Great by Eight Program

Social Innovation Fund

Year 4 (Final) Implementation and Impact Evaluation Report

June 2018

Prepared by:

Michael L. Michlin, PhD
Kate K. Diamond, EdM

Research, Development and Engagement to Improve Education

Acknowledgements

How to Cite This Report

Contact Information

Delia Kundin, PhD
Associate Director for Evaluation Services
Center for Applied Research and Educational Improvement
College of Education and Human Development
University of Minnesota
1954 Buford Avenue, Suite 425
Saint Paul, MN 55108
Telephone: 612-624-2932
Email: dkundin@umn.edu

Table of Contents

List of Tables

List of Figures

About this Report

This evaluation report is a final report for the Great by Eight program, and is intended to fulfill the SIF requirements to determine at least a moderate level of evidence for funded projects. It includes implementation and impact studies focused on subsets of the children and families served and supported by Way to Grow.

Executive Summary

The purpose of this evaluation was to examine the implementation and impact of Way to Grow's Great by Eight program spanning activities from May 1, 2016 through December 15, 2017, which were parts of the fourth and fifth years of funding from Social Innovation Fund (SIF) through Greater Twin Cities United Way in partnership with Generation Next. The implementation and impact studies were guided by specific research questions outlined below. The evaluation study seeks to achieve moderate evidence, using a quasi-experimental design with matched comparison groups.

Introduction

Way to Grow provides services to families with children prenatally through age 8; the Great by Eight program, an intensive home visiting program, is designed to provide parents with the knowledge and resources needed to support their three- to eight-year-old children's early development and educational progress in school. The program targets children most at risk for low academic achievement in the Minneapolis metro area. During home visits, family educators provide parents with information on child development and strategies to help children acquire critical academic skills. Major goals for the program are to ensure children are ready for kindergarten and are prepared to meet or exceed grade level benchmarks in kindergarten through the third grade.

The Center for Applied Research and Educational Improvement (CAREI) in the College of Education and Human Development at the University of Minnesota was contracted to conduct the program evaluation of Great by Eight in compliance with SIF requirements.

Overview of Prior Research on Home Visiting Programs. During the early years, the home is a child's primary developmental context and the processes that occur within this setting are considered to be important influences on children's subsequent development. There is evidence to suggest that providing home visiting services to low-income, at-risk families early on promotes positive experiences within the home for children during the first few years of life. Studies have documented the ability of home-based services to effect positive changes in parenting behaviors such as the ability to set appropriate limits and engage in responsive and cognitively engaging parenting strategies. Research also shows that incorporating health education in home visiting programs makes a positive difference in children's health, development, and ability to learn. In addition to providing information on childhood health and development, these programs connect families to a broader range of community services and supports and produce measurable outcomes for children and families, such as better health and greater school readiness and academic achievement.

Research on home visiting programs such as Parents as Teachers offers evidence on the potential effectiveness of Great by Eight family services. Rigorous research studies have shown that parents learned how to interact with their child more effectively, had a better understanding of child development, and spent more time with their children. Parents also engaged in more conversations with their children and were more likely to promote reading in the home. Children scored higher

on kindergarten readiness tests and standardized measures of reading, math, and language in elementary grades.

The literature search by Michlin and Schultz (2016) investigated the effects of home visiting programs on both child and family outcomes. They found that, "Overall home visiting programs seem to have a small but significant effect on various child and family outcome measures. Although these effects are small, the practicality of reducing child maltreatment or increasing access to health care may be meaningful no matter how small the impact." (p. 4) Michlin and Schultz also reported on programs finding significant effects on child development and school readiness, positive parenting practice, family economic self-sufficiency, child health, and reductions in child maltreatment. (p. 3)

Implementation Study

This final evaluation research included an implementation study as described in the revised Subgrantee Evaluation Plan (SEP) that was approved by CNCS on July 13, 2016. For the implementation study, data were collected primarily through a newly developed parent survey, the Life Skills Progression-Parent Scale, the Home Visit Rating Scales, and the Way to Grow longitudinal programming database. The implementation research questions were

1. What topics presented by family educators during home visits are perceived by parents to have the greatest impact on their knowledge concerning their children's school-related and health-related outcomes?

2. From the parents' perspective, what program delivery features of Great by Eight are most important for program effectiveness?

3. How well do family educators relate to participating parents and their families? What aspects of the parent-family educator relationship need improvement?

4. What is the home visit dosage (frequency, spacing, duration) currently received by parents participating in the Great by Eight program? How does the actual dosage compared to the dosage recommended in the program guidelines?

