CCRY Research Initiative Results

Introduction

This report is a continuation of the first report, CCRY Research Initiative Report (Appendix), which outlined the background and initial steps in the initiative. These initial steps included outlining the goals of the initiative, identifying the common data elements across agencies, obtaining the data elements identified, and some discussion around data translation and consolidation into a unified dataset. This report will continue a discussion:

Section D: Data translation and consolidation issues
Section E: Data analysis of combined dataset to answer the goals of the report
Section F: Feedback on the research process

As a continuation of the previous report, it continues from subsection D of the last report, looking at data translation and combining into a unified dataset for analysis.

D. Data translation and combining into a unified dataset for analysis

As discussed in the previous report, there were three datasets that were obtained from the three participating communities. The data in each of the datasets was built around a list of common data elements identified by the Initiative.

The previous report provided examples of issues faced in combining the data. It was a very time intensive activity and required around 100 hours to complete the data combination. There were three main issues identified in the combination process:

1. Understanding what variables in each of the three community datasets matched the common variables identified across all three communities. In some cases it was obvious what the matches were, for example, demographic variables tended to match up. However some variables did not match up cleanly and needed further clarification. In those cases, each community liaison was asked to clarify how their variables lined up with the common variables.

2. It was often the case, especially for the "time spent" variables, that there were multiple variables in each of the individual community datasets that contributed to one of the variables in the common list of variables. Once it was determined which set of variables contributed to the one common variable, the time spent in each of the separate variables were added together. For example, if the common variable was "Math Remediation" and the Hartford community had variables "Math Program 1" "Math Program 2" and "Math Program 3", the time spent in each of the three math programs for that individual were added together to get a total time spent for that individual in Math Remediation. This was particularly more difficult to do for the Hartford and Philadelphia sites because each record was a point-of-service, indicating the

time spent in a program. Records had to sorted by each program, and then by individuals, and then calculate a score for time spent for each common variable.

3. Data from each of the communities were split up across several Excel spreadsheets. Some amount of looking-up needed to be done in order to consolidate all the data for each of the individuals.

For future projects of this nature, there needs to be a significant amount of time and resources put into understanding each contributing communities datasets, how the data is presented, and also allocating sufficient resources to consolidating datasets as it is a very time intensive activity.
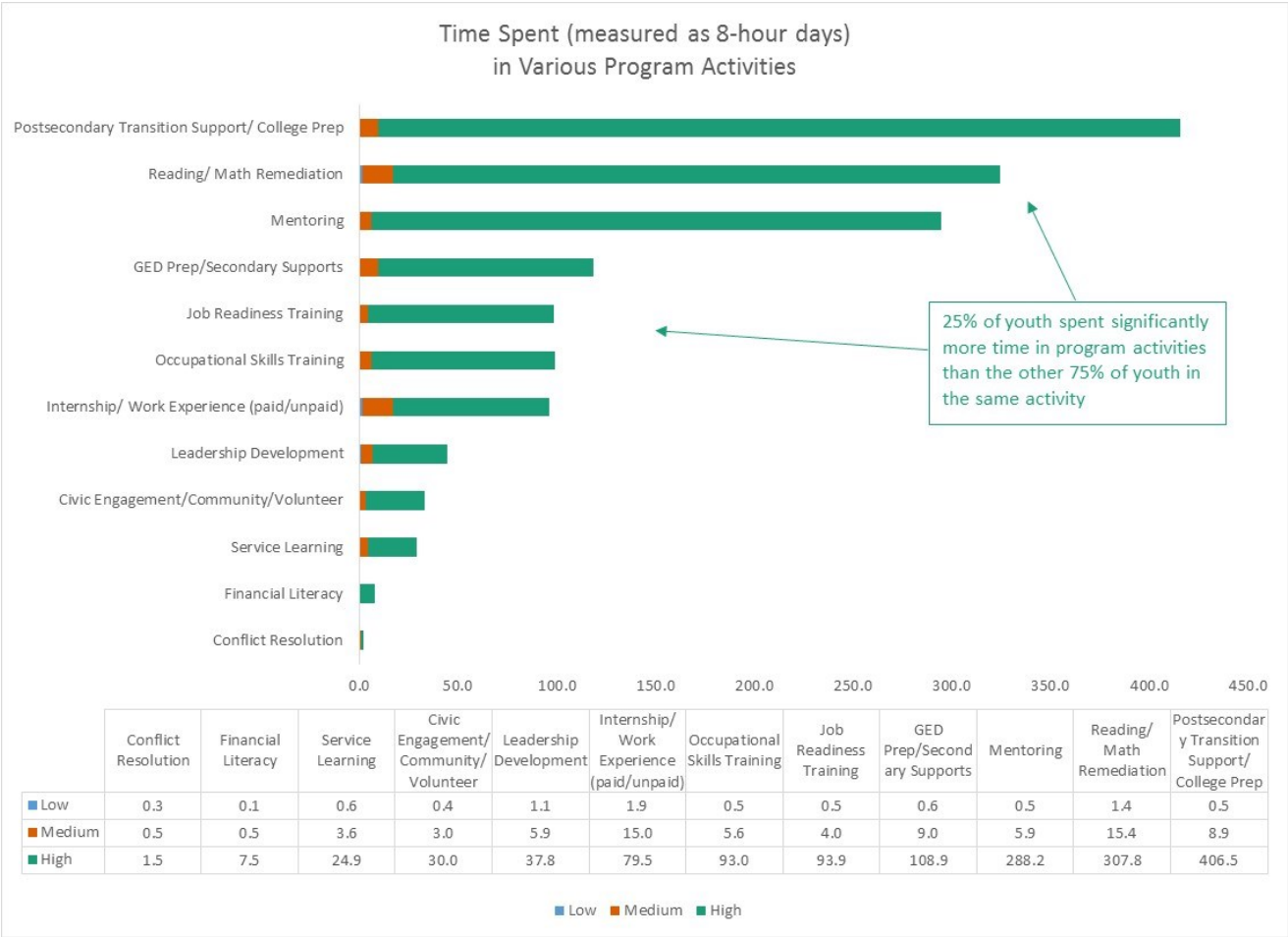
E. Data analysis of combined data set to answer the goals of the report

