



Realist Evaluation of Human Services in Chautauqua County: Findings from local evaluation of SAMHSA-funded Chautauqua Tapestry, 2008-2015

September 15, 2015

Mansoor A. F. Kazi, PhD, Lead Evaluator, Chautauqua Tapestry; Patricia Brinkman, Director Community Mental Hygiene Services, Chautauqua County; Rachel Ludwig, Project Coordinator, Chautauqua Tapestry; Anne Bartone, PhD, Research Assistant; & Yeongbin Kim, Research Assistant.

Mansoor A. F. Kazi, PhD
Chautauqua Tapestry,
2 Academy Street, Suite 202
Mayville, New York 14757
Cell: 716 417 0419
Email: mkazi@albany.edu

Introduction

Jamestown Public Schools, Chautauqua County Mental Health Clinic, Family Services of Chautauqua Region and Boys and Girls Club are among the 28 agencies that participate in the SAMHSA-funded Chautauqua Tapestry's local evaluation. The local evaluation strategy utilizes a new approach to evidence-based practice based on realist evaluation, with the central aim of investigating what interventions work and in what circumstances (Kazi, 2003). This approach essentially involves the systematic analysis of data on 1) the client circumstances (e.g. demographic characteristics, cultural differences and needs, environments in which people live and function, and the nature of baseline target problems); 2) the dosage, duration and frequency of each intervention in relation to each client; and 3) the changes in the outcomes as observed through the repeated use of reliable outcome measures with each client. This is a mixed methods approach, combining the traditions of epidemiology and effectiveness research in social work. As the research designs unfold naturally, data analysis methods are applied to investigate the patterns between the client-specific factors, the intervention variables, and the outcomes. This analysis can be repeated at regular intervals and helps agencies to better target their interventions, and to develop new strategies for users in the circumstances where the interventions are less successful. The Tapestry evaluators assist the partner agencies in the repeated use of a reliable outcome measure (e.g. school outcomes) and the regular updating of information on their existing management information systems. This report is based on the realist evaluation process with the merger of the school district data to help with the investigation of what interventions work and with whom.

This report is based on findings from the application of robust research methods drawn from both effectiveness research and epidemiology traditions on management information system data routinely collected by human service agencies in Chautauqua County in New York State, all partners in the Chautauqua Tapestry System of Care funded by the federal Substance Abuse and Mental Health Services Administration (\$9,000,000; 2008 to 2015). As part of this grant, this evaluation strategy was offered to all human service agencies in the county, and to date 28 have taken part, including the county's mental hygiene and social services departments, not-for-profit agencies, and 7 school districts. All services (social services, mental health, schools, and probation) are evaluated in all human service agencies taking part.

Brief description of the evaluation strategy

The main purpose of the evaluation is to partner with human service agencies and school districts to utilize the data that is already collected by the agencies, to investigate what programs work and for whom, enabling the development of effective programs and/or the linkage of existing effective programs to the schools. This approach was first developed in Moray Council (Scotland) and also applied in Chautauqua County, NY, and in 2010, Chautauqua Tapestry (NY) received SAMHSA's Gold Award for Outstanding Local Evaluation, having used realist evaluation strategies in schools, mental health, youth justice and other human service agencies. These agencies routinely collect data that is typically not used for evaluation purposes. The 100% evaluation strategy utilizes a new approach to evidence-based practice based on the realist evaluation paradigm, with the central aim of investigating

what interventions work and in what circumstances (Kazi, 2003). This approach essentially involves the systematic analysis of data on 1) the service users' circumstances (e.g. demographic characteristics); 2) the duration and frequency of each intervention in relation to each service user; and 3) the repeated use of reliable outcome measures with each service user e.g. academic outcomes already recorded by schools. This is a mixed methods approach, combining the traditions of epidemiology and effectiveness research in human services (Kazi, 2003; Videka, 2003), helping participating agencies to use the data they have already collected at any given time. The main strength of this approach is that the data is already collected by the participating agencies and routinely entered in their management information systems. Each participating agency continues to manage and hold their own data, and uses a de-identified copy of the data which is analyzed with and for each agency for evaluation purposes. Also, the data analysis is done three times a year in real-time. A limitation is that the findings are for the agency, and each agency decides if and when to share these findings with others.

What difference did the evaluation make?

The strategy that is used in Chautauqua is Realist Evaluation, as applied in schools, mental health, youth justice and other human service agencies. One of the first adopters was Jamestown Public Schools (JPS), the largest school district in the county. Studies have indicated that school interventions can be utilized to make numerous improvements in the overall well being of students (e.g. Goldenson, 2011). In order to evaluate interventions, most researchers utilize pre/posttest designs, looking at the initial baseline levels of the targeted behavior before and after the intervention is implemented. Chautauqua Tapestry's evaluation team has been utilizing whole-school data for evaluation. Kazi (2014) evaluated school-based programs in Jamestown, NY using whole-school data, comparing academic outcomes of those receiving interventions with those who were not. It was found that, although Caucasian children and youth achieved better academic grades compared to Hispanic and African Americans, the binary logistic regression indicated that the actual predictors for improved academic achievement were the tutoring and mental health services as well as individualized education programs. It was found that those receiving these interventions had significantly improved their Grade Point Average (GPA) scores when compared to those that had not received these interventions. The data analysis helps school administration to identify gaps in care, and helped to build partnerships for additional mental health counseling on-site for the school district. At this time, 7 out of 18 school districts have participated in local evaluation. A very important additional method has come out of the work that would be carried out through the next four years, namely "Innovate Collaborate Educate ICE-8", enabling the participating agencies and schools to coordinate and maximize public, non-profit and private resources to deliver critical services to students and their families, thereby increasing student achievement and generating other positive health and wellbeing outcomes. This new all County program has been signed on by all the 18 school districts in Chautauqua County under the joint direction and support of Tapestry and of Erie 2 BOCES. Strategy for expansion is to build the necessary champions within each organization and to develop learning communities to support specific learning goals, based on findings from the data regularly analyzed with each agency.

Why did this evaluation have the impact that it did?

Tapestry's local evaluation demonstrates that human service agencies can use their data for evaluation and apply robust research methods to investigate their effectiveness in real-time. This realist evaluation in both Scotland and USA shows how the actual data routinely collected by human service agencies can be used for evaluation. Findings at regular intervals from a 100% evaluation with the entire population helps to better target and develop the interventions in a longitudinal, continuous evaluation. However, it is important to recognize that the impact of this evaluation is high because this is a partnership between academics and the human services, helping these services to utilize the data in a suitably de-identified form, and not the usual approach of the academics taking the data away and producing reports after practice has moved on.

For example, school based interventions are often used to target specific student behavior, including socio-emotional problems and dealing with stress. Many studies in both USA and United Kingdom have indicated that school interventions can be utilized to make numerous improvements in the overall well being of students. However, most evaluations of interventions targeting school students do not use academic outcomes routinely collected on all students and instead concentrate on the use of standardized measures or surveys mainly with smaller samples. On the other hand, as part of the strategy demonstrated here, Kazi (2014) evaluated school-based programs in Jamestown, NY using whole-school data, comparing academic outcomes of those receiving interventions with those who were not. A 100% evaluation with the entire population helps to better target and develop the interventions in a longitudinal, continuous evaluation.

Human service agencies routinely collect data, but this data is not typically used for evaluation. Tapestry's local evaluation demonstrates how to make the best use of this available data from the management information systems of schools, mental health and other services to continuously evaluate alongside the repeated universal screening tools and to develop effective services. Evaluation can have a maximum impact when used effectively to expand collaboration between human service systems. The agencies are able to develop their services and to make informed decisions based on the evaluation findings at regular intervals, helping to enhance practice by utilizing findings from data analysis methods drawn from a combination of epidemiology and effectiveness research to investigate if the comprehensive array of services and supports are working to provide better outcomes for youth and families. This evaluation strategy has helped to enhance the internal evaluation capacities within each of the agencies with regard to improved data collection within their management information systems, and the regular use of this data to investigate what interventions work and in what circumstances.

First Evaluation: WCA Hospital Chemical Dependency and Mental Health Programs

Women's Christian Association Hospital in Jamestown provides a supportive environment to receive confidential, compassionate, and professional care along the road to recovery. Treatment is personalized taking into consideration unique needs

and characteristics. This includes someone who may need both mental health and addiction services. Family members are included in the rehabilitation process, and a positive atmosphere of support and fellowship is developed.

The evaluators helped WCA hospital to analyze data for 355 (100%) outpatients receiving these services in the period 2006-2009. 73% were male, and the age range was from 16 to 64, with a mean age of 36 years. 38.6% had been treated also for mental health in the past. The goal of stopping the use of drugs was achieved with 138 (38.9%), partially achieved with 82 (23.1%) and not achieved with 135 (38%). It was found that the following were significant factors associated with fully achieving the goal of stopping the use of drugs:

- Total treatment visits. Those who had achieved attended on average 10 and those not achieved 4 sessions on average ($r = .466$, $n = 355$, $p = .000$).
- Arrested or not—24.6% of those who had been arrested achieved as against 41.8% of the rest ($r = .133$, $n = 355$, $p = .012$)= .53)
- Employed at discharge or not—51.6% of those employed achieved as against 28.6% of the rest ($r = .235$, $n = 355$, $p = .000$)

When the binary logistic regression model was applied (table 1), it was found that those who were employed were twice as likely to achieve the goals. Not being arrested was important for the sample but not a predictor for the outcome. It was also found that every additional treatment session attended was associated with increased likelihood of achieving goals, indicating the effectiveness of the treatment sessions.

Table 1: Binary Logistic Regression Model (WCA Hospital)

		Variables in the Equation					
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	employdischornot(1)	.960	.243	15.599	1	.000	2.612
	Total Treatment Visits	.060	.011	31.360	1	.000	1.062
	Arrestedornot(1)	.665	.341	3.799	1	.051	1.944
	Constant	-.655	.214	9.322	1	.002	.520

a. Variable(s) entered on step 1: employdischornot, Total_Treatment_Visits, Arrestedornot.