Implementation Study Findings

Parent Survey. For all topics concerning children's school-related and health-related outcomes that are presented by family educators during home visits, at least 70% of respondents reported having learned *a lot* and that the topics were *very useful,* indicating a high overall level of reported learning and perceived usefulness. To be sure there was response variation to amount learned and usefulness of topics depending on the parent's child's age group: birth to 2 years old, three to five year olds, or kindergartners through third graders.

Evaluators also compared some Life Skills Progression Scale (LSP) data with the parent survey questions about amount learned and usefulness of topics, in order to explore whether triangulation of the two instruments was possible. We calculated Spearman correlations between the six LSP

variables and the parent survey items. Overall, most of the correlations between LSP variables and parent survey items were small. We explored several potential explanations for why greater triangulation was not reached.

The survey asked respondents to rate six aspects of the Great by Eight program using the options *very unimportant, unimportant, important,* and *very important*. Respondents could also select *unable to rate.* For all program components included in the survey, at least 70% of respondents selected *very important,* indicating a high overall level of perceived importance of the different aspects of the program.

Respondents were asked to rate several items related to their interactions with their family educator, using the options *never, seldom, usually, and always.* Respondents also had the choice of selecting *unable to rate.* Over 90% of respondents selected *always* for all seven aspects of the parent-family educator relationship, indicating a very high level of satisfaction in this area.

The survey also asked respondents to rate the overall *quality* of their relationship with the family educator, using the options *unable to rate, poor, fair, good,* and *excellent*. Nearly 90% of respondents selected *excellent*.

Home visit dosage. During the 5/1/16 to 12/15/17 timeframe of this final evaluation, family educators were extremely busy with 18,401 client/family contacts. In this report, home visit dosage, however, refers only to the early learning visits for families with children age 3 through pre-K (n = 8,967) and elementary program visits with children in K through third grade (n = 4,468), and not, for example telephone visits, e-mail contacts, group-based visits, or the like. For reasons of continuity in reporting implementation and impact results as well as the practicality of handling data and analyses, and importantly per the approved SEP, evaluators limited the analysis of frequency, duration, and spacing to the 100 sampled families for the linear mixed-effects (LME) models analyses.

Frequency of home visits. Of the 100 sampled families, the number of families in this analysis with one or more visits was 70. The range of visits (frequencies) per family was from 1 to 98 visits for the sample during the reporting period with a mean of 21 visits; the median was 20, but the mode of families per frequency was 1.

Duration of home visits. The range of minutes per family home visit was from 5 to 200 for the sample. The mean number of minutes per home visit was 48.5; the median was 45 minutes, and the mode was 60 minutes per visit.

Spacing between visits. We discovered that the spacing between visits for this sample during this timeframe cannot be usefully captured in any meaningful number(s) such as a mean number of days.

Impact Study One: Kindergarten Readiness and Third Grade Proficiency

Great by Eight kindergartners and third graders were matched with similar Minneapolis public school system children to study the program's effectiveness on educational progress. Propensity score matching procedures were used to select comparison group samples. Kindergarten children were compared on Minneapolis Public Schools' Beginning Kindergarten Assessment (BKA), a standardized assessment of literacy skills administered in the fall of the kindergarten year. Third graders' academic achievement was measured using the Minnesota Department of Education's standardized Minnesota Comprehensive Assessment (MCA) reading test. The Impact Study One research questions were

1. Are Great by Eight children more prepared for kindergarten than non-program children?

2. Are Great by Eight children more prepared to test proficient on the third-grade reading assessment than non-program children?

Impact Study One Findings

Kindergarten Readiness. We analyzed BKA total literacy scores. An independent samples t-test was used to determine if there were statistically significant differences between Great by Eight kindergartners ($n = 31$) and comparison group kindergartners ($n = 27$). The results showed that there was no significant difference between group mean scores on the BKA total literacy scores.

Third Grade Proficiency. We analyzed third-grade MCA reading scores. An independent samples t-test was used to determine if there was a difference on MCA reading scores between program third graders ($n = 69$) and comparison group third graders ($n = 63$). The results showed that Great by Eight third graders scored significantly higher than comparison group third graders. The effect size calculation showed a small to moderate program intervention impact.

Impact Study Two: Changes and Progress in Children and Parents

The program evaluation used linear mixed-effects models to examine changes in parents based on multiple measures in a longitudinal format. The evaluation was specifically interested in exploring and confirming changes based on home visitation interventions related to increased parent engagement with their children's development, learning, and schooling.

The evaluators randomly selected a subgroup of 100 families with children ages three to eight years (families 3-5 yrs. $n=36$; families K-3rd grade $n=64$) enrolled in the Great by Eight program. These are the same 100 sampled families for the linear mixed-effects (LME) models analyses described in the implementation analysis. A total of $n=177$ children are represented in the 100 families. The Impact Study Two research questions were

1. Do parents show increased engagement with their children's development, learning, and schooling?

2. Does child development and growth progress achieve appropriate milestones and in conjunction with preparation and success for formal schooling?