Demographics

The final combined dataset had 7,105 youth entries. From that total, 1,971 were from Baltimore, 4,381 were from Philadelphia, and 753 from Hartford. Note that while this is the grand total, there were missing data across the datasets so some of the other results presented will not reflect this grand total. The dataset was evenly split between males and females, with 2,823 males (51%) and 2,745 females (49%). The youth were predominantly African American, making up 84% of youth that had ethnicity data. The next largest group was Hispanic (8%) and then White (2%). The majority of these youth were also out-of-school youth that dropped out of high school (86%), were not in foster care (85%), were not homeless or living in a shelter (97%), did not have a mental or physical disability (94%), were not juvenile justice or court involved youth (87%), or have any major barriers (such as child/dependent care, medical insurance, transportation, housing, etc.).
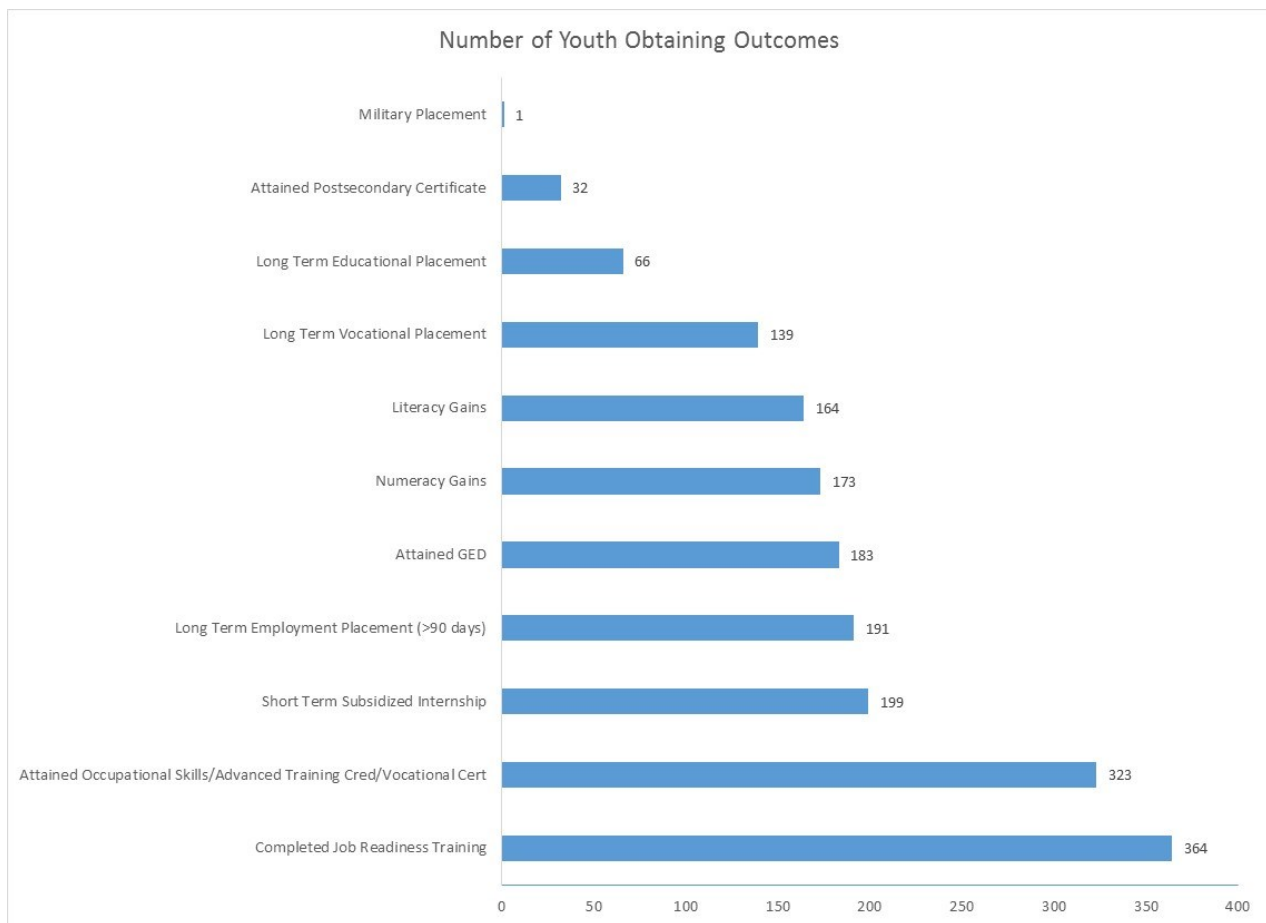
Time Spent

One group of variables looked at the time spent in various program activities. The dataset had information for time spent on various program activities for 3,355 youth. The time spent was measured by the number of minutes each youth spent in total for that program activity. For each of the program activities, the time spent was re-classified into 3 categories, **low**, **medium**, and **high**. Each of the categories represented a quartile range of time spent. So for example, the **low category** for time spent doing math and reading