Child Welfare Study: Effect of Solution Focused Training on the Permanency outcome in child welfare (excerpts from Kazi, Coombes, Krause and Anderson 2015)

The Chautauqua Tapestry evaluators helped the county to compare the outcomes of children in foster care whose caseworkers received Solution-Focused Casework (SFC) training or not. The evaluation analyzed the child welfare outcomes of children ($n = 97$) who were discharged from foster care in the calendar year of 2009. The correlations of children's 'lengths of stay in foster care' were significantly different between caseworkers who received training in SFC and those who did not participate in the training. Children had shorter lengths of stay in foster care when served by caseworkers who participated in Solution Focused Casework training compared to the children's caseworkers that did not participate. The initial results suggest that there

may be a positive impact of this training on child welfare, although other factors also impacted on the length of discharge from foster care. This study demonstrated that the effectiveness of social work training can be evaluated using client outcomes in a realist evaluation.

Solution Focused Casework is goal directed and future oriented, and focuses on engagement of families, while highlighting the client's strengths and resiliency when addressing problems. Developed by Insoo Kim Berg and associates at the Brief Family Therapy Center, Solution Focused Casework training was structured from the model outlined in *Building Solutions in Child Protective Services* (Berg & Kelly 2000). The training of supervisors and caseworkers began in 2006, and the agency developed a training plan grounded in principles of collaboration, sustainability, and systemic adaptation.

The data was obtained from the Department of Social Services COGNOS reports that include multiple data sets. The analysis focused on a sample of (N) 97 children who were discharged from the local child welfare system in 2009. This was approximately 38% female and 62% male; 71% were described as white and 17.5% as Hispanic. The children's ages ranged from several days old infants to youth over 18 years old, and the largest age group in care was 14 to 17 years of age (31%) and the next group was infants days old to children two years (29%) and the third age group was children 10-13 years of age (17.5 %). Most of the children were discharged from foster care to a parent (38%), the second most common discharge was to adoption and the third was to a relative. The smallest percentages were youth released to another situation and discharged to their own responsibility. The youth who were discharged to their own responsibility either aged out in care or chose not to remain in the child welfare system when they turned 18 years or older. The agency's 2008 discharges statistics were comparable; the most frequently reported discharge was to a parent (33%) and second was to adoption (28%) (NYS, OCFS 2010).

In 2009, most of the children were in licensed district foster boarding homes placements, and 10 % resided in approved relatives' homes and institutions, and 9% in other living situations, 3% in Agency Operated Homes, and only 1% was in group homes. The average child was almost nine years of age, with a standard deviation (SD) of 6.22 from the mean, and the average child remained in foster care for approximately 17 months with a range of less than a month (days) to 79 months. The caseworkers serving the families were primarily foster care staff; however a few children were in care for very brief periods of time and may have been primarily served by a child protection caseworker. Seventy-six percent of the foster care staff were trained in SFC and served this population in 2009.

For the children who had a caseworker not trained in SFC, the median stay in care was 26.91 months; whereas children served by caseworkers who received Solution Focused Casework training were in foster care approximately 13.96 months. An independent- samples t-test was conducted to compare outcomes of children's time in foster care whose caseworkers were solution focused trained or not. The t-test indicated a significance difference of months in foster care, $t(95) = 3.36$, $p = .01$ ($<.05$). The t-test had a 95% confidence interval ranging from 5.3 to 20.6 months in foster care. There was a statistically significant decrease in children's length of stay in care served by caseworkers trained in SFC than caseworkers not trained.

Further analysis with Pearson Correlations Coefficients found that fewer hours of caseworker SFC training had a positive correlation with higher number of months a child remained in foster care. Sixty three percent of the sample was in foster care up to 17 months, and 36% had an average stay in care higher than 17 months or more. There was a notable negative relationship between discharge reasons of the child and the actual number of months in care, higher than 17 months, and actual time in care, $r = -0.71$, $p < .05$. There were several small positive correlations between months in care, the child's age and/or age group at discharge and age group and facility type. The correlations can be interpreted as lower hours of training, or no SFC training had a negative association with the amount of time in care, $r = -0.34$, $p < .05$ or SFC trained or not at $r = -0.31$, $p < .05$. Also, children whose discharge plan was to return to a parent were more likely to spend 17 months or less in foster care, than those discharged to other sources and remained in care 17 months or more. This is not an unexpected finding because once it is determined by social services that a child is not returning to a parent, the legal procedures begin. Family court legal procedures are often lengthy when terminating parental rights and as a result, a child's permanency is stalled.

The initial findings suggested that the significant variables were whether or not a caseworker was trained in Solution Focused Casework, the number of training hours, and the child's reason for discharge. The variables in the equation provided some insight into the length of time children remained in foster care and as stated previously, children remained in care less time if they returned to a parent. Fifty-nine children stayed in foster care up to 17 months and 34 children remained in care 17 months or more. The Chi Square showed a goodness of fit, 77.591(df 4) and lengths of stay in foster care could be explained by the variables in the model.

Initial analysis indicated that males stayed longer in care than females if they did not receive the SFC. The chi-square analysis (.115 Eta Square) explored the relationship between the variables and found a .006 for the Yates correlation and Cramer V at .313 at significance level of $p = .003$ for those who did not participate in SFC training with 'time in months in care of 17 months or higher'. In an independent t-test of 'time in care' there was a significant association of SFC trained or not and the number of months a child remained in care. Twelve percent of the variance of 'time in care' could be explained by the variable 'SFC trained or not'. Initially, in a test of between subjects, it appeared that gender may have a significant effect when controlling for 'time in care' and 'SFC training or not', but upon further comparisons this effect dissipated. The initial finding indicated that male youth with caseworkers trained in SFC did spend two months less in care than females whose caseworkers' were trained in SFC. Stated differently, females, whose caseworkers were not trained in SFC, stayed in care for a median of 21.57 months, whereas the male youth with caseworkers not trained in SFC stayed in care the median was 29.25 months. However, the numbers were small and it would be important to monitor this effect in future analyses.

This Tapestry evaluation study found some notable effects in the outcome of children discharged from foster care in the 2009 calendar year by comparing the outcomes of children whose caseworkers participated in SFC training and the other children whose caseworkers did not participate in the training. This study focused on the SFC training

received by caseworkers and whether or not this training had an effect or not, but did not focus on whether or not they used or learned the SFC skills. The findings indicated that the more SFC training a caseworker had in SFC the less time a child remained in foster care. SFC is one type of the many interventions used in the child welfare system and the study's aim was to investigate the effect of this training on the permanency outcome. Historically, SFC was primarily used in mental health and addiction treatment settings and was later adapted and modified to better serve or address family needs in the child welfare system (Antle et al. 2008; Christensen, & Antle 2009).

Overall, there were significant findings of foster children who had caseworkers trained in SFC; the time in care was shorter (13.96 months) than those not trained in SFC (26.91 months). The noted negative correlation effect was when the caseworkers training hours increased, the children's months in care decreased. These findings supported the hypothesis that caseworkers who receive SFC training are more likely to have better permanency results with the children served in foster care, than caseworkers not trained in SFC. It was also found that older youth being discharged to their own responsibility remained in foster care 10.67% longer than other children. However, when controlling for these factors children with a caseworker trained in SFC spent less time in care.

The strength of this study is in the realist evaluation approach of utilizing the available data, and examining patterns in the various factors. This study can also be replicated in other counties or states where similar management information systems are used, and where the workers' participation in trainings can be monitored.

Evaluating Resource Center's Mental Health Programs

The Tapestry evaluators assisted the Resource Center to utilize the data they already collect to investigate the reasons behind the 'no-show' rates in mental health, and to help reduce them. The Resource Center's mission is to support individuals with disabilities and other challenges in achieving maximum independence, contributing to their community, experiencing lifelong growth, and enjoying the quality of life. Over the years, the Resource Center has developed a range of health, behavioral health and health-related services that are available to anyone, in order to ensure the well-being of our entire community.

It was found that 956 people had been served in the period July-August 2011; and of these 51.5% had attended 4th street clinic and 48.5% had attended Lake Shore Drive. The age range was from 0 to 97, with a mean age of 37 years; and of these 956, 61.5% were female and 80% were white Caucasian. The average number of appointments was 2.6, ranging from 1 to a maximum of 13. The largest proportion (51.2%) was diagnosed with bipolar, and the second largest (26.5%) with mood disorders. However, amongst those aged 6-18 years, the largest proportion (40.2%) was diagnosed with ADHD.

Table 2: Binary logistic regression (no show at least once versus the rest)

Variables in the Equation						
	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a						
age range			11.489	3	.009	
age range(19-21)	.960	.311	9.536	1	.002	2.611
Constant	-1.222	.090	184.527	1	.000	.295

With regard to the ‘no shows’, it was found that 262 (27.4%) had missed at least one appointment, as compared with 694 (72.6%) that had not missed any appointments. It was found that the predictor was age, (table 2) with those aged 19-21 years twice as likely to have at least one ‘no show’. Following this finding, the Resource Center developed strategies to reduce ‘no-shows’ through clinical engagement, motivational interviewing techniques, peer advocacy, cultural sensitivity, care coordination and improved communication across systems of care.

Tapestry Evaluation of County Mental Health Clinic

At the Dunkirk clinic, two years’ discharge data from January 2011 to December 2012 was analyzed, including 1231 people with an average age of 40 years, with just over half aged 21 years or younger. The average length of stay was 321 days, and the average number of appointments attended was 15.3, ranging from a minimum of 1 to a maximum of 130 appointments. 244 (19.8%) were discharged prior to admission, and of the rest, 27% were discharged with all or some goals met.

At the clinic in Jamestown, analysis was carried out on two years’ discharge data from January 2011 to December 2012, including 2368 people with an average age of 28 years, with just 40% aged 21 years or younger. The average length of stay was 405 days, and the average number of appointments attended was 12.1, ranging from a minimum of 1 to a maximum of 119 appointments. 812 (34.3%) were discharged prior to admission, and out of those admitted, all or some goals were met with 35%.

