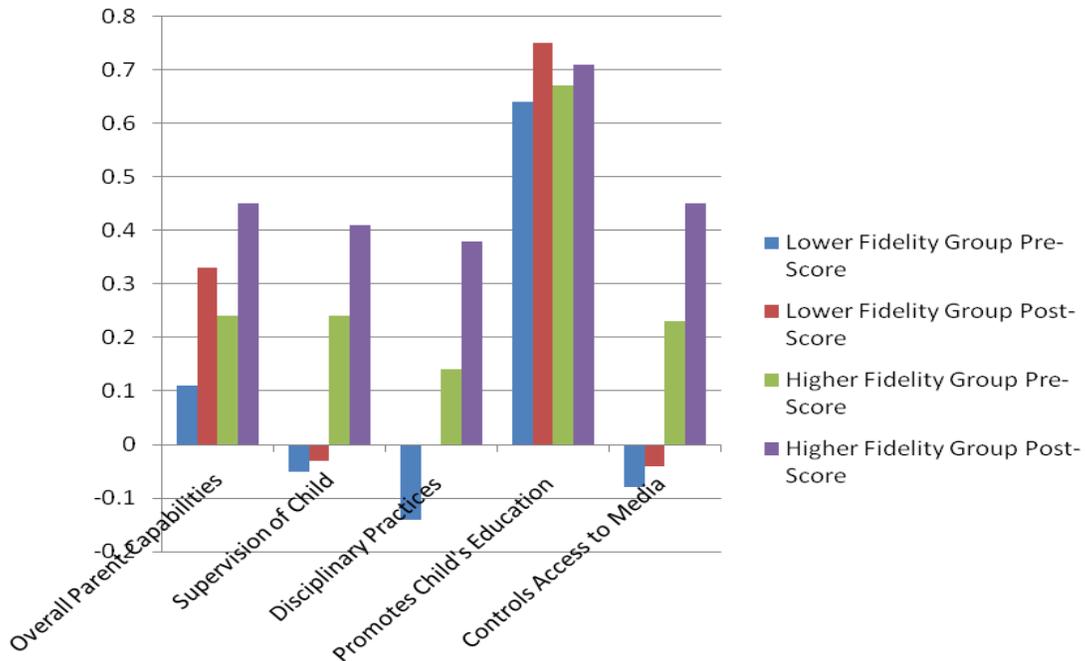
Overall Environment--Lower ($p < .01$) and Higher ($p < .05$)

2. Improvement in Parent Capabilities: Higher model fidelity created larger change-scores child supervision, disciplinary practices, and controls access to media/reading materials. Though two groups were similar in their change-scores for overall parent capabilities, the average score at discharge was higher for the Higher Fidelity group. Only the Higher Fidelity group had significantly higher scores at discharge compared to intake for disciplinary practices and controlling access to media/reading materials.