remediation program activities means that 25% of youth spent 670 minutes or less doing math and reading remediation. In other words, if a youth spent 670 minutes or less doing math and reading mediation, they were classified as low. The **medium category** for math and reading remediation is from 25% to 75% of youth in that category and equaled to the time spent from 671 minutes to 7,410 minutes in that program activity. Therefore if a youth spent between 671 minutes to 7,410 minutes doing math or reading remediation, they were classified as medium. And finally, for high, it would be those above the 75% quartile, or those who spent more than 7,410 minutes doing math or reading remediation. This method of classification is true for all time spent program activity measures. The chart on the next page shows the ranges of time spent for each of the program activities.

Time Spent (measured as 8-hour days) in Various Program Activities

> 25% of youth spent significantly more time in program activities than the other 75% of youth in the same activity

	Conflict Resolution	Financial Literacy	Service Learning	Civic Engagement/ Community/ Volunteer	Leadership Development	Internship/ Work Experience (paid/unpaid)	Occupational Skills Training	Job Readiness Training	GED Prep/Second ary Supports	Mentoring	Reading/ Math Remediation	Postsecondar y Transition Support/ College Prep
Low	0.3	0.1	0.6	0.4	1.1	1.9	0.5	0.5	0.6	0.5	1.4	0.5
Medium	0.5	0.5	3.6	3.0	5.9	15.0	5.6	4.0	9.0	5.9	15.4	8.9
High	1.5	7.5	24.9	30.0	37.8	79.5	93.0	93.9	108.9	288.2	307.8	406.5