The Tapestry partners decided that repeated use of outcome measures would provide a better reflection of effectiveness; and some of these measures included the Child and Adolescent Needs and Strengths (CANS), and Strengths and Difficulties Questionnaire (SDQ). Utilizing the Tapestry local evaluation, the mental health providers also found that school outcomes could be utilized, such as academic outcomes and behavior, to measure the effectiveness of the system of care.

Merging data with Jamestown Senior High School

The active Jamestown Mental Health Clinic clients in the 2011-2012 year were merged with the Jamestown Senior High School, and it was found that, when

compared with others not receiving county mental health services (table 3), 62% had improved their GPA from the school year 2012 to the school year 2013.

Table 3: Active County MH * Improved GPA 2012 and 2013 Crosstabulation

Active County mental Health or not		ImpGPA2012and2013		Total
		Not improved	Improved	
Not county MH	Count	255	247	502
	% Within active County MH	50.8%	49.2%	100.0%
County MH	Count	19	31	50
	% Within active County MH	38.0%	62.0%	100.0%
Total	Count	274	278	552
	%	49.6%	50.4%	100.0%

The evidence indicates that the county mental health clinic at Jamestown is achieving all or some goals with over a third of those discharged, and just under a fifth are lost to contact. It was also found that, in the Jamestown Senior High School, the majority of those receiving county mental health clinic services had improved their GPA from the baseline year (2011-12) into the following school year (2012-13), and that a higher proportion had improved compared with the others not receiving these services.

Evaluation of Family Service of Chautauqua Region: Utilizing Repeated Outcome Measures and School outcomes

The Tapestry evaluators assist the partner agencies in the repeated use of a reliable outcome measure (e.g. school outcomes) and the regular updating of information on their existing management information systems. Family Service of the Chautauqua Region is a not for profit mental health agency and the agency mission is to provide high quality mental health services, and one program is School Based Social Work. This program intervenes with academically and/or behaviorally challenged students. The current format was adopted during the 2008-2009 school year due to the concerns of a rising dropout rate and a rapidly changing culture in the community, including but not limited to an increased Hispanic population. The school system also has a very large free and reduced lunch population, a marker for low-income families. The referrals are made by school liaisons, either school counselors or school-based Health Center professional. Once a referral is made permission is secured from legal caregivers to treat the youth, including a consent for communication between relevant school staff, the student, and their family. Students are seen onsite in discreet locations. Treatment goals are to provide counseling and case management services until student clients demonstrate reduced symptoms of the problems that initiated the referral. Additional goals may emerge from the counseling process. Students may be linked with community resources that support their individualized treatment plan. FSCR prioritize interventions that develop life/social skills that enhance the students' ability to be more independent. Therapists use a variety of interventions including but

not limited to: Cognitive Behavioral Therapy, Mindfulness Based Stress Reduction, Solution Focused therapy, insight oriented and relaxation techniques. The combination of approaches can be adjusted based on the needs and response of each student client, leading to individualized use of research supported interventions.

CANS-MH Improved or Not

This evaluation is based on data from August 2005 to March 2012, including 753 young people with an average age of 13 years. Just over half (53.5%) were female, and 84.5% were Caucasian. Child and Adolescent Needs and Strengths—Mental Health (CANS-MH) has a reported reliability ranging from 0.73 to 0.85 (Lyons et al. 1999). This measure was used every six months, and it was found that out of 475 young people where this measure was repeated, 329 (69.3%) actually improved. The paired samples t test also indicated that the improvement was statistically significant, with a change from a baseline average score of 34.02 to 28.87, an improvement of 5.15 for the whole group on the CANS-MH $\{t(474) = 11.661, p = .000\}$. When the database was divided between those who improved on CANS-MH total scores and those that did not, the binary logistic regression model indicated that those with worse baselines were 1.03 times more likely to improve than those with better scores for every single point on the score; those with adjustment disorder were twice as likely to improve, and those with anxiety were 4 times more likely to improve, than the others. It was also found that youth with oppositional defiant disorder were 3 times more likely not to improve on the CANS-MH.

Interim Findings from Jamestown Senior High School and the other Middle Schools Within Jamestown Public Schools in relation to FSCR

This evaluation is based on findings from the use of a whole-school database with Jamestown Senior High School, Jefferson Middle School, Persell Middle School, and Washington Middle School within Jamestown Public Schools. At the start of the evaluation, the school year 2010/11 was used as a baseline year, and the academic achievement of the students and the number of incidents of indiscipline in this baseline year was compared with the first quarter results in the following school year 2011/12. Then the year 2010/11 was compared with the full year 2011/12.

In the year 2010/11, 124 youth (5%, out of a total population of 2464 from grade level 6 to 12 within Jamestown Public Schools) received the Family Service of Chautauqua Region intervention. Within Jamestown Senior High School ($n = 1344$; FSCR intervention received by 102), those who received the FSCR intervention improved on average by 4.5% ($n = 99$) in the first quarter of 2011/12, as compared to 1.9% for those who did not receive this intervention ($n = 1,195$). This difference in improvement was statistically significant $\{t(1293) = 3.479, p = .001\}$. Amongst those who received the intervention, it was also found that 73 (73.7%) had improved their GPA in this period as compared with 752 (62.9%) of those that had not received the FSCR intervention.

Next, the GPA was compared from the entire year 2010/11 to the entire year 2011/12 in the Senior High School and the three middle schools taken together. It was found that 1805 students who did not receive the FSCR intervention reduced their GPA on

average -1.34%, whereas the 99 students who had received the FSCR intervention actually improved their GPA on average by 0.22%. This difference was statistically significant $\{t(1883) = 2.274, p = .000\}$. It was also found that 36 (45.6%) of those who had received the FSCR intervention improved their GPA as compared with 668 (37%) of the rest who did not receive the FSCR intervention.

When the incidents of behavior were compared between those who had at least one incident in the academic year 2010/11 with their number of incidents in the academic year 2011/12, it was found that those who had received the FSCR intervention increased by 1.5 incidents on average, as compared with 1.7 for the rest of the schools ($n = 2098$). It was also found that 69 (50.4%) of those receiving the FSCR had improved their behavior in the same period, as against 776 (37%) of the rest who had not received this intervention.

The evidence indicates that the Family Service of Chautauqua Region's intervention program is having a positive significant impact on school students' academic achievement when compared with those who did not receive the intervention, and that the FSCR intervention is also having a positive impact on the behavior of the students by reducing the number of incidents of indiscipline from one academic year to the next, when compared with the other students not receiving the FSCR intervention within Jamestown Senior High School and the three Jamestown Public Schools' middle schools.

Evaluation of Gustavus Adolphus (GA) Family Services and Learning Center

GA has had a tradition of caring and compassion toward youth since 1886, providing a full continuum of services for at-risk youth including residential, educational, and preventive services. All services are individualized, child-centered and permanency-focused. The guiding principles of the GA are as follows:

- **Growth** - All youth will be given the opportunity to reach their full potential.
- **Achievement** - All youth will experience a positive environment in which they can enjoy success.
- **Family** - We will work with the entire family moving all members towards permanence.
- **Safety** - We will treat youth, families, and each other with dignity and respect, creating a safe environment for everyone.

The Tapestry evaluators partnered with the GA to access 100% data from 2008 to 2013, with all 927 youth served, including those from Erie County (44%); day students (16%); and Chautauqua (8.6%). 55.7 % were male, 48.4% Caucasian, and 25.7% African-American. The mean length of stay was 286 days, ranging from 2 to 2086 days. The outcome measure used was Child and Adolescent Needs and Strengths (CANS-MH); reported reliability 0.73 to 0.85 (Lyons et al. 1999); then in 2011 moved to CANS Comprehensive. With regard to the residential youth, CANS-MH was used at least twice with 117 youth, and of these 96 (82.1%) improved on the total score. The mean improvement was 16 points on the total score, indicating a significant improvement $\{t(116) = 9.5, p = .000\}$. Those who had worse baseline scores improved ($r = .226, n = 117, p = .014$). The CANS-COMP was used at least twice with 105 youth, and of these 72 (68.6%) improved on the total scores. The

mean improvement was 12.7 points, and again this was a statistically significant improvement $\{t(104) = 4.9, p = .000\}$. The factors associated with improvement on CANS-COMP were ethnicity ($r = .198, n = 101, p = .047$); and county ($r = .203, n = 100, p = .043$). 78% of the Caucasian youth (42) improved, as against 53.7% of African American ($n = 41$). 56.4% from Erie County ($n = 55$) improved, as against 81% from Chautauqua County ($n = 21$); and with regard to ethnicity 47.8% of the youth from Erie County were African American as against 6.8% of those from Chautauqua County. Therefore, the evidence is that the GA as Tapestry System of Care partner is improving with all youth, but the improvements are far greater with those from Chautauqua County.

Tapestry Evaluation with Chautauqua Opportunities Inc. (COI) and Dunkirk Schools

This evaluation is based on data from the Dunkirk school district on K grade level students for the academic year 2011 to 2012; and data from the COI Head Start's management information system. We are able to identify the Beginning of Year (BOY) DIBELS literacy and mCLASS numeracy assessments for the school district's grade level K, and begin a comparison between those who had received pre-K Head Start and those that had not.

There were 163 at grade level K in 2011/12, and of these children, special educational needs were identified with 17 (10.4%); 65 (39.9%) were Hispanic and 107 (69.9%) were on free or reduced lunch. Out of total 163, 61 (37.4%) had received COI Head Start; and out of the 107 on free or reduced lunch, 45 (42.1%) had received COI Head Start. Comparison was made between COI Head Start and non-COI Head Start with the 107 on free or reduced lunch at the Beginning of Year (BOY) assessments.

With regard to the DIBELS composite assessment score levels, it was found that the Head Start Group was almost the same as the non-Head Start group. However, with regard to sound fluency levels, it was found that the Head Start group was performing better (35.6% achieved benchmark as against 27.4% of those that had not received Head Start). The differences were also in favor of the Head Start group with regard to the mCLASS numeracy assessments. It was found that 66.7% of the Head Start group had achieved emerging as against 56.5% of those that had not received head Start. The Head Start group also performed slightly well with regard to the mCLASS overall level (40% as against 37.1% strategic). The differences in respect of mCLASS counting levels were greater, with 51.1% of the Head start group achieving the merging level as against only 40.3% of the non-Head Start Group

Table 4 indicates the differences in the mean quantitative scores between the Head Start and non-Head Start groups (free/reduced lunches only, $n = 107$). It was found again that the Head Start group were doing better in respect of DIBELS sound fluency, mCLASS quantity discrimination, counting and missing number scores. However, the independent t test indicated that these differences were not statistically significant.