3. Do Great by Eight health components (health referrals and education measured by dosage) improve and sustain health promoting behaviors?

4. Do Great by Eight parents and children access resources and navigate health and education systems?

Impact Study Two Findings

Parent engagement with their children's development, learning, and schooling. The construct of parent engagement was measured with the Life Skills Progression (LSP) Scale. Indicators in the "Relationship with Children" LSP category were used to assess parent engagement with their children over time. These scales included 1) Nurturing (Skills and Ability); 2) Discipline (Appropriateness); 3) Support of Child Development; 4) Safety (protecting children from environmental harm); 5) Use of Resources; 6) Child Care (quality of child care environment): and 7) Cognitive Ability (cognitive understanding of children's needs).

Results for parent engagement from measurements of LSP scale, Parent Engagement Code, and Parent Involvement with Children's School are described below.

LSP scale (parent engagement). We found very little change over time in the parent engagement outcome variables, the seven LSP scales speaking to parent engagement. The LSP ratings that parents received were consistently high, even during the earliest observations. Because parents were rated so highly from the first time they were assessed, there was little room for the average score to increase over time. This meant that the LSP instrument *could not* capture substantial change over time. The lack of statistically significant covariate coefficients – or, in the case of the *number of visits* variable, coefficients that did not differ significantly from zero – may be due to the very small amount of variation in the outcome variables that could be explained in the first place.

Parent Engagement Code. The Parent Engagement Code is a zero to 5 rating made by the family educator at each home visit on the *quality* of a parent's engagement with her child. Our analysis revealed significant differences between parents of different ethnicities. Specifically, we found that Asian and Hispanic/Latino parents were 99 percent less likely to receive a rating of five compared to the reference group (i.e., White parents).

Parent Involvement with Children's School. We found that most parents were involved at their child's school through parent teacher conferences. Forty-five percent of the K-3 parents (*n*=64) in the 100 sampled families during the May 1 through December 15, 2017 timeframe, attended an average of two parent-teacher conferences, 16 percent of these parents attended less than two school events, and less than one percent volunteered at school.

Child development and growth in preparation for and success in formal schooling. We found that the Great by Eight four year olds were on a positive trajectory for meeting expectations at each time point for kindergarten readiness. Literacy skills improved significantly as children moved from one IGDI and DIBELS assessment period to another.

Parents' access to resources and navigation of health and education. The same 100 families sampled for the other impact analyses served as the sample for this analysis. Way to Grow's programming data for these families for the date range May 1, 2016 through December 15, 2017, showed that 50 program parents were referred to one or more community services, and parents were often referred to the same resource more than once. For these 50 parents, 118 community service referrals were made. Twenty-five parents had referrals to an early learning preschool program for their children who were 3-5 years of age, and 56 percent of parents had referrals related to family medical care.

LSP scale (health care). We examined four LSP scales related to health care. Two scales assess whether parents have obtained appropriate medical care (immunizations and dental care) for their children. The third scale rates parents on accessing health/medical insurance, an outcome that family educators stress during home visits. The last indicator evaluates general parental child wellness care. The same independent variables that were included in the parent engagement LSP variables were used for this analysis: measurement time, number of home visits, primary home language, and ethnicity.

We found that, with the exception of the Medical Health Insurance scale, the mean value at each measurement time (including time 1) was between 4 and 5, which meant that there was again little room for the scores to increase over time. For all scales except for Medical Health Insurance, the means were higher at measurement time 7 (the last measurement time with more than 1 score recorded) compared to measurement time 1, but the increases were small in magnitude. For the Medical Health Insurance variable, there was a slight decrease in the average score over time.

On average, between one measurement time to the next, there was a small but significant point increase in Child Immunization, Child Dental Care, and Child Wellness Care. There were several ethnicity coefficients that were statistically significant.

Summary and Conclusions

Evaluators presented an overall summary of findings and limitations, a brief comparison of results with previous interim evaluations, and finished with conclusions and next steps. Among the next steps, they recommended that Way to Grow look long and hard at how they are undertaking quality control in the Great by Eight program; assess how heavily they should rely on public school collected data for their in-house evaluation of Great by Eight program; be cautious in undertaking comparisons with groups of non-participant children on any metric; consider tracking participant children and families over a longer timeframe; assess the appropriateness of their current outcome measurements; and consider less focus on educational/academic outcomes and more focus on non-school outcomes.

Introduction

Purpose and Audience

The purpose of this evaluation was to examine the implementation and impact of the Great by Eight program for parents of children ages 3 through 8 spanning activities from May 1, 2016, through December 15, 2017 (the timeframe), which were parts of the fourth and fifth years of funding from Social Innovation Fund (SIF) through Greater Twin Cities United Way in partnership with Generation Next.