NCFAS-G Parent Capabilities Change Scores



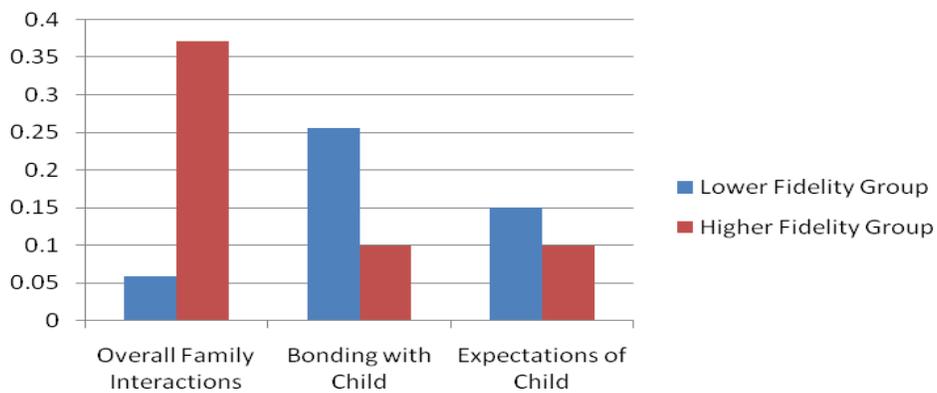
NCFAS-G Parent Capabilities Scores



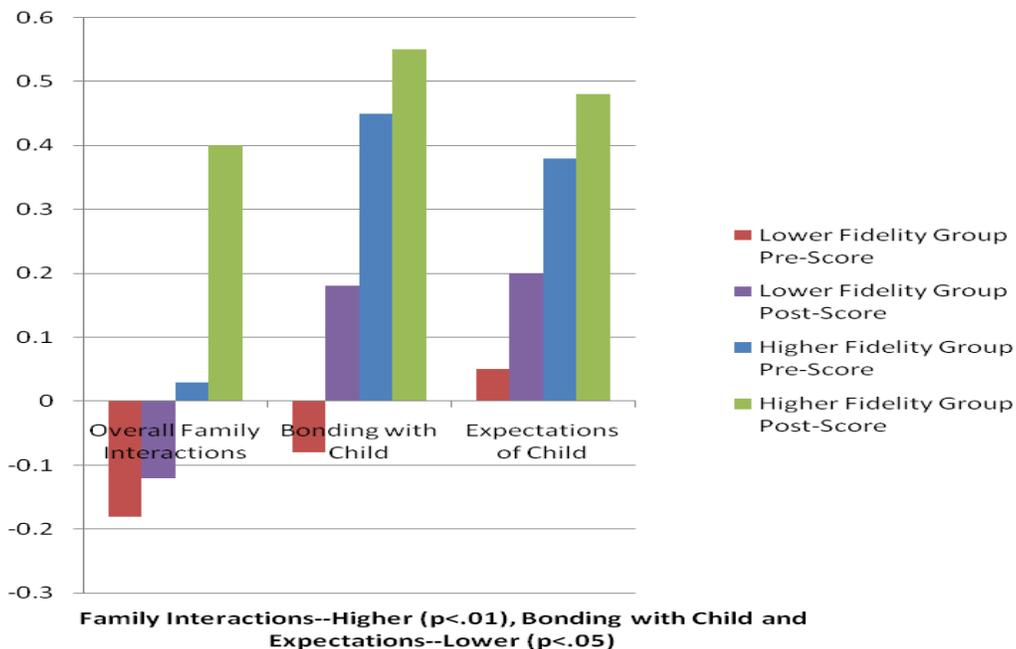
Disciplinary Practices--Higher (p<.05), Access to Media--Higher (p<.05)

3. Improvement in Family Interactions: Higher levels of fidelity to the WSIH model lead to larger change-scores for overall family interactions and significantly higher scores at discharge than intake. Interestingly, lower fidelity to the model lead to larger change-scores for bonding with the child and setting expectations for the child, suggesting either that any treatment could improve this dimension of family interaction, that the pre-assessment was so low for those within the Lower Fidelity group that more improvement was possible, or some combination of both. Compared to the Higher Fidelity group, the Lower Fidelity group had lower average scores for both bonding and setting child expectations at discharge; therefore, despite showing greater improvement, they're still not functioning at the higher level of the Higher Fidelity group.