Finally, the DIBELS and mCLASS assessment levels were compared with regard to the actual number of days attended at Head Start, but only in relation to the group that had received Head Start. It was found that the mean number of days attended at

HEAD Start were greater for those that had achieved better in both DIBELS and mCLASS assessments. It was also found that there was a statistically significant difference between those that achieved intensive and those that had achieved higher than intensive in the mCLASS overall score levels $\{T(37) = 3.08, p < .004\}$.

Table 4: Mean scores between Head Start and non-Head Start Groups in K grade level

	Head Start or not	N	Mean	Std. Deviation	Std. Error Mean
DIBELS Sound Fluency Scores	Not Head Start	62	7.2258	9.44079	1.19898
	Head Start	45	9.0667	10.45423	1.55842
mCLASS Number Identification Scores	not Head Start	62	7.9032	8.20977	1.04264
	Head Start	45	6.6444	7.15782	1.06702
mCLASS Quantity Discrimination Scores	not Head Start	62	7.0000	6.20153	.78760
	Head Start	45	8.2444	7.54729	1.12508
mCLASS Counting Scores	not Head Start	62	22.6129	16.78832	2.13212
	Head Start	45	26.4000	18.15764	2.70678
mCLASS Missing Number Scores	not Head Start	62	1.8871	3.07344	.39033
	Head Start	45	2.3111	4.45085	.66349

Conclusion

When the Head Start group was compared with the non-Head Start group in the K grade level in Dunkirk School District, it was found that the Head Start group was not achieving better results in both the literacy (DIBELS) and numeracy (mCLASS) beginning of the year assessments in the school year 2011-2012. However, when the students on free or reduced lunch were selected, and within that group those who had received Head Start were compared with those that had not, the achievements were better for the Head Start Group. The indications are that Head Start is providing better pre-K preparation for children from the lower family income groups when compared with their peers who were also in similar family income groups but had not received Head Start. As Head Start is targeted at low-income groups, this appears to be a fairer comparison. It was also found that, with regard to the mCLASS overall numeracy assessments, there was a significant difference in mean attendance days between those that had achieved higher than intensive level (in the mCLASS overall numeracy assessment) and those that had achieved only the intensive level, indicating that the number of days attended does make a difference to the beginning of year assessments.

Tapestry Evaluation with Jamestown Community Learning Council (JCLC)

JCLC's mission statement is "to strengthen the capacity of families to nurture, support and provide for their child's academic development, emotional growth and physical well-being by establishing effective relationships between home and school." The Tapestry evaluation team helped JCLC to access and to analyze its management information system data for the period 2004/05 to 2008/09, including 1,419 children.

Of these, it was found that 398 (28%) had received mental health or social services; 437 (30.8%) had a sibling who had received mental health or social services; 178 (28.5%) were in a family where at least one family member had health problems; 201 (14.2%) had an IEP; 668 (47.1%) had no pre-K program; 230 (16.2%) had a disability; 546 (38.5%) were from single-mother families; and 703 (49.5%) had been on free lunch at some stage. When those receiving JCLC services were followed into the elementary schools, it was found that in the first DIBELS literacy test 27% were at the intensive level, but 73% had achieved higher than intensive, and the following factors were correlated with higher than intensive or not:

- Marital status. Married 78%; single 68% ($r = .096$, $n = 628$, $p = .016$).
- Number of addresses—77% of those who had not moved was higher as against 64% who had moved even once ($r = .123$, $n = 578$, $p = .003$)= .53)
- Child disability—76.1% without disability was higher, as against 54.8% of those with a disability ($r = .164$, $n = 628$, $p = .000$)
- Parent disability—75.4% of those with no disability were higher, as against 57.9% of those with disability ($r = .129$, $n = 628$, $p = .001$)
- Free lunch at some stage—68.6% of those on free were higher as against 87.6% of those not on free lunch ($r = .184$, $n = 628$, $p = .000$)
- IEP or not—41.5% with IEP higher, as against 81% of those with no IEP ($r = .354$, $n = 628$, $p = .000$)
- Pre-K or not—68.2% of those without pre-K were higher, as against 76% of those who had been in a pre-K program ($r = .085$, $n = 629$, $p = .034$)

The binary logistic regression model indicated that the predictors for achieving higher than intensive were lunch status, IEP and whether they had pre-K or not. Children on free lunch at some stage were twice as likely not to achieve higher than intensive; those on IEP were 5 times more likely not to achieve higher than intensive; and those with no pre-K were 1.5 times more likely not to achieve higher than intensive (table 5).

Table 5: Binary logistic regression: Achieving higher than intensive or not

		Variables in the Equation						
		B	S.E.	Wald	df	Sig.	Exp(B)	
Step 3 ^c	Free_Lunch(1)	.863	.279	9.551	1	.002	2.370	
	IEP or not(1)	1.655	.228	52.645	1	.000	5.234	
	prek or not(1)	.460	.210	4.794	1	.029	1.584	
	Constant	-2.303	.274	70.377	1	.000	.100	

Conclusion

73.3% of all children receiving any JCLC program had achieved higher than the intensive level at the first Dibels test in school; and the predictors for those not achieving higher than intensive were free lunch, IEP and whether or not they had been in a pre-K program.

Tapestry Evaluation in Partnership with Jamestown Public Schools: Utilizing whole-school data to investigate what works and for whom: Part 1.

An example of the realist evaluation local strategy is the use of a whole-school database with Jamestown Senior High School. The school year 2008–09 was used as a baseline year. The overall average cumulative school grade for 1,205 youths was found to be 80.19, but Hispanic youths had an average of 73.8 ($n = 117$) compared with 77.4 ($n = 88$) for Black and African American youths, and 81.2 ($n = 980$) for white non-Hispanic youths. Apart from ethnic origin ($r = .109$, $n = 1202$, $p = .000$), the other significant factors were found to be lunch status – whether or not they were eligible for free/reduced price ($r = .323$, $n = 1200$, $p = .000$), and the father’s level of education ($r = .076$, $n = 1155$, $p = .010$). All of these significant factors were included in a forward-conditional binary logistic regression model. In this model, youths were divided into those who achieved an average grade of 70 or less (18.2%) and those who had achieved an average grade greater than 70 (81.8%).

Table 6: Binary Logistic regression Model for those 70% or less, and those achieving school grades greater than 70% .

Predictors	B	S.E.	Wald	df	Sig.	Exp(B)
Fathers with 4 years of college	2.119	.615	11.863	1	.001	8.324
Fathers with more than 4 years of college	2.677	1.028	6.788	1	.009	14.543
Lunch Status (1)	1.445	.214	45.782	1	.000	4.240

Analyses found that predictors of school grades (table 6) included lunch status and the father’s level of education. Those receiving free lunches were four times more likely to have an average grade of 70 or less, whereas those with fathers with four years of college education were eight times more likely to have an average grade greater than 70. Ethnic origin was thrown out of the equation, indicating that the factors associated with family income were more important than ethnic origin. When the data for the first marking period in 2009–10 became available for 1,049 students, analyses found that the cumulative average grade had increased from 81.4 in the baseline year 2008–09 to 83.6 for this group and that this improvement was statistically significant $\{t_{1048} = 9.568$, $p = .000\}$ and that 64.7% of these students had improved their average grade. The school year 2008–09 was used as a baseline year.

This comparison between the baseline year and the first quarter of the second school year enabled the evaluation of particular school-based interventions using changes in cumulative average grade as the outcome. At this time, data were available on three interventions, as follows:

1. A behaviour specialist provided counselling that focuses on the management of emotional issues and anxiety. The intervention is based on the particular concerns of the student and they range from social and emotional issues to domestic problems, drugs and alcohol, and others, which manifest themselves in a failure to achieve in classes and difficulty adapting in school social situations. The Behaviour Specialist uses a variety of approaches to address the concerns including therapeutic counselling approaches, mentoring in the Masonic and Check and Connect models, and family outreach. Students receiving this programme improved their cumulative average grade by 5.9 (n

- = 40) compared to 2.1 (n = 1,009) for those who did not. This improvement was statistically significant {t 41.89 = -.3068, p = .004}.
2. Safari academic services provided by Striders consists of tutoring and athletics. Safari Tutors are academic support partners provided to the school through a partnership with Striders of Jamestown. The partnership is supported by a grant from the Sheldon Foundation and Jamestown Public Schools funding, and is designed to provide academic tutoring support to students in need. Guidance counsellors and teachers make the referrals, and the students are scheduled into the appropriate blocks of time within the school day. Tutoring is provided for all academic subjects and the tutors work with the classroom teachers to get the texts, assignments, projects and study guides that the students need to stay on top of their work. There is an after-school component of the same service that is called “Striders Tutors” and it operates daily. The youths receiving this intervention improved on average by 7.3 (n = 45) as compared to 2.0 for those who did not (n = 1,004). This improvement was statistically significant {t 1047 = 4.709, p = .000} in this school year, and subsequent school years also indicated a consistent pattern of improvement for those receiving the intervention, and remained an important predictor for improvement.
 3. Family Service of Chautauqua Region, a Tapestry partner agency, provided school-based therapy program. Through partnership with the agency and a United Way grant, students are able to access mental health services in the school building. This assures that the students get to their appointments and it helps to return them to their classes quicker and with strategies to help them cope. Grades of the youths who received this programme (n = 35) improved by an average of 4.5, as compared with 2.1 for those who did not (n = 1,010). However, this difference was not statistically significant in this school year. In subsequent school years, this intervention progressed to becoming a predictor for improvement.