The evaluation study seeks to achieve moderate evidence, using a quasi-experimental design with matched comparison groups.

This evaluation report is submitted to the Corporation for National and Community Service (CNCS) and its SIF. It is also directed to the SIF grantees Generation Next and Greater Twin Cities United Way, and to the subgrantee Way to Grow.

Background of Way to Grow and the Great by Eight Program

Way to Grow was founded as a collaborative effort to address disparities in early education in Minneapolis through activities that empower parents to be their children's primary teachers. Responding to research that shows educational disparities emerge before children step into a classroom and widen throughout their first years in school, Great by Eight programming targets families and their children ages three to eight who are most at risk for low academic achievement. This targeted work aims to reduce the academic achievement gap, ensuring that these children enter kindergarten with the literacy and social and emotional strengths needed for successful learning and that each student meets key benchmarks for success in reading by the third grade.

The program's education model targets children experiencing the highest levels of poverty with the assumption (and evidence) that poverty will have a deep impact on broader economic, educational, and social systems if not addressed during the earliest years of children's lives. To have the deepest and most lasting impact on the educational outcomes of these children, Great by Eight works closely with parents assuming that their increased knowledge and engagement will lead to improved child success. It is central to this assumption that parents will open their homes to an outside organization, desiring to be their children's primary teachers in an effort to see their children succeed in school and life.

Great by Eight has structured a parent-centric model around home visiting, basing interventions on studies supporting the effectiveness of home visiting on improving academic achievement. The program operates under the assumption that children will be more school ready and meet key grade level benchmarks by families engaging in the home visiting model year-round. Further, the program is driven by metrics and outcomes that inform ongoing program development. Without these metric-driven, systematic solutions to address educational disparities, Way to Grow assumes these disparities will persist.

Great by Eight family educators are collectively fluent in seven languages and culturally reflect the children and families they serve. During home visits, family educators model for parents how to create a culture of learning in the home. Research and evidence-based curricula and assessments guide Great by Eight's home visiting practices and instruction. Additionally, the curriculum is guided by Minnesota's Early Childhood Indicators of Progress. The early learning program addresses educational disparities and builds a parent-centric model for school readiness by 1) nurturing and encouraging parents to be their children's first teachers, 2) providing early childhood development and parent education, and 3) implementing research-based literacy and numeracy instruction in the home (http://education.state.mn.us/MDE/dse/early/ind/).

Program Intended Outcomes

- Parents engage in high quality interactions with their children
- Improved home environment that supports learning
- Parents have the skills to support their children's learning and development
- Increased and sustained parent engagement with their children's development and school system
- Children are prepared for school (kindergarten)
- Children demonstrate improved engagement in school and develop skills that support learning
- Children are prepared to meet or exceed grade level benchmarks in kindergarten through third grade
- Children demonstrate proficiency on the statewide third grade reading assessment
- Parents and children have improved health and demonstrate health promoting behavior
- Parents access resources and navigate health and education systems

Overview of Prior Research

Home Visiting Programs. During the early years, the home is a child's primary developmental context and the processes that occur within this setting are considered to be important influences on children's subsequent development (Bronfenbrenner & Morris, 1998). There is evidence to suggest that involving low-income, at-risk minority families in home visiting services early on promotes positive experiences within the home for children during the first few years of life (Azzi-Lessing, 2011; Kirkland & Mitchell-Herzfeld, 2012). Studies have documented the ability of home-based services to effect positive changes in parenting behaviors such as the ability to set appropriate limits (Mitchell-Herzfeld, DuMont, Lee, & Spera, 2005) and engage in responsive and cognitively engaging parenting strategies (Rodriguez, DuMont, Mitchell-Herzfeld, Walden, & Greene, 2010). Research also shows that incorporating health education in home visiting programs makes a positive difference in children's health, development, and ability to learn (Caldera, Burrell, Rodriguez, Crowne, Rohde, & Duggan, 2007). In addition to providing information on childhood health and development, these programs connect families to a broader range of community services and supports and produce measurable outcomes for children and families, such as better health and greater school readiness and academic achievement (Children's Defense Fund, 2013).

Research on home visiting programs such as Parents as Teachers offers evidence on the potential effectiveness of Great by Eight family services. Rigorous research studies have shown that parents learned how to interact with their child more effectively, had a better understanding of child development, and spent more time with their children (Pfannenstiel, Seitz, & Zigler, 2002; Research and Training Associates, 2002). Parents also engaged in more conversations with their children and were more likely to promote reading in the home (Albritton, Klotz & Roberson, 2004; Pfannenstiel, Seitz, & Zigler, 2002; Research Training Associates, 2006; Zigler, Pfannenstiel & Seitz, 2008). Additionally, positive academic results for children were identified. Children scored higher on kindergarten readiness tests and standardized measures of reading, math, and language in elementary grades (Pfannenstiel, Seitz & Zigler, 2002; Zigler, Pfannenstiel & Seitz, 2008).