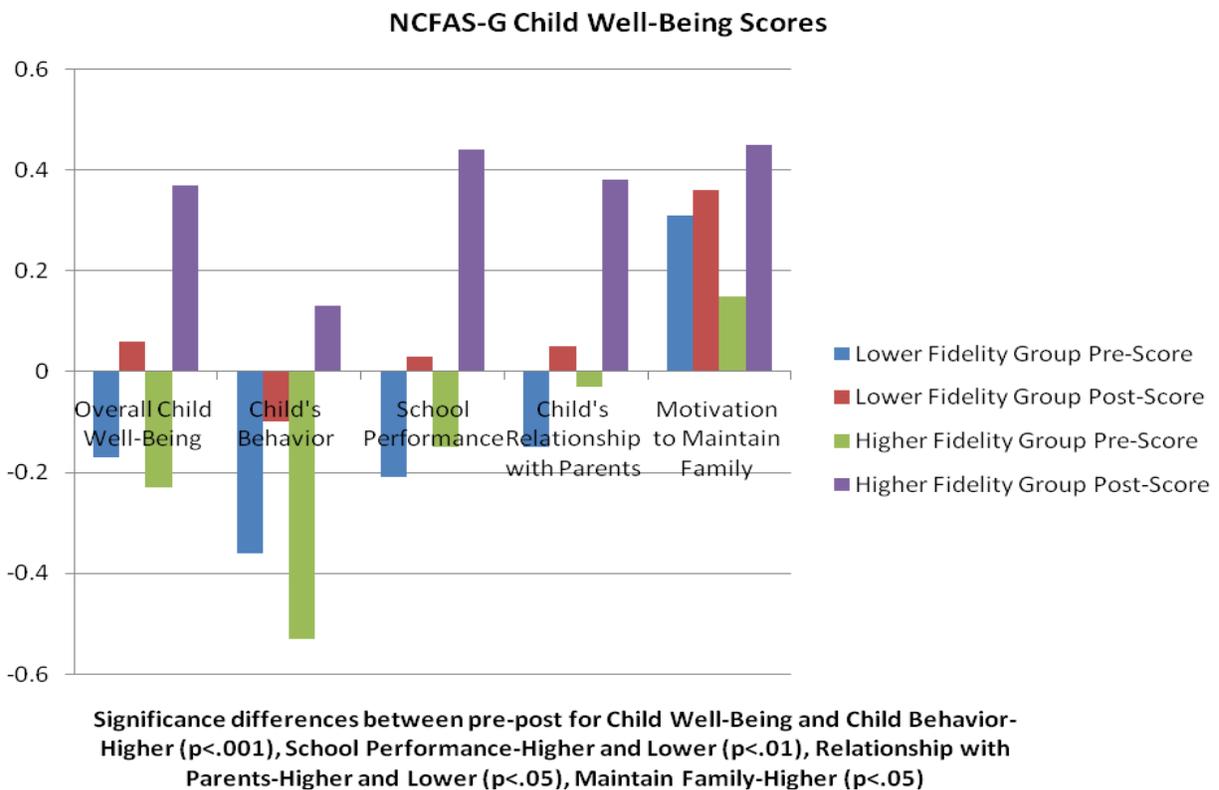
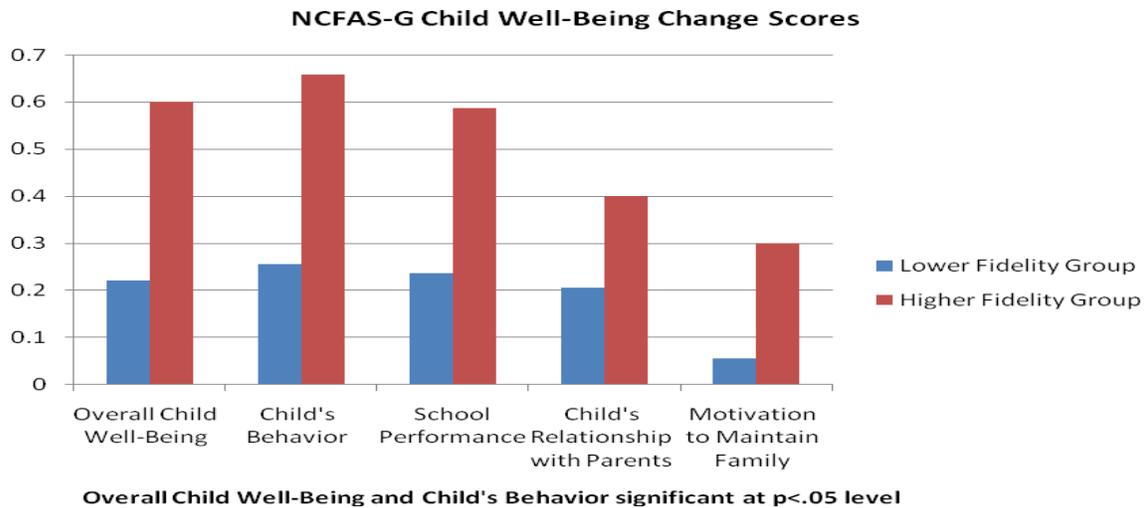
NCFAS-G Family Interactions Change Score