This data analysis can be repeated at every marking period and also every school year when data from academic assessment tests and other exams becomes available. In this way, the partner agencies and school districts can work together to evaluate the impact of interventions on school outcomes. The local evaluation’s main purpose is to utilize the data that each partner agency already collects, undertake analysis of this data in a suitably de-identified form, and to help each agency investigate where it is more or less likely to be effective in meeting the needs of youths and families. It is making excellent progress in developing management information systems within each agency to examine the patterns in the data and to report every three months or whenever the outcome measures are repeated. This local evaluation strategy can address the twin problems of application of evidence-based practice and the evaluation of practice to investigate what works and in what contexts, providing regular analyses to inform practice as it unfolds.

Conclusion

In this example a large majority of the children had achieved over 70%, and therefore the investigation of patterns in the data was rather limited. However, the realist enquirer is never satisfied with appearances, and seeks to investigate the hidden processes that may explain the observable phenomena. The realist evaluation

paradigm has a central aim of investigating what interventions work and in what circumstances. This approach essentially involves the systematic collection of data on 1) the client circumstances, the nature and duration of the intervention; and 2) outcomes repeatedly measured with reliable instruments. The realist investigation includes repeated analyses of the patterns in the data in relation to these three sets of variables.

One method that can be applied in most practice circumstances is the development of binary logistic regression models where the service recipients can be divided into those that improve and those that do not at regular intervals during a prospective evaluation. The binary logistic regression method identifies patterns in the data where multiple factors are influencing the outcome, and selects the main factor or factors responsible for the outcome, with a prediction of the odds of achieving a given outcome in particular circumstances. This evidence provides information about the effectiveness of the models of intervention in terms of what works, for whom and in what contexts. The evaluation helps the project workers to better target their interventions, and to develop new strategies for users in the circumstances where the interventions are less successful. Chautauqua Tapestry System of Care received SAMHSA's Gold Award for Outstanding Local Evaluation in July 2010 for the implementation of this realist evaluation strategy.

Tapestry Evaluation in Partnership with Jamestown Public Schools: Utilizing whole-school data to investigate what works and for whom: Part 2.

Research indicates that the reform of the system of care and the use of initiatives such as wraparound are effective in improving mental health and functioning in school (Kutash, 2006; Reback, 2010; Goldenson, 2011). However, most studies have focused on at risk groups rather than the total school populations. The purpose of this evaluation was to help the school district undertake a 100% evaluation of school-based services, utilizing data on the entire school populations, in a longitudinal study.

The Sample and Demographics

This evaluation (Kazi & Kim, 2015) utilized a 100% sample of all students in Jamestown Senior High School who had a grade point average in the academic year 2010-11 (n = 1299). The grade levels ranged from 9 to 12 (table 7), and the students were almost equally divided between the four grade levels. A total of 51.1% (n = 664) were male, 73.1% (n = 950, table 8) were white Caucasian, 59.9% (n = 778) were on free or reduced lunch, 12.3% (n = 160) had an Individualized Education Program, i.e. that the young person had a disability and needed special education; 17.6% (n = 228) had mothers who had achieved less than school graduation (i.e. did not go to college); and 12.4% (n = 161) had fathers who had not achieved school graduation (i.e., did not go to college).

Table 7. Frequencies of Grade Levels in Baseline (2010-11)

Grade Level	N	%
9th	349	26.9
10th	340	26.2
11th	309	23.8
12th	301	23.2
Total	1299	100.0

Table 8. Frequencies of Ethnicities in Baseline (2010-11)

Ethnicity	N	%
Asian	11	0.8
African-American	84	6.5
Hispanic	193	14.9
Native American	10	0.8
Mixed Race	51	3.9
White Caucasian	950	73.1
Total	1299	100.0

Baseline outcomes for school year 2010-11

Grade Point Average (GPA) was calculated by the School District for all Jamestown Senior High School students who were enrolled in the school year 2010-11. The mean GPA for the students ($N = 1299$) was 81.35, ranging from 9.00 to 99.45 ($SD = 10.65$) and 23.3% ($n = 301$) had achieved a GPA of less than 75.

Patterns in the data were explored between the baseline GPA achieved 75 or higher, and the demographic (or contextual) variables to investigate the significant differences in achievement. Using the Spearman test, it was found that:

- More females (80%, n = 508) had achieved a GPA of 75 or higher when compared with males (73.8%, n = 490), and there was a significant correlation between gender and whether a GPA of 75 or higher was achieved or not ($r = .074$, $p < .01$, $n = 1299$, power = .99).
- Ethnicity was an important factor (table 9). More White Caucasian students (80.7%, n = 998) had achieved 75 or higher when compared with all others (e.g. Hispanic 63.7%, n = 123) with the exception of Asian children who had all achieved 75 or higher, but they were a small minority (n = 11). When the database was divided between White Caucasian and all others, it was found that there was a significant correlation (80.7%, n = 767; $r = .153$, $p < .01$, $n = 1299$, power = .99).

Table 9. Disposition of Ethnicities Based on GPA for School Year 2010-11

Ethnicity Category		GPA		Total
		Less than 75	75 or Higher	
Asian	Count	0	11	11
	% Within Ethnicity	.0%	100.0%	100.0%
Black African-American	Count	32	52	84
	% Within Ethnicity	38.1%	61.9%	100.0%
Hispanic	Count	70	123	193
	% Within Ethnicity	36.3%	63.7%	100.0%
Native American	Count	6	4	10
	% Within Ethnicity	60.0%	40.0%	100.0%
Mixed Race	Count	10	41	51
	% Within Ethnicity	19.6%	80.4%	100.0%
White Caucasian	Count	183	767	950
	% Within Ethnicity	19.3%	80.7%	100.0%
Total	Count	301	998	1299
	% Within Ethnicity	23.2%	76.8%	100.0%

- Free or reduced lunch was also a significant factor. A higher proportion of those on paid lunches (90.8%, n = 473) had achieved 75 or higher than those on free or reduced lunch (67.5%, n = 525; $r = .271$, $p < .01$, $n = 1299$, power = .99).

- More students who did not have an IEP achieved 75 or higher (80%, n = 911) than those who had an IEP, and the two variables were strongly correlated (54.4%, n = 87; $r = .199$, $p < .01$, $n = 1299$, power = .99).
- Whether students' mothers had less than school graduation or a higher level of education was also a significant factor, with a smaller proportion with mothers with less than school graduation achieving 75 or higher (60.1%, n = 137) when compared to those with mothers who had school graduation or higher levels of education (82%, n = 801; $r = .207$, $p < .01$, $n = 1205$, power = .99).
- Fathers' level of education was also a significant factor. More students with fathers who had school graduation or more achieved 75 or higher (82.4%, n = 664) when compared with those with fathers who had school graduation or higher levels of education (62.7%, n = 101; $r = .180$, $p < .01$, $n = 967$, power = .99).

The above significant factors were entered in a forward-conditional binary logistic regression model with 'achieving GPA 75 or higher or not' as the dependent variable (table 10). It was found that those on free or reduced lunch were 3.2 times less likely to achieve 75 or higher than their fellow students on paid lunch; those with an IEP were 2.5 times less likely to achieve 75 or higher than the others with no IEP; students with mothers who had less than school graduation were 1.8 times less likely to achieve 75 or higher than those with mothers who had achieved school graduation or a higher level of education; and those who were not White Caucasian (i.e. Black African-American, Hispanic and others) were 1.9 times less likely to achieve 75 or higher than those who were White Caucasian.

Table 10. Binary Logistic Regression Results for Variables Predicting GPA 75 or Higher

Variables in Equation	B	SE	Wald	df	Sig	Exp(B)
Step 4 (c)						
Free or Reduced Lunch (1)	1.16	0.22	28.15	1	0.00	3.18
IEP (1)	0.90	0.24	14.68	1	0.00	2.47
Mothers' Education Level (1)	0.58	0.21	7.95	1	0.01	1.79
White	0.63	0.19	10.95	1	0.00	1.88
Constant	-2.55	0.18	205.17	1	0.00	0.78

Note. IEP = Individualized Educational Programs.

GPA Achieved in the following School Year 2011-12

When the students who were enrolled in 2010-11 were followed into the end of the next school year ($N = 910$, as grade level 12 had moved on), it was found (table 11) that mean GPA was reduced by 1.7% from the year 2010-11 to the year 2011-12, and that this small difference was significant $\{t(909) = 7.23, p < .01\}$.

Table 11. Mean GPA in Baseline (2010-11) and Following School Year (2011-12).

	Mean	N	SD	SE
GPA in 2010-11(Baseline)	81.84	910	10.24	0.34
GPA in 2011-12	80.14	910	10.56	0.35

However, when the individual students' change from the previous year into the next was taken into account, it was found that 35.7% ($n = 325$) had actually improved their GPA and the mean change for those who improved was 4.6 GPA points, as compared with a mean reduction of 5.2 GPA points for those who had not improved (64.3%, $n = 585$).

It was also found that a higher proportion (50%, $n = 29$) of the students receiving FSCR ($N = 58$) had improved when compared with those not receiving FSCR (34.7%, $n = 296$, improved; $N = 852$). Similarly, it was found that 62.7% ($n = 32$) of those receiving Safari ($N = 51$) had improved their GPA when compared with the students not receiving Safari (34.1%, $n = 293$, improved; $N = 859$). However, as the mean improvement for all students who had improved was 4.6 GPA points, the patterns in the data were investigated between those who had improved by 5 GPA points or more and those who had not; and the Spearman test indicated that the following were the significant factors:

- 22.1% ($n = 25$) of students who had an IEP ($N = 113$) improved their GPA by 5 points or more when compared with 10.3% ($n = 82$) of those that did not ($N = 797$) in the school year 2010-11; and the variable 'IEP or not' was significantly correlated to 'improved GPA by 5 points or more or not' ($r = .121, p < .01, n = 910$).
- Safari tutoring program or not. It was found that 31.4% ($n = 16$) of those who had received Safari ($N = 51$) in the school year 2010-2011 had improved their GPA by 5 points or more in the following year when compared with those who had not received Safari (10.6%, $n = 91$; $N = 859$), and that receiving Safari or not was correlated to GPA improved by 5 points (or more) or not ($r = .148, p < .01, n = 910, \text{power} = .72$).