The literature search by Michlin and Schultz (2016) investigated the effects of home visiting programs on both child and family outcomes. They found that, "Overall home visiting programs seem to have a small but significant effect on various child and family outcome measures. Although these effects are small, the practicality of reducing child maltreatment or increasing access to health care may be meaningful no matter how small the impact." (p. 4) Michlin and Schultz also reported on programs finding significant effects on child development and school readiness, positive parenting practice, family economic self-sufficiency, child health, and reductions in child maltreatment. (p. 3)

Finally, a 2013 report by Child Trends described the early implementation of Great by Eight's first two cohorts of children to receive the expanded services beyond early learning and into the elementary grades. The report showed that 93% of Great by Eight children were meeting the Minneapolis Public School literacy standard at the beginning of kindergarten. The report also showed that most parents were engaged in their child's education, with 96% attending a school conference and 60% reading to their child at least five times a week.

Logic model. The program's logic model shows the resources and activities Way to Grow utilizes to reach intended short-term outcomes and long-term impact. It lays out specific activities that Great by Eight family educators implement during home visits, resource connections, community partnerships, and community-based visitations. These activities describe the content, methods, strategies, and ongoing actions that engage families, schools, and the community to be a part of the systemic change required to ensure children are ready to succeed in school and in life (see Appendix A).

Research Questions

The evaluation was designed to answer key questions about the program, based on the program's theory of change and logic model.

Implementation

1. What topics presented by family educators during home visits are perceived by parents to have the greatest impact on their knowledge concerning their children's school-related and health-related outcomes?

2. From the parents' perspective, what program delivery features of Great by Eight are most important for program effectiveness?

3. How well do family educators relate to participating parents and their families? What aspects of the parent-family educator relationship need improvement?

4. What is the home visit dosage (frequency, spacing, duration) currently received by parents participating in the Great by Eight program? How does the actual dosage compared to the dosage recommended in the program guidelines?

Impact Confirmatory

1. Are Great by Eight children more prepared for kindergarten than non-program children?

2. Are Great by Eight children more prepared to test proficient on the third grade Minnesota Comprehensive Assessment (MCA) than non-program children?

3. Do parents show increased engagement with their children's development, learning, and schooling?

4. Does child development and growth progress achieve appropriate milestones and in conjunction with preparation and success for formal schooling?

5. Do Great by Eight health components (health referrals and education measured by dosage) improve and sustain health promoting behaviors?

6. Do Great by Eight parents and children access resources and navigate health and education systems?

Impact Exploratory

There were two exploratory research questions in the SEP. These were

1. What populations are most likely to benefit from Great by Eight programming?

2. Which program components impact program outcomes?

The Great by Eight program targets families and their children ages three to eight who are most at risk for low academic achievement. No other populations are available through programming for comparison. Thus, the first question cannot be addressed based on the Great by Eight programming and data available for this final evaluation or in either of the two previous interim evaluations (Daugherty & Edwards, 2015, and Dretzke et al., 2016). Regarding the second question, all of the program components listed as outputs (activities) in the program logic model, impact program outcomes in some way and to some degree. These questions cannot be answered, and thus we departure from the SEP.

The evaluation examined aspects of the implementation and impact of the Great by Eight program outlined in the SEP for children ages 3 through 8 spanning activities from May 1, 2016 through December 15, 2017, which were parts of the fourth and fifth years of SIF funding.

Evaluation Design

This final evaluation included an implementation study as described in the revised SEP (approved by CNCS on July 13, 2016) and two impact studies identified in the original SEP. For the implementation study, data were collected primarily through a newly developed parents survey, the Life Skills Progression-Parent Scale, the Home Visit Rating Scales, and the Way to Grow longitudinal programming database.

The first impact design centered on children in the program and their educational progress. Program impact confirmatory questions were 1) Are children ready for kindergarten as compared to similar children who did not receive program services; and; 2) Are third grade children proficient in meeting academic literacy standards relative to comparison group children? A quasi-experimental between groups design was used to study the program's impact on these two outcomes. Comparison groups were selected based on propensity score matching.