■ Low ■ Medium ■ High

As the chart shows, the top 25% of youth are spending significantly more time in program activities than the lower 75% of youth for that same program activity. The top 25% of youth also put in more time into academic activities related to remediation and postsecondary transition, as well as, into mentoring. There is a moderate amount of time devoted to workforce development (Job Readiness Training, Occupational Skills Training, and Internship/Work Experience) as well as GED Prep.

Outcomes

The other group of variables was related to outcomes. From the dataset, there was information on outcomes for 1,353 youth. The majority of youth in the dataset just had one outcome (66%), followed by youth with 2 outcomes (25%), and a smaller amount of youth with 3 or more outcomes (9%). The chart below shows the number of youth obtaining different outcomes. The chart on the next page shows the number of youth that obtained the different outcomes.

Number of Youth Obtaining Outcomes

Outcome	Number
Military Placement	1
Attained Postsecondary Certificate	32
Long Term Educational Placement	66
Long Term Vocational Placement	139
Literacy Gains	164
Numeracy Gains	173
Attained GED	183
Long Term Employment Placement (>90 days)	191
Short Term Subsidized Internship	199
Attained Occupational Skills/Advanced Training Cred/Vocational Cert	323
Completed Job Readiness Training	364

How Is Time Spent in Various Program Activities Related to Outcomes?

We were interested in answering the question, does the amount of time a youth spends doing a particular program activity related to getting a certain outcome. In order to answer this question, simple crosstabs were constructed looking at the categorical time spent in each program activity (low, medium, or high) and outcomes. After running all of the permutations of time spent per program activity by each outcome, particular emphasis was given to the rates of youth that spent a high amount of time in a program activity and had higher rates of outcomes than either the medium or low categories. For example, for the outcome "attained occupational skill/advanced training credential/vocational certificate" 65 youth (40% of all youth that achieved this outcome) spent a "high" amount of time in leadership development, whereas only 46 youth (29%) spent a "medium" amount of time, and 50 youth (31%) spent a "low" amount of time doing leadership development program activities. Therefore youth who spent a lot of time doing leadership development program activities were more likely to get to attain an occupational skill/advanced training credential/vocational certificate than those who did not.

Using this analysis, three distinct patterns emerged when thinking about what types of program activities are related to outcomes.

First, reading and math remediation are important to workforce development outcomes. Youth who spent a high amount of time in math/reading remediation were more likely to complete a job readiness

training (45%), complete a subsidized internship (46%), and obtain occupational skill/advanced training credential/vocational certificate (51%).

Second, job readiness training lays a strong foundation for obtaining long term outcomes. Youth who spent a high amount of time in job readiness training were more likely to attain a high school diploma (77%), have long-term employment placement of at least 90 days (39%), and a long-term advanced training placement/vocational placement.

Third, youth who spent a high amount of time in postsecondary transition support/college prep was related to completing job readiness training (45%) and attaining a GED (51%).

F. Feedback on the research process

One of the big questions that this research initiative attempted to answer was, would it be possible to combine datasets from different communities in order to draw some conclusions about what youth development activities are successful in leading to outcomes. This report shows some preliminary findings of a combined dataset among three different communities and provides some evidence that a collective dataset across different communities can wield meaningful results. These results should be viewed as preliminary, highlighting patterns in the data, rather than been viewed as definitive. A much more rigorous design would be needed in order to produce more definitive, robust results.