- Family Service of Chautauqua Region (FSCR) mental health services received or not. It was found that 29.3% (n = 17) of those who received these services (N = 58) in the school year 2010-11 improved their GPA by 5 points or more in the following year when compared with those who had not received FSCR services (10.6%, n = 90; $r = .142$, $p = .019$, $n = 910$, power = .72).

When the above significant factors were entered in a forward-conditional binary logistic regression model with the dependent variable GPA improved by at least 5 points or not (table 12), it was found that those who had received Safari were 2.8 times more likely to improve their GPA by 5 points or more when compared with those not receiving Safari; and the other predictors were receiving FSCR (3 times more likely) and having an IEP (2.3 times more likely).

Table 12. Binary Logistic Regression Results for Variables Predicting GPA improved by at least 5 points or not

Variables in Equation	B	SE	Wald	df	Sig	Exp(B)
Step 3 (c)						
Safari	1.03	0.34	9.18	1	0.00	2.79
FSCR	1.09	0.32	11.38	1	0.00	2.98
IEP	0.83	0.26	9.98	1	0.00	2.29
Constant	-.605	0.43	1.95	1	0.16	0.54

Note. FSCR= Family Service of Chautauqua Region. IEP = Individualized Educational Programs.

Conclusion

This evaluation demonstrates how the actual data collected by schools and human service agencies can be used for evaluation. In this example, the interventions were having a positive impact on the children and youth in Jamestown Senior High School. When a baseline year is selected with the academic achievement results for all students, the demographic factors such as lunch status, parents' level of education, ethnicity and special educational needs are correlated with achievement. However, when the students are followed though into the next academic year, the predictors for significant improvement become the individualized education programs, tutoring, and mental health services provided in the school. Findings at regular intervals from a 100% evaluation with the entire population helps to better target and develop the interventions in a longitudinal, continuous evaluation. This approach to realist evaluation can be achieved by human services utilizing the data they routinely collect, and the findings can have a high utility for improving the practice and impact of the

human services. However, it is important to recognize that this is a partnership between academics and the human services, helping these services to utilize the data in a suitably de-identified form, and not the usual approach of the academics taking the data away and producing reports after practice has moved on. This evaluation demonstrates that schools and human service agencies can use their data for evaluation and apply robust research methods to investigate their effectiveness at regular intervals, providing evidence in real-time to help to improve and to better target their programs of intervention.

Tapestry Evaluation in Partnership with Jamestown Public Schools: Utilizing whole-school data to investigate what works and for whom: Part 3

Universal Screening and Evaluating Services With School Outcomes

Strengths and Difficulties Questionnaire (SDQ) is a brief measure of pro-social behavior and psychopathology designed for 3-16 year-olds. It has very good reliability of 0.73 (Goodman, 2001). The SDQ has five factors: emotional symptoms, conduct problems, hyperactivity, peer relationships and pro-social items. In partnership with Tapestry, this universal screening tool has been used by Jamestown Public Schools as a teacher assessment for Grade levels K to 4 in April beginning 2012 & repeated in subsequent years. It was found that 20.6% had abnormal total scores on the SDQ in 2012 and 20.2% in 2013 (table 13), with teachers indicating in both years that conduct and hyperactivity were the leading factors contributing to these abnormal scores.

Table 13: SDQ total scores in 2012 and 2013

SDQ total score	2012	2012%	2013	2013 %
Abnormal	448	20.6	421	20.2
Borderline	288	13.3	246	11.8
Normal	1434	66.1	1414	67.9
Total	2170	100.0	2081	100.0

For the baseline year 2012, it was found that the predictors for abnormal scores on the SDQ were gender, lunch status, fathers' level of education and IEP (table 14). Boys were twice as likely to have an abnormal score than girls; students on free and/or reduced lunch were almost 3 times more likely to have abnormal scores; students with fathers who had an education less than school graduation were almost twice as likely to achieve abnormal scores; and finally, those with an IEP were 4 times more likely to achieve abnormal SDQ total scores than those without and IEP.

Table 14: Binary logistic regression model with SDQ total score abnormal or not

Variables in the Equation						
	B	S.E.	Wald	df	Sig.	Exp(B)
Step 4 ^d gender1(1)	.725	.122	35.351	1	.000	2.065
freeorreducedlunch(1)	.993	.149	44.166	1	.000	2.700
fatherHSgraduateornot(1)	.569	.140	16.539	1	.000	1.766
lep1(1)	1.432	.213	45.393	1	.000	4.188
Constant	-1.480	.229	41.958	1	.000	.228

When the teacher's SDQ assessments were compared from 2012 to 2013 (n = 1511 as no repeated SDQ for grade K and 4 in the following year), it was found that 36.5% of those who had an abnormal SDQ total score in 2012 had actually improved to a normal SDQ total score in 2013, but 17.7% had become worse (table 15).

Table 15: Change from 2012 to 2013 on the SDQ total scores

SDQtotnormalornot2012 * SDQ2013totnormal Crosstabulation					
			SDQ2013totnormal		Total
			not normal SDQ total score	normal SDQ total score	
SDQtotnormalornot2012	not normal SDQ total score	Count	318	183	501
		% within SDQtotnormalornot2012	63.5%	36.5%	100.0%
	normal SDQ total score	Count	179	831	1010
		% within SDQtotnormalornot2012	17.7%	82.3%	100.0%
Total	Count		497	1014	1511
	% within SDQtotnormalornot2012		32.9%	67.1%	100.0%

Impact of Tapestry partners' services on SDQ total score levels

It was found that out of the 2170 (grade levels K to 4) who had been screened in 2012, 104 had received services from the county mental health clinic, and of these 48.1% had abnormal SDQ total scores in 2012 when the services began. When the SDQ total scores were compared in the following year, it was found that there was a statistically significant difference between those receiving county mental health services and those that did not (table 16), with those receiving county mental health services improving on the SDQ total score by a mean of 2 points on the SDQ when compared with the rest who did not change on the SDQ scores on average. It was also found that whilst 45.5% of the students not receiving county mental health services

had improved on the SDQ total scores, a larger proportion (56%) of those receiving these services had improved.

Table 16: Difference in SDQ total scores between those who received County mental health clinic services and those that did not

Group Statistics					
JMHornot		N	Mean	Std. Deviation	Std. Error Mean
DifftotalSDQ20122013	Not JMH	1461	.00	6.597	.173
	JMH	50	-2.12	7.507	1.062

$T(1509) = 2.23, p < .05$

There was also a statistically significant difference between those receiving Striders' Elementary Programs and those that were not, with the Striders' recipients improving by a mean total score points of 1.3 as compared with none for the rest.

Table 17: Difference in SDQ total scores between those who received Striders' Elementary Programs those that did not

Group Statistics					
stridersselementary2012		N	Mean	Std. Deviation	Std. Error Mean
DifftotalSDQ20122013	not Striders Elementary	1399	.03	6.660	.178
	Striders Elementary	112	-1.30	6.235	.589

$T(1509) = 2.05, p < .05$

When the binary logistic regression model was used (table 18), it was found that the predictors for improvement on the SDQ total scores were baseline score levels and Striders. Those who had abnormal SDQ total scores in 2012 were 4 times more likely to improve their scores in the following year when compared to the rest who did not have abnormal scores. Similarly, those who received Striders' elementary programs were almost twice as likely to improve on the SDQ total score in the following year.

Table 18: Binary logistic regression model for improved or not on the SDQ total scores from 2012 to 2013

Variables in the Equation						
	B	S.E.	Wald	df	Sig.	Exp(B)
Step 2 ^b SDQtotnormalornot2012(1)	1.437	.143	100.815	1	.000	4.210
stridersselementary2012(1)	.535	.233	5.297	1	.021	1.708
Constant	-.651	.079	68.602	1	.000	.522

Conclusion

The main contributors to abnormal SDQ total score levels were hyperactivity and conduct subscales. The predictors for abnormal/borderline SDQ total scores were IEP, lunch status, gender, and father's level of education. There were significant improvements with Jamestown County Mental Health interventions, but the

predictors for SDQ total score improvement were Striders intervention and baseline abnormal/borderline scores. However, a higher proportion of those with abnormal scores in 2012 were in the group that received county mental services, and therefore that service is also a contributor for improving the scores in the following year.

Tapestry Evaluation in Partnership with Jamestown Public Schools: Utilizing whole-school data to investigate what works and for whom: Part 4

Tapestry Partners' Services are now Predictors for improved Behavior in Jamestown Public Schools

This evaluation is based on the total school population from grade levels K to 12 in the school academic year 2014-15, comparing changes in the school outcomes from the previous year as the baseline year.

Table 19: Ethnicity

	Frequency	Percent
Valid Asian	24	.5
Black African American	216	4.1
Hispanic	1023	19.4
Native American	23	.4
Mixed race	484	9.2
Pacific Islander	2	.0
White Caucasian	3497	66.4
Total	5269	100.0

The largest majority (66.4%) was white Caucasian, and the largest minority group was students of Hispanic origin (table 19). It was also found that 3524 (66.9%) of the students were described as from 'low income families'. Additionally, it was found that 739 (14%) of the students had an IEP, and 254 (4.8%) had a section 504 disability.

Interventions

All three interventions included in this evaluation were provided in the baseline school year (2013-14) and the changes in the outcomes were investigated in the following school year (2014-15) between those who received these interventions and those that did not, thereby providing evidence of effectiveness (tables 20 to 22). It was found that 309 (5.9%) school students had received county mental health clinic services, 97 (1.8%) had received mental health services from FSCR, and 1098 (20.8%) had attended received services at the Boys & Girls Club in the school year 2013-14 (including summer attendance).