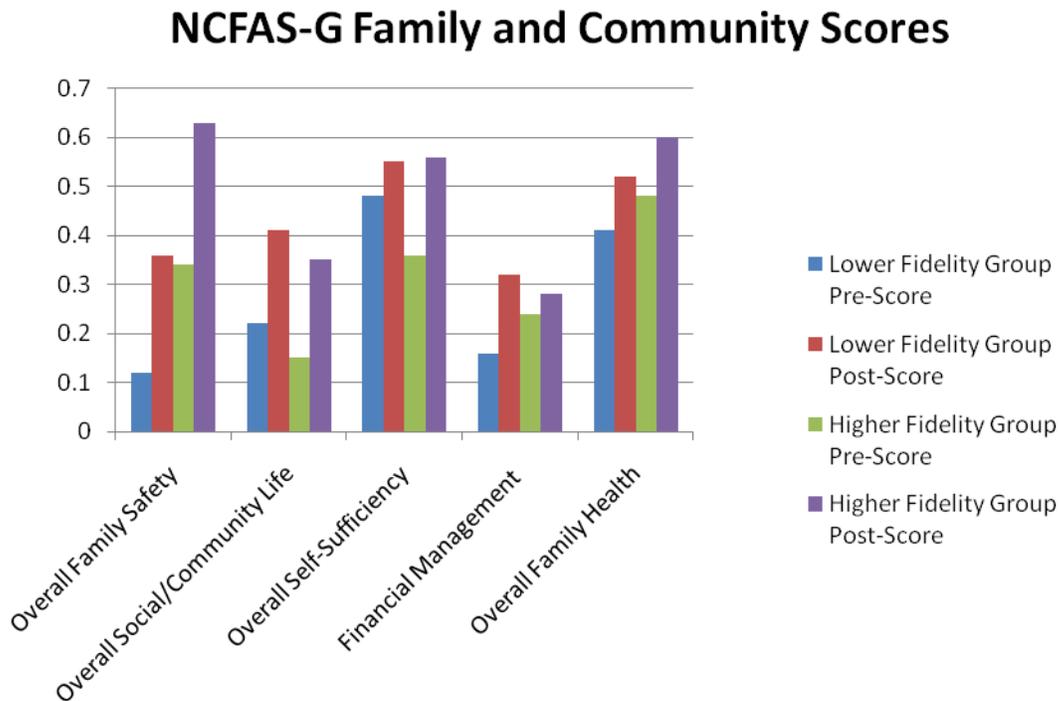
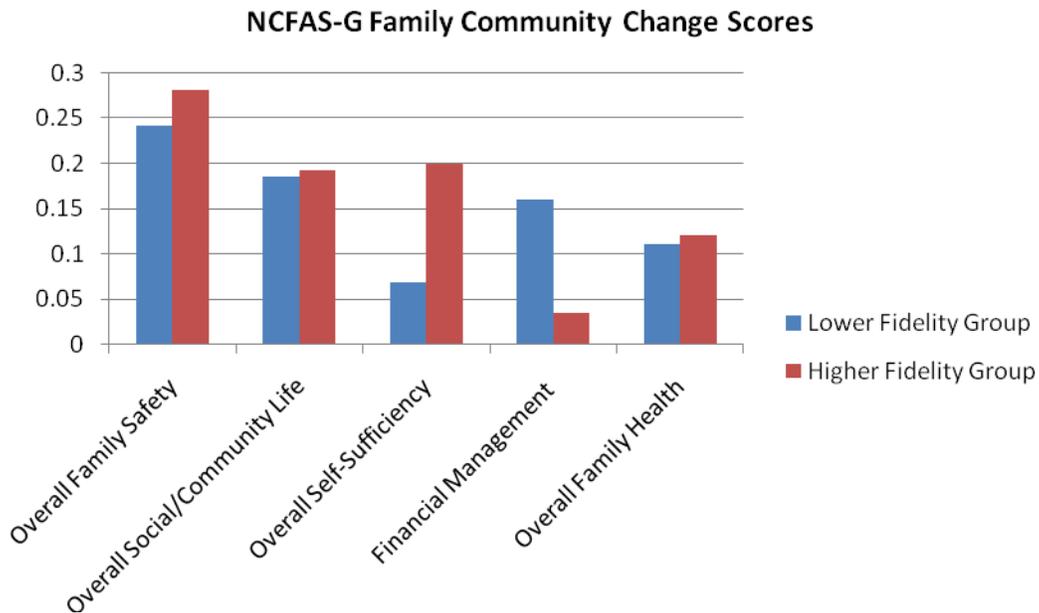
NCFAS-G Family Interactions Scores



4. Improvement in Child Well-Being: Children whose therapist adhered to the WSIH model with higher fidelity had on average larger change-scores in child well-being than those children whose therapist had lower levels of model fidelity. These differences were significant for overall child well-being and child behavior. The Higher Fidelity group, compared to the Lower Fidelity group, had higher average scores at discharge for overall child well-being, child behaviors, school performance, child relationships with parents and motivation to maintain family.