However, as a learning process, this initiative also highlighted some of the challenges faced in attempting to undertake this research project. Several of these are discussed below:

1. At the very outset, one of the big questions is, what questions are we trying to answer? An additional related question is, what are the common data elements between the different communities datasets to answer this question. Much more time needs to be spent understanding how each of the communities programs work and in addition understanding how the data recorded in each of these communities is related to the programs. This research made big assumptions that because data elements were labeled in a similar way, it implied that they could be aggregated across communities. Having a clear understanding of the questions that are trying to be answered, the program activities in each community and the related data for each activity will be helpful in the later aggregation and data analysis, as well as provide more reasonable assumptions about what data elements can and cannot be combined.
2. Understanding how the data is structured in each of the communities and how those data elements related to each other across communities. For this project we started off with some identified universal data elements and then tried to match each communities specific data elements to the universal ones. This process needs to be much more deliberate, both to understand how multiple data elements from one community might be contributing to a single universal variable, but also to understand what data elements are missing from each of the communities. If one community is unable to provide data in the required format, then it reduces the ability to gain meaningful results in the end.
3. Significant investment of resources is important from the start of the project, to formulate the research questions, understanding communities programs, understanding the data in each community, combining data across communities, analyzing the data and writing a report. As we

saw from this project, there was significant amount of time required simply to put the data together. To do another project of this size or larger well, there would need to be larger investments of time and resources at all stages of the research.

4. Coordination across communities is also important. This includes identifying the correct staff among the various communities that will be able to dedicate some time to the project, having regular meetings to drive the project forward, having a centralized space (probably online) to coordinate activities, a system of checks and balances along the way to ensure that the data is handled correctly, and results are being interpreted accurately.

This research initiative provides promise that the collective power of CCRY can create meaningful reports that can inform a wide variety of audiences about youth development and helping youth achieve meaningful outcomes.

Report prepared by Sean Seepersad, PhD
Senior Director of Research and Organizational Performance
Our Piece of the Pie, Inc.
20-28 Sargeant St
Hartford, CT 06070

Questions or comments about this report can be directed to: sean.seepersad@opp.org

Appendix

CCRY Research Initiative Report

Background

The Communities Collaborating to Reconnect Youth (CCRY) Network is a group of workforce and youth development professionals, working in communities across the country.[1] One of the questions that the CCRY Network is trying to answer is; can the research efforts by various agencies within the Network be combined to provide some degree of evidence on what type of programming is effective in producing positive outcomes for out of school youth. In an attempt to answer this question, the Network started a research initiative with agencies in three metropolitan areas in the Network: Baltimore (*Mayor's Office of Employment Development*), Hartford (*Our Piece of the Pie* and *Capital Workforce Partners*), and Philadelphia (*Philadelphia Youth Network*). These areas and agencies were chosen because they all have some research and evaluation capacity, thus have datasets that can be used in an analysis to answer questions about demonstrated effectiveness of youth programming.

The research initiative started on April 23, 2012 with a conference call of representatives from the various agencies along with additional CCRY leadership. The main purpose of the initial conference call was to clarify the goals of the initiative as well as determine the next steps forward. The end deliverable of the research initiative would be a report that would outline the steps of the research initiative along with any findings that may be obtained from the combined datasets of the various participating agencies.

In proceeding forward, the research initiative undertook the following broad steps:

A. Creating a shared understanding of the goals of the report to be produced by the initiative
B. Identifying what are the common data elements across agencies that would fulfill these goals
C. Obtaining the data elements identified
D. Data translation and combining into a unified dataset for analysis
E. Data analysis of combined data set to answer the goals of the report
F. Producing the research initiative report for review and feedback by participating agencies and CCRY leadership
G. Report dissemination