Table 20: Receiving Active County Mental Health Services (2013-14)

	Frequency	Percent
Not county MH	4960	94.1
County MH	309	5.9
Total	5269	100.0

Table 21: Receiving Family Services of Chautauqua Region Services (2013-14)

	Frequency	Percent
Valid Not receiving FSCR services	5172	98.2
Receiving FSCR services	97	1.8
Total	5269	100.0

Table 22: Receiving Boys & Girls Club Services (2013-14)

	Frequency	Percent
Valid Not Boys and Girls Club	4171	79.2
Boys and Girls Club	1098	20.8
Total	5269	100.0

Table 23: SDQ Total Difficulties (2014-15)

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Abnormal	426	8.1	20.6	20.6
Borderline	269	5.1	13.0	33.6
Normal	1373	26.1	66.4	100.0
Total	2068	39.2	100.0	
Missing System	3201	60.8		
Total	5269	100.0		

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Abnormal	426	8.1	20.6	20.6
	Not abnormal	1642	31.2	79.4	100.0
	Total	2068	39.2	100.0	
Missing	System	3201	60.8		
Total		5269	100.0		

Outcomes

This report is based on two school outcomes: the use of Strengths and Difficulties Questionnaire (SDQ) as a universal screening tool completed by teachers for elementary school students in grade levels K to 4, and the recording of behavior incidents in all the schools in JPS in the two school years. The SDQ is used repeatedly in April of each year, and in 2014-15 it was found that 426 (20.6%) were in the abnormal range, mainly due to hyperactivity and conduct subscales within the SDQ (table 23).

It was found that there were significant patterns in the correlations between the SDQ total difficulties (abnormal or not) and the demographics of the school students. Higher proportions of school students from low income families (table 24), mixed race (table 25), IEP (table 26) and Section 504 disabilities (table 27) were found to be in the abnormal SDQ total scores.

Table 24: Cross tabulation between low income and SDQ total score abnormal or not

			Total abnormal or not		
			Abnormal	Not abnormal	
Stateprogram1415poverty	Not from low-income family	Count	48	492	540
		% Within stateprogram1415poverty	8.9%	91.1%	100.0%
	Poverty - from low-income family	Count	378	1150	1528
		% Within stateprogram1415poverty	24.7%	75.3%	100.0%
Total		Count	426	1642	2068
		% Within stateprogram1415poverty	20.6%	79.4%	100.0%

Table 25: Cross tabulation between Ethnicity and SDQ total score abnormal or not

			Total abnormal or not		Total
			Abnormal	Not abnormal	
Ethnicity1	A	Count	2	8	10
		% Within Ethnicity1	20.0%	80.0%	100.0%
	B	Count	15	54	69
		% Within Ethnicity1	21.7%	78.3%	100.0%
	H	Count	88	348	436
		% Within Ethnicity1	20.2%	79.8%	100.0%
	I	Count	1	5	6
		% Within Ethnicity1	16.7%	83.3%	100.0%
	M	Count	49	157	206
		% Within Ethnicity1	23.8%	76.2%	100.0%
	P	Count	0	2	2
		% Within Ethnicity1	.0%	100.0%	100.0%
	W	Count	271	1068	1339
		% Within Ethnicity1	20.2%	79.8%	100.0%
Total		Count	426	1642	2068
		% Within Ethnicity1	20.6%	79.4%	100.0%

Table 26: Cross tabulation between IEP and SDQ total score abnormal or not

			Total abnormal or not		Total
			Abnormal	Not abnormal	
IEP1	Not IEP	Count	338	1525	1863
		% within IEP1	18.1%	81.9%	100.0%
	Has IEP	Count	88	117	205
		% within IEP1	42.9%	57.1%	100.0%
Total		Count	426	1642	2068
		% within IEP1	20.6%	79.4%	100.0%

Table 27: Cross tabulation between Section 504 and SDQ total score abnormal or not

			Total abnormal or not		Total
			Abnormal	Not abnormal	
Section504a	Not section 504	Count	412	1616	2028
		% Within Section504a	20.3%	79.7%	100.0%
	Has section 504	Count	14	26	40
		% Within Section504a	35.0%	65.0%	100.0%
Total		Count	426	1642	2068
		% Within Section504a	20.6%	79.4%	100.0%

Table 28: Change in Total Difficulties levels from the SDQ in 2013-14 and in 2014-15

SDQTotalscore2014and15Change					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Got worse	643	12.2	43.7	43.7
	No change	176	3.3	12.0	55.6
	Improved	653	12.4	44.4	100.0
	Total	1472	27.9	100.0	
Missing	System	3797	72.1		
Total		5269	100.0		

When the SDQ total difficulties' score levels were compared between the school years 2013-14 and 2014-15 (table 28), it was found that 44.4% had improved but in the same period a similar proportion had become worse. However, cross tabulations between those receiving the three services and those that were not found that more improvements were made by elementary school students that were receiving mental health services from county mental health and FSCR when compared with those not receiving these services. However, the proportions were similar for those who attended Boys and Girls Club and those that did not (tables 29-31).

Table 29: Change in Total SDQ and FSCR Services

			SDQTotalImprovedornot14and15		Total
			Not improved	Improved	
FSCRServicesto2014	Not receiving	Count	807	639	1446
	FSCR services	% Within FSCRServicesto2014	55.8%	44.2%	100.0%
	Receiving	Count	12	14	26
	FSCR services	% Within FSCRServicesto2014	46.2%	53.8%	100.0%
Total		Count	819	653	1472
		% Within FSCRServicesto2014	55.6%	44.4%	100.0%

Table 30: Change in Total SDQ and Boys & Girls Club Services

			SDQTotalImprovedornot14and15		Total
			Not improved	Improved	
BoysandgirlsclubAug2014	Not	Count	630	505	1135
	Boys and Girls Club	% Within BoysandgirlsclubAug2014	55.5%	44.5%	100.0%
	Boys and Girls Club	Count	189	148	337
		% Within BoysandgirlsclubAug2014	56.1%	43.9%	100.0%
Total		Count	819	653	1472
		% Within BoysandgirlsclubAug2014	55.6%	44.4%	100.0%

Table 31: Change in Total SDQ and County Mental Health Services

		SDQ Total Improved or not 14 and 15		Total
		Not improved	Improved	
activeCountyMH2011314 Not county MH	Count	802	636	1438
	% Within activeCountyMH2011314	55.8%	44.2%	100.0%
County MH	Count	17	17	34
	% Within activeCountyMH2011314	50.0%	50.0%	100.0%
Total	Count	819	653	1472
	% Within activeCountyMH2011314	55.6%	44.4%	100.0%

Table 32: Change in Recorded Incidents of Behavior

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Count Incidents 2013-14	1.29	5269	4.996	.069
	Count Incidents 2014-15	1.27	5269	4.62785	.06376

Change in Incidents 2013-14 and 2014-15

	Frequency	Percent
Got worse	683	13.0
No change	3848	73.0
Improved	738	14.0
Total	5269	100.0

Behavior Incidents

When the recorded incidents of behavior were compared between the school years 2013-14 and 2014-15, it was found that the overall mean of 1.3 remained almost the same in both years (table 32). Although 14% of the school students did improve on the number of recorded incidents, 13% actually became worse. It was found that there were significant correlations between improvement in behavior incidents and

grade level, ethnicity, IEP, Section 504 disability, and all three interventions.

When the binary logistic regression was used with those who had improved and those who had not (tables 33-38), it was found that the predictors were 504 disabilities, low income, grade level and the three services. Students from low income families were almost 3 times more likely to improve; and those who received services from county mental health, Boys & Girls Club and FSCR were almost twice as likely to improve on incidents of behavior than those not receiving these services. Additionally, it was found that higher-grade levels were associated with improvement.

Table 33: Binary Logistic Regression with behavior incidents improved or not

		Variables in the Equation					
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 6 ^f	Section504a(1)	.407	.163	6.244	1	.012	1.502
	stateprogram1415poverty(1)	1.055	.113	87.797	1	.000	2.871
	activeCountyMH2011314(1)	.563	.139	16.351	1	.000	1.756
	BoysandgirlsclubAug2014(1)	.541	.099	29.602	1	.000	1.718
	FSCRServicesto2014(1)	.567	.242	5.486	1	.019	1.763
	Grade Level	-.293	.014	427.733	1	.000	.746
	Constant	1.709	.325	27.576	1	.000	5.525

Table 34: Section 504 Disabilities and Improvement in Behavior Incidents

			Incidents improved or not 14 and 15		Total
			not improved	improved	
Section 504a	not section 504	Count	4342	673	5015
		% within Section 504a	86.6%	13.4%	100.0%
	has section 504	Count	189	65	254
		% within Section 504a	74.4%	25.6%	100.0%
Total		Count	4531	738	5269
		% within Section 504a	86.0%	14.0%	100.0%

Table 35: Low income and improvement in Behavior Incidents

			Incidents improved or not 14 and 15		Total
			not improved	improved	
stateprogram1415poverty	Not from Count		1626	119	1745
	low- income family	% within stateprogram1415poverty	93.2%	6.8%	100.0%
	Poverty - from low- income family	Count % within stateprogram1415poverty	2905	619	3524
			82.4%	17.6%	100.0%
Total	Count		4531	738	5269
	% within stateprogram1415poverty		86.0%	14.0%	100.0%

Table 36: County Mental Health and improvement in Behavior Incidents

			Incidents improved or not 14 and 15		Total
			not improved	improved	
activeCountyMH2011314	not county MH	Count % within activeCountyMH2011314	4325	635	4960
			87.2%	12.8%	100.0%
	County MH	Count % within activeCountyMH2011314	206	103	309
			66.7%	33.3%	100.0%
Total	Count		4531	738	5269
	% within activeCountyMH2011314		86.0%	14.0%	100.0%

Table 37: Boys & Girls Club and improvement in Behavior Incidents

			Incidents improved or not 14 and 15		Total
			not improved	improved	
Boys and Girls Club Aug 2014	not receiving	Count	3652	519	4171
	FSCR services	% within FSCR Service to 2014	87.6%	12.4%	100.0%
Boys and Girls Club Aug 2014	Receiving	Count	879	219	1098
	FSCR services	% within FSCR Service to 2014	80.1%	19.9%	100.0%
Total			4531	738	5269
			86.0%	14.0%	100.0%

Table 38: FSCR and improvement in Behavior Incidents

			Incidents improved or not 14 and 15		Total
			not improved	improved	
FSCR Service to 2014	not receiving	Count	4467	705	5172
	FSCR services	% within FSCR Service to 2014	86.4%	13.6%	100.0%
FSCR Service to 2014	Receiving	Count	64	33	97
	FSCR services	% within FSCR Service to 2014	66.0%	34.0%	100.0%
Total			4531	738	5269
			86.0%	14.0%	100.0%

It was found that 33.3% of the JPS students receiving county mental health clinic services improved on the behavior incidents, as compared with 12.8% of the others not receiving these services (table 36). It was also found that 19.9% of those that had attended events at the Boys & Girls Club improved as compared with 12.4% of those that had not attended these events (table 37). The pattern with FSCR was similar as

the county mental health—34% of those receiving FSCR’s mental health services improved as compared with 13.6% of the rest (table 38).