5. Improvement in Family Self-sufficiency: Higher fidelity to the WSIH model lead to larger change-scores in family safety, social/community life, self-sufficiency, and family health, but not in financial management. The large change-score for overall family safety represented a significant increase in the Higher Fidelity group's scores from intake to discharge.



Family Safety-Higher (p<.05)

Comparison of Wesley Spectrum Services In-Home Family Therapy (WSIH) and the Brand-name Program, Multisystemic Therapy (MST)

Using data from the 2010 report of the MST Institute, comparisons were conducted between two groups of Wesley Spectrum Services In-Home Family Therapy cases (all cases and only Higher Fidelity cases) and the MST outcomes⁴. The results for both groups of WSIH cases showed similar or better outcomes compared to the MST Blueprint program.

Demographics and Length of Treatment MST and WSIH similarities and differences: In terms of demographics, MST and both groups of WSIH cases were similar for average age of youth. Both groups of WSIH cases had significantly more males and more cases referred by Social Services compared to MST. When comparing the statistically adjusted lengths of treatment⁵, there were no significant differences between WSIH and MST.

Demographic Comparisons between MST and WSIH

| Demographics | MST | WSIH—Higher Fidelity Only | WSIH-All |
|-------------------------------------|-------|---------------------------|----------|
| Average Age in years | 15.1 | 14.7 | 15.2 |
| Percent of Males | 66% | 40%** | 49%** |
| Percent African American | 26% | 44%+ | 46%** |
| Percent Caucasian | 42% | 48% | 47% |
| Percent Juvenile Justice Referral | 53% | 54% | 66%* |
| Percent CYF and WCCB Cases | 20% | 46%*** | 34%** |
| Average Length of Treatment in days | 129.5 | 173.5 | 136.8 |

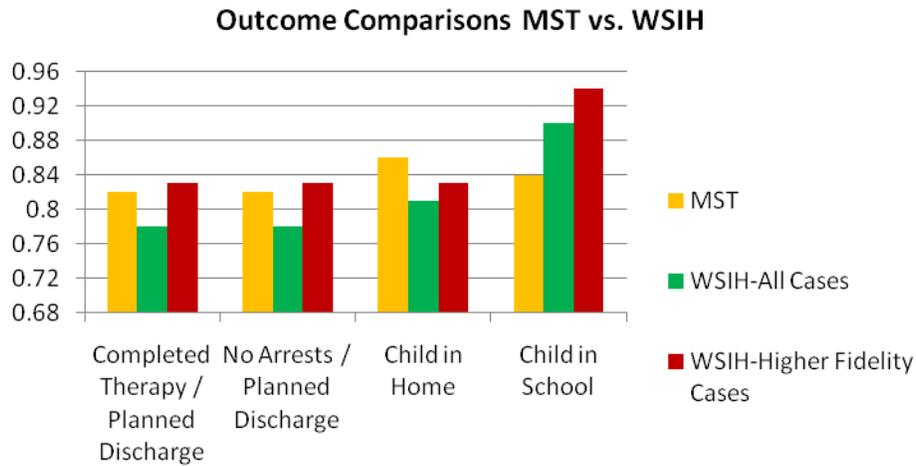
*p<.05, **p<.01, ***p<.001

Outcome Data: MST and WSIH similarities: Outcome data for the Wesley Spectrum Services In-Home Family Therapy program were statistically similar to the MST reported data for the percentage of children/youth who completed treatment, stayed with their families and were not arrested (defined within the WSIH data as a planned discharge). However, WSIH youth in

⁴ MST Institute (2008). MST Final Data Report. www.mstinstitute.org/2008-mst_data_report-summary.pdf.

⁵ Length of treatment for WSIH was positively skewed especially for the Higher Fidelity group (X=189, SD=125.5, Skewness=1.25, SE Skewness=.35). Two outliers that were greater than 2 SD above the mean were removed reducing skewness but the variable was still not normally distributed. The data were transformed using log₁₀ to address non-normal distribution prior to running the one sample t-test. The one sample t-test with the transformed values was non-significant at the p=.14 level.

the Higher Fidelity group were significantly more likely to be in school at the completion of therapy.⁶



Child In School significant for MST vs. WSIH Higher (p<.05).