A. Creating a shared understanding of the goals of the report to be produced by the initiative

The research initiative group met several times to discuss what should be the goals of the initiative and the final report. Three goals were identified:
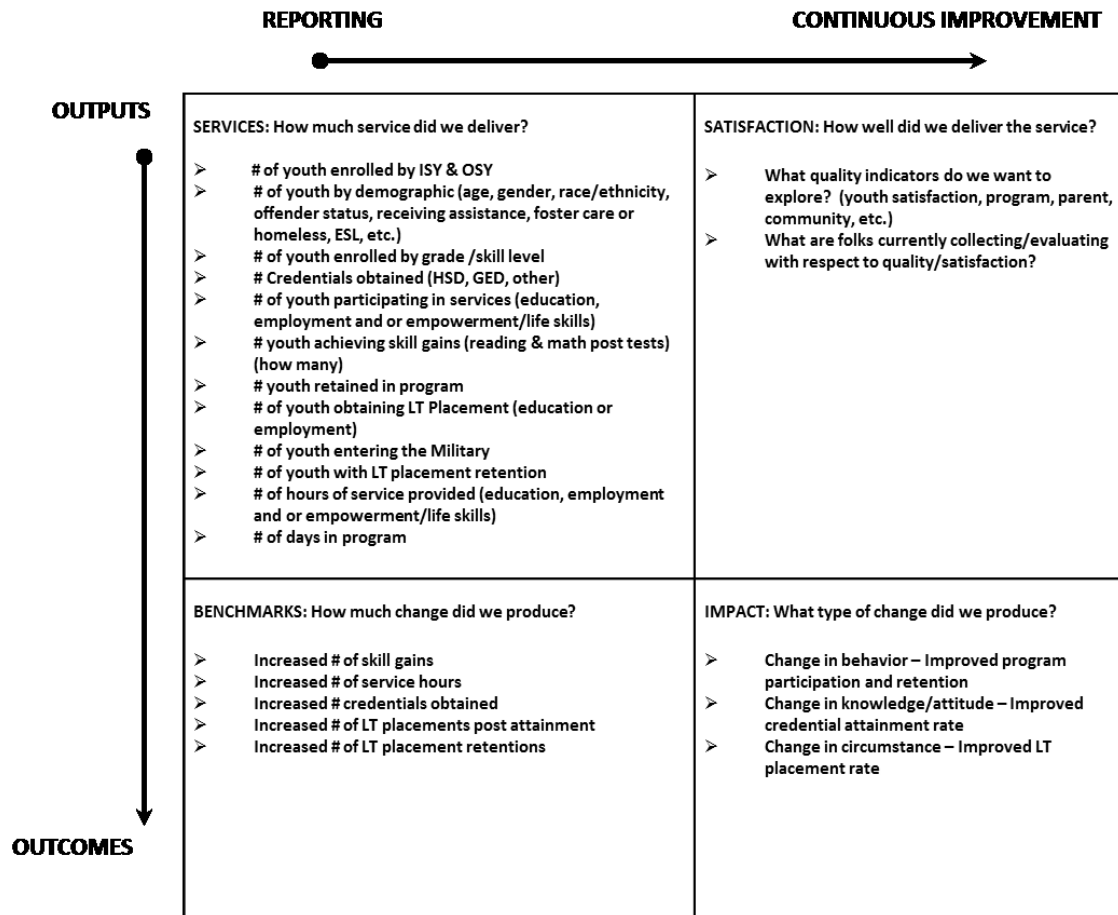
---

[1] http://www.ccrynetwork.org/about

1. To outline the process of data extraction and compilation across various agencies to answer assorted research questions, such as demonstrated effectiveness.
2. To provide a profile of the youth served, programmatic efforts, and outcomes achieved by various agencies.
3. To determine and document any demonstrated effectiveness that can be obtained by looking at the combined datasets of participating agencies.

B. Identifying what are the common data elements across agencies that would fulfill these goals

The next step of the initiative was to determine what data elements to be obtained from each agency. The group initially started with an ideal list of data elements to be obtained from each agency. For example, with regards to Goal 2 (an agency profile), the group identified questions such as: average number of service hours required to gain a skill, and average age of youth served. One of the tools we used to help us think about what the ideal list of data elements should be was Results Based Accountability (RBA). RBA allows for understanding of the types of data needed to accomplish the goals of an initiative. In using RBA, we considered questions around the RBA topics of Effort and Effect. Effort asks questions about how much effort was expended and how well was that effort delivered. Effect asks questions about how much change was accomplished and how well was the type of change produced. Below is a diagram showing the types of questions generated using this method.