The second impact design focused on Great by Eight children and parents' progress towards meeting specific program goals. These goals were 1) Parents engage in high quality interactions with their child; 2) Child development and growth progress achieving appropriate milestones and in conjunction with preparation and skills for formal schooling; and 3) Children and parents have access to high quality health and educational services. An interrupted times series design was proposed in the SEP to study the effects of Great by Eight programming on these objectives. The SEP proposed the use of both interrupted time series (ITS) analysis and linear mixed-effects (LME) models in analyzing the longitudinal Way to Grow data. The evaluators, however, subsequently determined that an ITS design was not feasible, and we thus departed from the approved SEP.

In ITS analysis, the outcome variable of interest – for example, the quality of a parent's engagement with her child – is measured several times before an intervention is introduced and again several times after the intervention is introduced. The repeated measures data is then examined to see if there was a significant change in the variable (either in its average value or its rate of change) from before to after the intervention (Biglan, Ary, & Wagenaar, 2000).

The Way to Grow data does not contain repeated measures data that were collected before the intervention (home visits) began. Instead, data collection for the LSP and parent engagement code variables occurred in parallel with the home visits. The lack of pre-intervention repeated measures data makes an ITS analysis impossible. We were however able to use LME models, as proposed in the SEP, because LME analysis is a more flexible framework that can be used with repeated measures data regardless of when the intervention occurred.

Implementation Evaluation

This final evaluation research included an implementation study as described in the revised SEP.

Research Questions

1. What topics presented by family educators during home visits are perceived by parents to have the greatest impact on their knowledge concerning their children's school-related and health-related outcomes?

2. From the parents' perspective, what program delivery features of Great by Eight are most important for program effectiveness?

3. How well do family educators relate to participating parents and their families? What aspects of the parent-family educator relationship need improvement?

4. What is the home visit dosage (frequency, spacing, duration) currently received by parents participating in the Great by Eight program? How does the actual dosage compared to the dosage recommended in the program guidelines?

Method

The primary data source for answering the first three questions was a newly developed parent survey, the Life Skills Progression-Parent Scale, the Home Visit Rating Scales, and, for home visiting dosage, the Way to Grow programming database. Descriptions including psychometric characteristics of the measurement instruments are detailed in Appendix B.

Parent survey administration

The revised SEP proposed administering the survey to a random sample of approximately 300 parent participants selected in equal numbers from the three program components defined by the age/grade categories of the participating children: birth to 2 years = 100; ages 3 to 5 years = 100, and grades kindergarten to 3 = 100. The survey was to be administered online and could be taken on a home computer or cellphone. Computers would also be available for survey completion during parent events scheduled at the Way to Grow offices, and Way to Grow consultants would be present during these events to assist parents when literacy skills were a concern. Contrary to the revised SEP, parent participants taking the survey had not been sampled. Also, the survey was constructed and administered to allow a survey taker to respond to the items for their child or children in each of the three age/grade categories.

Way to Grow staff and CAREI evaluators drafted and field-tested the online version in English. They then completed a Spanish version. Both online surveys were hosted on the University of Minnesota's Qualtrics system. Although proposed, Somali and Hmong versions were not available. The Way to Grow staff person working on the Somali translation indicated that the Somalis preferred to complete the survey in English, and the Way to Grow staff person working on the Hmong version stated that the Hmong parents participating in Way to Grow often do not read

---

Hmong. Instead, Hmong survey participants were asked to complete the survey at the Way to Grow office where Hmong-speaking consultants would be available to provide translations. Consultants (not staff) also completed hardcopy English and Spanish versions for use when paper-and-pencil was preferred.

Sample

The parent survey sample was the group of parents taking the survey and providing responses to some or all items on the survey ($n = 128$). Their responses provided the data for addressing the first three implementation research questions. Ninety-eight percent of the surveys were taken between April and November 2017. The last parent survey was entered in the Qualtrics electronic database on 11/15/17. In the end, there were 103 English version surveys and 25 Spanish version surveys with usable data. This covered 81 prenatal to 2-year-old children, 105 3- to 5-year-old children, and 98 kindergarten through 3rd grade children. The target was 100 children covered in each of the three age groups. Most of the 128 survey takers had two or more children in the three age/grade categories and gave survey responses to age/grade appropriate categories.

The sample for the fourth implementation research question was the random sample of 100 families from the group of children aged three to eight years participating in the Great by Eight program and their parent(s) selected to participate in the linear mixed-effects (LME) evaluation plan. These are the ages of interest for the SIF-supported Way to Grow programming in which family educators are delivering home visits for early learning and elementary program curricular visits.

<div align="center">

Implementation Study Findings

</div>

This section presents survey results that help to answer the first implementation research question:[1]

1. What topics presented by family educators during home visits are perceived by parents to have the greatest impact on their knowledge concerning their children's school-related and health-related outcomes?