Cost Comparison: MST and WSIH

Given that the Wesley Spectrum In-Home program and MST programs experience no statistical difference in outcomes - outcomes which include a decrease in arrest rates, an increased ability for the child to remain in their home and an increased ability for the child to stay in school, the question remains as to what is the cost of achieving comparable outcomes or the relative value of each. This analysis equates the benefits of the MST programs and the Wesley Spectrum in-home program.

The Washington State Institute for Public Policy has published a highly influential paper on the benefits for early intervention programs for youth that puts dollar amounts of the benefits to participants, their families and the communities they live in (Aos, Lieb, Mayfield, Miller, and Pennucci, 2004). It finds that the main financial benefit for these programs is in reduction of crime. This estimate is conservative and recognizes that there are other benefits that result from these programs. In addition, because the benefit is attributed to those who complete the programs, only those who completed these programs were included in the analysis. Only 49% of Allegheny County clients with closed cases who received MST treatment completed it versus 78% of closed cases in the locally developed, Wesley Spectrum In-Home program.

⁶ Measured by the NCFAS-G sub-scale of School Performance with “in school” being coded as having a score of a “2, clear strength” through a score of “-1, mild problem.”

Table 1: Costs and Benefits of the In-Home Program vs. MST

| Program | Benefit | Completed cases | Average hrs/client | Cost/unit service | Cost per completed case | Return on investment (per \$1 invested) |
|-------------------------------------|--------------------------|--|--------------------------------|-------------------|--------------------------|---|
| MST (Allegheny County) ⁷ | \$14,996 per participant | 22 or 49% of all served (12 months 2010) | 55.5 hrs/client | \$138.96/hr | \$8,531.474 ⁸ | 0.76 |
| Wesley Spectrum In Home | \$14,996 per participant | 69 or 78% of all served (six months 2011) ⁹ | 48.16 hrs/client ¹⁰ | \$51.94/hr | \$1,936.78 | 6.74 |

Each dollar invested into MST programs paid for by Allegheny County Department of Human Services sees a \$0.76 return on their investment. This means for every dollar invested, the investor receives their dollar back and an additional \$0.76. However, for every dollar invested in the Wesley In-Home Program, investors see a \$6.74 return on their investment. This return on investment is attributable to two main

⁷ These numbers come from the Allegheny County Department of Human Services and are how much they pay an organization to provide MST and how many clients completed treatment during FY 2010.

⁸ This cost includes the cost per client as well as the \$18,000/year that an organization must pay for a license to use the MST program. This information is found at Duncan, Melanie. (2007). "Multisystemic Therapy (MST) for Juvenile Offenders". *SAMHSA's National Registry of Evidence Based Program's Practices*. Even without this cost, the return on investment is still much less than the Wesley Spectrum program (1:0.94). In addition, there are other fees that an organization might have to pay (start-up fees, trainings, etc), but were not included in this analysis given the limits of time and scope.

⁹ This number was calculated by multiplying the number of closed cases (89) by the percentage of those who completed treatment (78%).

¹⁰ This number was calculated using the average paid hours for juvenile court clients, which includes both service hours and travel time since both are included in the costs of the program.

factors – three times as many young people (and their families) completed the Wesley program than the MST programs administered in Allegheny County and the rate per hour is over two and half times more for MST programs than for the Wesley program.

This analysis uses a conservative estimate for the benefits of these juvenile intervention programs, though other reports have valued the benefits at a much higher rate. The value of staying in school and staying in their home is not quantified, though an extremely important outcome and may be quantified in the future.

Conclusion

Our work in building evidence-based, locally-developed youth service models convinces us that many programs can achieve a higher level of “evidence supported practice” using our methodology, Fidelity Management. In so doing, we think there are benefits at various levels:

Benefits to the community/state/funder:

- Services could be purchased with continued guarantees of positive outcomes at a lower cost than the current, limited portfolio of brand-named models or practices;
- Increased assurance that those needing services would be able to access the service. Criteria for accessing evidence-based programs based on randomized-clinical trials restrict access);
- Fidelity metrics (measuring what practitioners do) would be available to understand the use of evidence-based interventions and could be used to increase the scaling up of such practices within many more community-based programs;
- Caseworkers/Probation Officers would have improved information for making referrals to effective programs – ones they already know and with which they have relationships. They would have access to simple metrics showing the degree to which a program or service is adhering to evidence-based practices and the outcomes it is achieving. They would have a way of “trusting” the promised outcomes.
- Increased return-on-investment in social service programs could be realized.