REPORTING → CONTINUOUS IMPROVEMENT

OUTPUTS

SERVICES: How much service did we deliver?	SATISFACTION: How well did we deliver the service?
➢ # of youth enrolled by ISY & OSY ➢ # of youth by demographic (age, gender, race/ethnicity, offender status, receiving assistance, foster care or homeless, ESL, etc.) ➢ # of youth enrolled by grade /skill level ➢ # Credentials obtained (HSD, GED, other) ➢ # of youth participating in services (education, employment and or empowerment/life skills) ➢ # youth achieving skill gains (reading & math post tests) (how many) ➢ # youth retained in program ➢ # of youth obtaining LT Placement (education or employment) ➢ # of youth entering the Military ➢ # of youth with LT placement retention ➢ # of hours of service provided (education, employment and or empowerment/life skills) ➢ # of days in program	➢ What quality indicators do we want to explore? (youth satisfaction, program, parent, community, etc.) ➢ What are folks currently collecting/evaluating with respect to quality/satisfaction?
BENCHMARKS: How much change did we produce?	IMPACT: What type of change did we produce?
➢ Increased # of skill gains ➢ Increased # of service hours ➢ Increased # credentials obtained ➢ Increased # of LT placements post attainment ➢ Increased # of LT placement retentions	➢ Change in behavior – Improved program participation and retention ➢ Change in knowledge/attitude – Improved credential attainment rate ➢ Change in circumstance – Improved LT placement rate

OUTCOMES

The list of ideal data elements got to be rather lengthy and difficult to manage. We re-strategized and organized the data elements in bigger buckets or categories. The four big categories were Demographics, Program Services/Dosage, Short-Term Outcomes, and Intermediate/Long-Term Outcomes. Participating agencies were asked to identify what data elements they have in each of the four categories. Once the feedback was obtained from all participating agencies, each list of data elements were reviewed to determine if there were consistent elements across each of the lists. A final list was constructed based on these common elements including the format of the data elements. The list is below:

Name of Field/Variable	Variable Dimensions/Levels
Demographics	
Date of Birth	
Gender	
Race/Ethnicity	
Zip Code	
Parent	Yes/No
Foster Care	Yes/No
Homeless/Living in Shelter	Yes/No
Disability (Physical/Mental)	Yes/No
Juvenile Justice/Court Involved	Yes/No
OSY w/HSD	Yes/No
OSY w/GED	Yes/No
OSY w/o HSD or GED	Yes/No
Program Services/Dosage	
Education	
Reading Remediation	Time Spent in Minutes
Math Remediation	Time Spent in Minutes
GED Prep/Secondary/Alt Secondary Supports	Time Spent in Minutes
Post Secondary Transition Support/College Prep	Time Spent in Minutes
Employment/Workforce	
Job Readiness Training	Time Spent in Minutes
Internship/Work Experience (paid/unpaid)	Time Spent in Minutes
Occupational Skills Training (Vocational Training)	Time Spent in Minutes
Empowerment/Youth Development	
Civic Engagement/Community/Volunteer	Time Spent in Minutes
Service Learning	Time Spent in Minutes
Leadership Development	Time Spent in Minutes

Financial Literacy	Time Spent in Minutes
Conflict Resolution	Time Spent in Minutes
Mentoring	Time Spent in Minutes
Case Management/Follow-Up/Barriers	
Child or Dependent Care	Yes/No
Medical/Insurance	Yes/No
Transportation	Yes/No
Housing	Yes/No
Job Assistance	Yes/No
Peer Support	Yes/No
Outcomes	
Attained HSD	Yes/No
Attained GED	Yes/No
Attained Occupational Skills/Advanced Training Credential/Vocational Certificate	Yes/No
Attained Post-Secondary Credential	Yes/No
Literacy Gains (ISY and OSY)	Yes/No
Numeracy Gains (ISY and OSY)	Yes/No
Long Term Educational Placement/Enrolled in Post Secondary Setting (college, uni)	Yes/No
Long Term Employment Placement (been employed for at least 90 days)	Yes/No
Long Term Occupational Skills/Advanced Training Placement/Vocational Placement	Yes/No
Military Placement	Yes/No
Complete Educational level/class OR Grade to Grade Promotion	Yes/No
Completed JRT	Yes/No
Complete short-term subsidized internship	Yes/No

At this point, while we have successfully identified a list, one of the most pressing concerns is the combination of data from different agencies that do similar, yet slightly different types of programming. Therefore, when you consider things such as time spent on Job Readiness Training, what each agency does in terms of job readiness training may be vastly different from one another. While it is being combined into a single dataset to answer questions of effectiveness, the question about what exactly about job readiness training produces positive outcomes would require further research given the different programming involved across agencies. This is true of any of the other data elements included in the data set under program services.