Specifically, the survey asked respondents (n=128) to indicate **how much they had learned** about various topics through the home visits, as well as **how useful they found the topics to be**. Some topics applied to parents with children of all age ranges, whereas others were specific to parents with children in a certain age or grade range (expecting a baby to 2 years, 3 to 5 years, and kindergarten through 3rd grade). For each topic, respondents could indicate that they had learned *nothing, a little, a moderate amount,* or *a lot;* and could indicate that the topic was *not useful, somewhat useful,* or *very useful.* They could also select *unable to rate* if they did not recall receiving information on the topic or if they did not know how much they learned or how useful it was. For all topics included in the survey, at least 70% of respondents reported having learned *a lot* and that the topic was *very useful,* indicating a high overall level of reported learning and perceived usefulness.

---

[1] Frequencies of responses to all the survey items are available in Appendix C.

Topics applying to all age/grade ranges. Among the nine topics that applied to all age/grade ranges (*n*=128), the ones with the highest percentage of parents reporting **having learned** *a lot*, as demonstrated in Figure 1, were the parent's role as the child's first teacher (86.78%), learning in the home (85.25%), and reading aloud to the child (85.25%). There were four topics for which less than 80% of respondents indicated having learned *a lot*: connections to resources and information (72.13%), immunizations (78.15%), nutrition (79.17%), and parenting skills (79.51%).
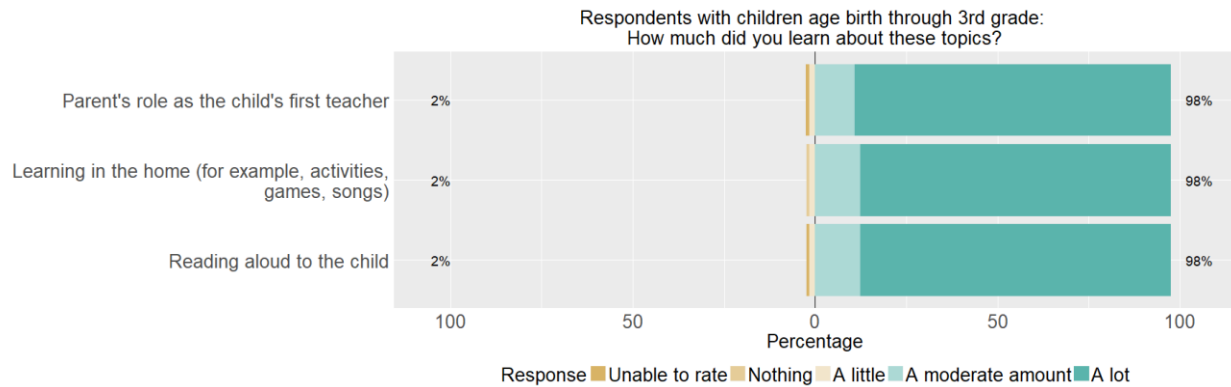
Figure 1. Respondents with children age birth through 3ʳᵈ grade: Perceived amount of learning about topics

The survey asked respondents about **how useful** they thought the same topics were. More than 80% of respondents selected *very useful* for all topics included in this section. The topics with the highest number of respondents selecting *very useful* were the parent's role as the child's first teacher (92.17%), parenting skills (90.52%), reading aloud to the child (89.66%), and learning in the home (88.70%). These results are displayed in Figure 2.
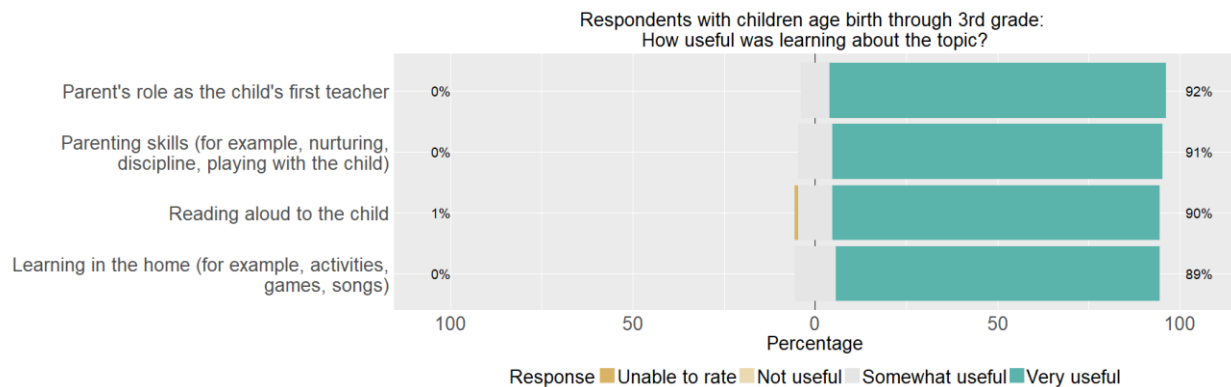
Figure 2. Respondents with children age birth through 3ʳᵈ grade: Perceived usefulness of topics