Benefits to youths and consumers:

- Access to services that are more likely to produce positive results would increase;
- Exactly what services, practices, dosages, and costs that are needed to produce positive results would be more understandable;
- Advocacy for research supported practices and results would be more available.

Benefits to service providers:

Service providers need to innovate through austerity. Becoming a home-grown evidence-based model has a number of advantages in such an environment including:

- Increased confidence that what the program does in fact influences (or contributes substantially to) positive outcomes (increased evidence for cause-effect relationship);
- Increased confidence that program costs for the produced outcomes are competitive with others in the industry and most likely substantially less than the proprietary brand-name models;
- Increased ability to negotiate with funders on VALUE not just on mission and outputs alone;
- Increased ability to preserve program/organizational heritage, seasoned staff, investment in organization/program identity and reputation. Build on it instead of doing away with it;
- Increased ability to adapt to new populations of consumers and new needs;
- Increased ability to use the fidelity tools and outcome measures to incorporate newly emerging evidence-based practices;
- Increased sustainability of positive practices, even with changing funding scenarios;
- Increased stakeholders' wisdom and abilities to support existing programs.

It all comes down to this:

- There are hundreds of studies showing “what works” in youth services. What we need are the steps practitioners can use to implement these findings (and future ones) well.
- Brand-name evidence-based models may be useful for launching a program in a system that lacks existing program services, but they need not be used to replace existing programs or to maintain a program to achieve high levels of practitioner fidelity to researched interventions.
- The only way to know if these researched practices (and the ones that will be generated in the future) are used is to train and supervise/monitor, and provide feedback to practitioners to assure that they are consistently using researched based practices. Program-specific tools for tracking fidelity to researched practices are essential for this next evolution of evidence-based practice.
- Client outcomes will be more positive the degree to which a program achieves high fidelity to the evidence-based practices most relevant to its service population and community.

References

Aos, T., Lieb, R., Mayfield, J., Miller, M., and Pennucci, A. (2004). Benefits and costs of prevention and early intervention programs for youth. Washington State Institute for Public Policy. Olympia, WA. <http://www.wsipp.wa.gov/rptfiles/04-07-3901.pdf>

Best Practices in Juvenile Justice Reform (2011). Princeton-Brookings, http://www.princeton.edu/futureofchildren/publications/highlights/18_02_Highlights.pdf

Campbell, K. and Menezes, R. (2011). Study of 100 fast-growing organizations finds striking consistency within four critical areas. Alliance for Children & Families Magazine, Issue 1 — 2011, pp. 14-18.

Chambers, B. (2010). Bringing Evidence to Practice in Juvenile Justice: Mark Lipsey, <http://blog.reclaimingfutures.org/juvenile-justice-system-evidence-based-practices-Mark-Lipsey> November 16, 2010.

Fagan, A.A., Hanson, K., Hawkins, J.D., and Arthur, M.W.(2008).Implementing effective community-based prevention programs in the community youth development study. Youth Violence and Juvenile Justice.

Fixsen, D.L., Naoom, S.F., Blase, K.A., Friedman, R.M., and Wallace, F.(2005).Implementation research: A synthesis of the literature. Tampa: University of South Florida, Louis de la Parte Florida Mental Health Institute, National Implementation Research Network (FMHI Publication #231).

Fixsen, D. and Blasé, K (2011). National Implementation Research Network. <http://www.fpg.unc.edu/~nirn/>

Gottfredson, D.C., and Gottfredson, G.D.(2002).Quality of school-based prevention programs. Journal of Research in Crime and Delinquency.

Hallfors, D., and Godette, D. (2002).Will the “principles of effectiveness” improve prevention practice? Early findings from a diffusion study. Health Education Research.

Lipsey, M. (2009). The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. Victims and Offenders, volume 4, issue 2, pp. 124-147.

Lipsey M., Howell, J., Kelly, R., Chapman, G., Carver, D. (2010). Improving the Effectiveness of Juvenile Justice Programs: A New Perspective on Evidence-Based Practice. Center for Juvenile Justice Reform

Chen, H.T. & Garbe, P. (2011). Assessing program outcomes from the bottom-up approach: An innovative perspective to outcome evaluation. In H.T. Chen, S.I. Donaldson, & M.M. Marks (Eds.), *Advancing validity in outcome evaluation: Theory and practice*. New Directions for Evaluation, 130, 93-106.