Conclusion: Tapestry Partners’ Services are Predictors for Improved Behavior

This evaluation is based on findings from the universal screening (SDQ) with grade levels K to 4, and the changes in the incidents of behavior between the school years 2013-14 and 2014-15. It was found that there was some evidence of effectiveness in improving the SDQ abnormal total difficulties score with the mental health services in relation to both county mental health and FSCR. With regard to changes in reported incidents of behavior, all three services were predictors for improvement from one school year to the next. In relation to each of the three services—county mental health, Family Services of Chautauqua Region and the Boys & Girls Club—it was found that those receiving these services were almost twice as likely to improve on the incidents of recorded behavior than those not receiving these services.

The examples of evaluation described here provide the evidence of how Chautauqua County is a world leader in the regular analysis of data already collected by human services, and combined with outcome measures and school outcomes from the whole school data, are helping to improve and develop more effective services. That is why we received SAMHSA’s first Gold Award for Outstanding Local Evaluation in 2010. Since 2008, the Evaluation Team (with Dr. Kazi as the lead evaluator) has been committed to developing a rigorous evaluation of the Tapestry Expansion Initiative that is consistent with the principles of ethical and culturally sensitive human service research. This evaluation strategy is based on Tapestry 2.0’s overall aim of moving the system of care towards wide scale adoption, and as one of the 27 rural counties in New York State, Tapestry 2.0 will demonstrate how the system of care can be developed from innovation through to wide-scale adoption.

The purpose of this program is to improve mental health outcomes for children and youth (4 to 21 years of age) with serious emotional disturbances (SED) and their families. The program will support the wide scale operation, expansion and integration of the system of care. SAMHSA’s Theory of Change framework is organized into **five phases**: Innovation, Translation, Dissemination, Implementation, and Wide-Scale Adoption. Each phase encompasses a range of strategies, activities, programs, and tasks, which pave the way toward strategic and evidence-based behavioral health system change. Surveillance and evaluation underpin each element of this process, and in Chautauqua County through the work of Chautauqua Tapestry since 2008, there has been a strong emphasis on evaluation, helping participating agencies to use the data they already collect in their management information systems to investigate what works and in what circumstances at regular intervals during the year, and to use the findings from evaluation to inform the development and expansion of the system of care. This evaluation strategy has become a component part of Tapestry 2.0’s system of care, and wide scale adoption also means the wide scale adoption of this evaluation strategy. At present, Chautauqua County is the only one in the world that combines all human service data regularly with all school district data, using robust evaluation research methods in partnership with each agency to investigate what works and for whom at quarterly intervals in each year.

Evaluation has been a stand out point for Tapestry, and will continue to do so for Tapestry 2.0. The strategy that is used in Chautauqua is Realist Evaluation, as applied in schools, mental health, youth justice and other human service agencies. One of the first adopters was Jamestown Public Schools (JPS), the largest school district in the county. Studies have indicated that school interventions can be utilized to make numerous improvements in the overall well being of students (e.g. Goldenson, 2011). In order to evaluate interventions, most researchers utilize pre/posttest designs, looking at the initial baseline levels of the targeted behavior before and after the intervention is implemented. Chautauqua Tapestry's evaluation team has been utilizing whole-school data for evaluation. Kazi (2014) evaluated school-based programs in Jamestown, NY using whole-school data, comparing academic outcomes of those receiving interventions with those who were not. It was found that, although Caucasian children and youth achieved better academic grades compared to Hispanic and African Americans, the binary logistic regression indicated that the actual predictors for improved academic achievement were the tutoring and mental health services as well as individualized education programs. It was found that those receiving these interventions had significantly improved their Grade Point Average (GPA) scores when compared to those that had not received these interventions. The data analysis helps school administration to identify gaps in care, and helped to build partnerships for additional mental health counseling on-site for the school district.

At this time, 7 out of 18 school districts have participated in local evaluation. A very important additional method has come out of the work that would be carried out through the four year Expansion, namely ICE-8. This new all County program has been signed on by all the school districts in Chautauqua County under the joint direction and support of Tapestry and of Erie 2 BOCES. Evaluation is an integral component of the ICE-8 initiative, and as a result of this SOC expansion grant, all 18 school districts will be invited to take part in this evaluation in the same way as JPS has been. Strategy for expansion will be to utilize the newly developed Chautauqua Evaluation and Translational Research innovation Center (CETRiC) to build the necessary champions within each organization and to develop learning communities to support specific learning goals.

Plan for Chautauqua Tapestry X 2 (2015-2019)

- 1. To date, Chautauqua County is the only one to receive SAMHSA's Gold Award for Outstanding Local Evaluation. An integral part of the plan is to expand this work to include all school districts and human service agencies in the county.** Chautauqua County is a world-leader in realist evaluation, and to date 28 agencies (including 7 school districts) have collaborated in applying robust research methods to the data they already collect to investigate what works and for whom.
2. The evaluation team will work in partnership with each participating agency to undertake regular analysis of their data every three months, and to prepare reports on the findings to help better target and to develop the human services.
3. Data from the participating agencies' management recording systems, including all 18 school districts, will be accessed to analyze what interventions work and in what circumstances, every three months. The outcomes would include all school outcomes (academic outcomes, attendance and reported incidents of behavior), the Strengths and Difficulties Questionnaire (SDQ) universal screening in schools,

and other standardized measures selected by the agencies and/or required for the SAMHSA grant. Data analysis will be carried out by the evaluation team with each agency, and then reported by the agencies with the participation of the evaluation team to the governance body and the advisory committee in each three-month period. In this way, all the data will become an important part of the Continuous Quality Improvement process.

4. The Evaluation Team is committed to utilizing a collaborative, empowerment approach with an advisory committee comprised of families and youth from within the County system of care. Through proactive collaboration with *Tapestry's* cultural competence consultant, as well as members of the evaluation team serving on the cultural competence committee, it is intended that representatives on the evaluation advisory committee will represent all ethnic groups served, as well as be inclusive of other groups served (e.g., sexual orientation, gender, rural). The evaluation will be based on consumer perspectives as well as evidence of the impact of the services on the needs and aspirations of the service recipients.
5. Sustainability is an important goal for the Evaluation team. This will be achieved through Chautauqua Evaluation and Translational Research innovation Center (CETRiC) which will strengthen the partnership of all Tapestry partners (including United Way and Chamber of Commerce) in integrating this evaluation strategy into the daily practice of participating human service agencies, as well as enabling agency partners to apply for other sources of funding. The Evaluation team will assist each agency to enhance its own capacity to sustain evaluation, undertaking regular analyses of the data from their management information systems, to investigate what works and for whom in real time.

References

Antle, B., Barbee, A. P., Christensen, D. N., Martin, M. (2008) "Solution-Based Casework in Child Welfare: Preliminary Evaluation Research." *Journal of Public Child Welfare* 2(2): 197-227.

Berg, I. K. & Kelly, S. (2000) *Building Solutions in Child Protective Services*. New York, NY US: W W Norton & Co.

Christensen, D. & Antle, B. (2009) "Engaging Child Welfare Families: A Solution-Based Approach to Child Welfare Practice." University of Louisville: Center for Family Resource Development.

Goldenson, J. (2011). When there is no blueprint: The provision of mental health services in alternative school programs for suspended and expedited youth. *Child & Youth Services*, 32(2): 108-123.

Goodman, R. (2001) 'Psychometric Properties of the Strengths and Difficulties Questionnaire (SDQ)', *Journal of the American Academy of Child and Adolescent Psychiatry*, 40, pp. 1337-1345.

Kazi, M.A. F. (2003). *Realist Evaluation in Practice*, London: Sage Publications.

Kazi, M. A. F. (2014) "Realist Evaluation and Effectiveness Research: An Example from School-based Interventions" in Kalliola, S. (ed.) 'Evaluation as a Tool for Research, Learning and Making Things Better', Newcastle-Upon-Tyne, UK: Cambridge Scholars Publishing, pp. 31-44.

Kazi, M. A. F., Coombes, M., Krause, D., & Anderson, J. (2015). Effects of Solution-Focused Training on the Permanency Outcome in Child Welfare: A Realist Evaluation. *International Journal of Interdisciplinary Social and Community Studies*, 10(4), pp. 27-39.

Kazi, M. A. F. & Kim, Y. (2015). Realist Evaluation of What Programs Work, for Whom and in What Circumstances: Utilizing Whole School Data to Improve Student Outcomes. *Journal of Academic Perspectives* (accepted May 2015, Oxford University).

Kutash, K., Duchnowski, A. J., & Lynn, N. (2006). School-based mental health: An empirical guide for decision makers. Tampa, FL: University of South Florida, Research and Training Center for Children's Mental Health.

Leon, SC, Lyons, JS, Uziel-Miller, ND, Tracy, P. (1999). Psychiatric hospital utilization of children and adolescents in state custody. *Journal of the American Academy of Child and Adolescent Psychiatry*, 38, 305-310.

New York State Office of Children and Family Services. (2010) Cognos Admissions Report 2009 Chautauqua. GPO.

Reback, R. (2010). Schools' mental health services and young children's emotions, behavior, and learning. *Journal of Policy Analysis and Management*, 29(4), 698-725.

Videka, L. (2003). 'Accounting for variability in client, population, and setting characteristics: moderators of intervention effectiveness', in Rosen, A. & Proctor, E. K. (eds.) *Developing Practice Guidelines for Social Work Intervention*, New York: Columbia University Press, p. 169-192.