C. Obtaining the data elements identified

Each participating agency was responsible for providing the data elements identified in Section B. Agencies provided anonymous, individualized student data. Not all agencies were able to provide all of the 45 data elements identified in Section B. The datasets were relatively large: Baltimore provided approximately 1,100 cases, Philadelphia provided approximately 1,800 cases, and Hartford provided approximately 400 cases. Each of the datasets was provided through Excel spreadsheets.

D. Data translation, and combining into a unified dataset for analysis

This section required that all of the datasets be combined in a single dataset for analysis. It involved understanding the data elements within each dataset, how it related to the master list of common data elements outlined in Section B, and translating the data to fit a unified model. The first step in the process involved creating a master key list which contained the elements identified in Section B, along with related data elements from each agency. An example is provided below:

Data Elements	Balitmore	Philadelphia	Hartford
Date of Birth	birth_date	Date of Birth	Date of birth
Gender	gender	Gender	GENDER
Race/Ethnicity	race_name	Race	Race
Zip Code	zip	Zip	zip
Living in Foster Care	Foster Care	Served Demos - Have you been or are you currently in living in a foster home	DCF Status
Homeless/Living in Shelter	Homeless	Are you homeless or living in a shelter	Barriers
Disability (Physical/Mental)	Column I - Disability Status	Do you have a disability medical conditions	Learning Disability & Physical Disability

A tentative list was constructed and distributed to participating agencies to ensure that the master key list correctly mapped the data elements in each data set to the list in Section B. Once the list was verified, the next task was to put each of the individual datasets into a single dataset. The single dataset is being constructed in the statistical software program, SPSS. Within each dataset, there was some amount of data translation that needed to be done. For example in Baltimore, several of the data elements provided often linked to a single data element from the list in Section B. For example, the element *GED Prep/Secondary/Alt Secondary Supports* for the final dataset is a combination of two elements in the Baltimore database (*YO(BLT)-GED Prep (YD/PP)* and *YO(BLT)-BCCC GED Prep (YD/PP)*). Or in another example, Hartford, the data is provided in multiple datasets, where each dataset can have multiple efforts of the same participant. This is a product of the way that the data management system exports its data.

Once data has been translated within each dataset, the data is then transferred over to the master file. In cases where there is categorical data, such as with the ethnicity/race data element, each categorical

data element within each dataset needs to be translated into a unified set of categories. So for example, Baltimore had the following categories for ethnicity: *American Indian/Alaskan Native, Asian, Black/African American, Other, White*; Philadelphia had the categories: *African-American, Asian, Bi-Racial, Caucasian, Hispanic, Multi-Racial, and Other*; and Hartford had the categories: *African-American, Asian, Hispanic, Bi-Racial, and Multi-Racial*. In transferring the data, there needs to be a consistent translation of these categories across datasets.

This task of data translation and unifying requires significant time and effort to complete. The group is currently at this step.

E. Data analysis of combined data set to answer the goals of the report

Data analysis is geared towards fulfilling the last two goals of the initiative, namely:

2. To provide a profile of the youth served, programmatic efforts, and outcomes achieved by various agencies.
3. To determine and document any demonstrated effectiveness that can be obtained by looking at the combined datasets of participating agencies.

Data analysis will involve a combination of descriptive statistics, cross-tabulations, correlations, t-test/ANOVAs, and possibly a regression analysis. Descriptives and cross-tabulations will be used mainly to answer Goal 2. The other statistical procedures will be used to investigate any programming effectiveness common to all three agencies.

F. Producing the research initiative report for review and feedback by participating agencies and CCRY leadership

This report will be expanded to include the steps involved from Step E onward. An initial draft will be produced and then vetted by participating agencies and CCRY leadership to ensure accuracy and that it fulfills the objectives and expectations set forth by CCRY.

G. Report dissemination

A final copy of the report will be disseminated to the Network and others potentially interested stakeholders.

Report prepared by Sean Seepersad, PhD
Senior Director of Research and Organizational Performance
Our Piece of the Pie, Inc.
20-28 Sargeant St
Hartford, CT 06070

Questions or comments about this report can be directed to: sean.seepersad@opp.org