Attachment: NCFAS in Detail

| NCFAS-G Scales and Sub-scales | Lower vs. Higher Fidelity | Lower Fidelity Only | Higher Fidelity Only |
|-----------------------------------|---------------------------|---------------------------|---------------------------|
| | Larger Change-scores | Pre vs. Post Significance | Pre vs. Post Significance |
| Overall Environment | High | p<.01 | p<.05 |
| Environmental Risks | Low | | |
| Learning Environment | High | | |
| Overall Parent Capabilities | Same | | |
| Supervision of Children | High | | |
| Disciplinary Practices | High | | p<.05 |
| Promotes Child's Education | Low | | |
| Controls Access to Media | High | | p<.05 |
| Overall Family Interactions | High | | p<.01 |
| Bonding with Child | Low | p<.05 | |
| Expectations of Child | Low | p<.05 | |
| Overall Child Well-Being | High | | p<.001 |
| Child's Behavior | High | | p<.001 |
| School Performance | High | p<.01 | p<.01 |
| Child's Relationship with Parents | High | p<.05 | p<.05 |
| Motivation to Maintain Family | High | | p<.05 |
| Overall Family Safety | High | p<.05 | p<.05 |
| Overall Social/Community Life | Same | | |
| Overall Self-Sufficiency | High | | |
| Financial Management | Low | | |
| Overall Family Health | Same | | |

| NCFAS-G Scales and Sub-scales | Percent All Cases with 75% improvement in NCFAS-G Scores | Percent Higher Fidelity Cases with 75% improvement in NCFAS-G Scores |
|-----------------------------------|---|---|
| Overall Environment | 22.9 | 26.4 |
| Environmental Risks | 13.4 | 12.5 |
| Learning Environment | 12.9 | 15.8 |
| Overall Parent Capabilities | 21.7 | 21.3 |
| Supervision of Children | 18.7 | 18.9 |
| Disciplinary Practices | 20.4 | 21.6 |
| Promotes Child's Education | 9.6 | 8.3 |
| Controls Access to Media | 15.2 | 22.7 |
| Overall Family Interactions | 20.2 | 25.7 |
| Bonding with Child | 22.8 | 17.5 |
| Expectations of Child | 17.6 | 17.5 |
| Overall Child Well-Being | 39.4 | 48.6 |
| Child's Behavior | 35.1 | 52.6 |
| School Performance | 37.6 | 50 |
| Child's Relationship with Parents | 29.1 | 32.5 |
| Motivation to Maintain Family | 22.3 | 30 |
| Overall Family Safety | 23.1 | 25.1 |
| Overall Social/Community Life | 20.8 | 23.1 |
| Overall Self-Sufficiency | 14.9 | 16 |
| Financial Management | 16.7 | 13.8 |
| Overall Family Health | 15.4 | 16 |

Attachment: Additional Data Useful for Continuous Quality Improvement

NCFAS intake profile: The cases that resulted in higher model fidelity and more positive outcomes also showed significantly higher “bonding” among family members (NCFAS) at the start of services; in other words, NCFAS low-bonding scores at intake were predictive of lower model fidelity (and subsequently poorer outcomes). Therefore, WSIH may want to focus on identifying new cases in which there is lower attachments among family members at the start of services for additional intervention work.

85% of the clients/families active in the In-Home program will participate in at least 4 family sessions per month until discharge. Based on these completed cases, this benchmark is not being met at this time. For all completed Therapist Checklists, the percentage of cases that had four or more family session of therapy per any given month ranged from 29% of cases during their second month to 88% of the cases during their seventh month of treatment. The overall average percent of cases meeting this benchmark of family visits across the 9 months of data is 51%.

85% of clients/families who have successfully completed the program will demonstrate 75% improvement overall on the NCFAS-G assessment at discharge. Based on these completed cases, this outcome is not currently being met. For all cases, the NCFAS-G scale that showed the highest percentage of cases with a 75% improvement is Overall Child Well-Being scale with 39.4% of the cases meeting this outcome. For those cases that were identified as having higher fidelity to the model this outcome is being met for 48.6% of cases for the Overall Child Well-Being Scale and for 52.6% of the cases for the Child’s Behavior Sub-Scale.