Topics applying to respondents expecting a baby or with children age birth through 2 years. The survey explored four topics that were specific to parents who were expecting a baby or who had children age birth through 2 years. A high percentage of the parent respondents (*n*=81) in this group (83.75%) reported having learned *a lot* about early childhood screenings and infant/toddler growth and development (see Figure 3). Two additional topics were explored for these parents including caring for newborn/infants and healthy pregnancy/prenatal care. The majority of respondents said they learned *a lot* about caring for newborn/infants (73.75%) and healthy

pregnancy/prenatal care (70%). A few respondents indicated they learned *a little* about these topics: caring for newborn/infants (6.25%) and healthy pregnancy/prenatal care (7.50%). Fewer than 5% of respondents indicated they learned *nothing* (3.75%) or they selected *unable to rate* (3.75%) this topic if they did not recall receiving information about these topics or they did not know how much they learned.
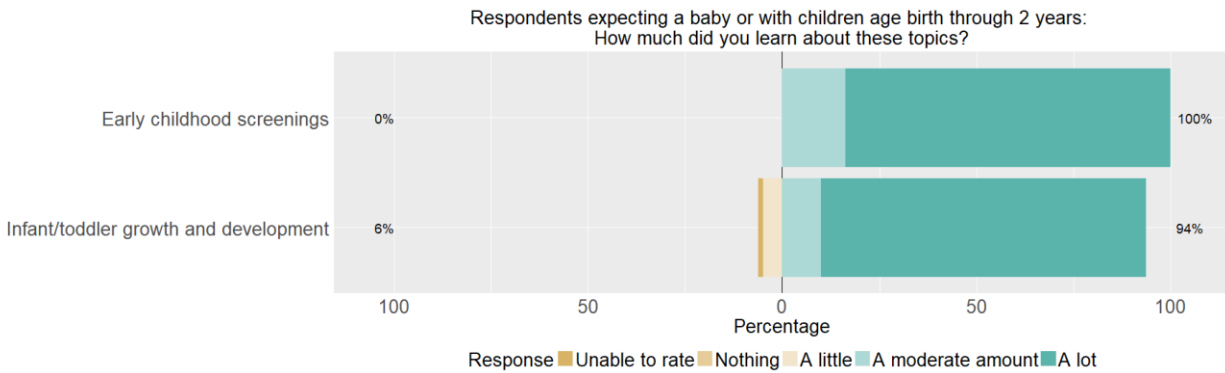
Figure 3. Respondents expecting a baby or with children age birth through 2 years: Perceived amount of learning about topics

The same two topics shown in Figure 3 above had the highest percentage of respondents indicating they had been *very useful*: early childhood screenings (89.74%) and infant/toddler growth and development (83.33%). Figure 4 displays the results for these two items. In addition, nearly twenty percent (19.48%) of parents for the age group indicated that learning about caring for a newborn/infant was somewhat useful, while one respondent (1.30%) thought the information was *not useful.* A few respondents (3.90%) selected *unable to rate* for this topic.
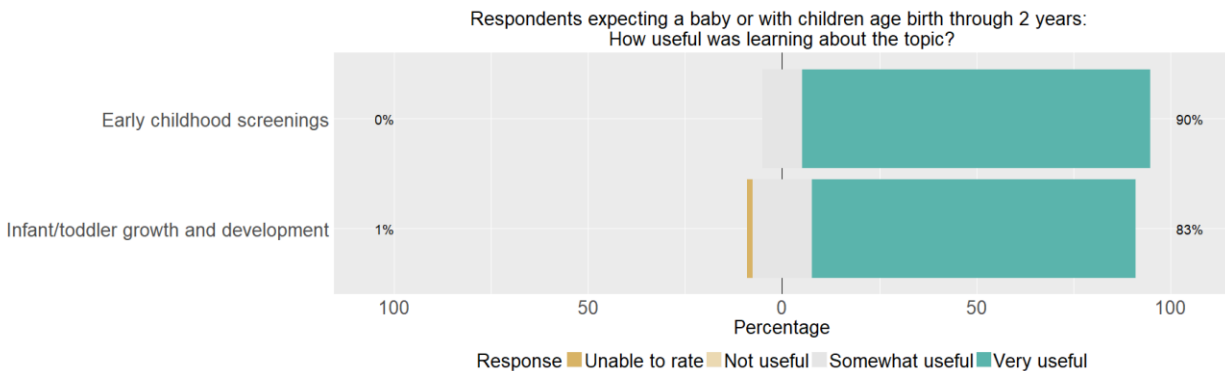
Figure 4. Respondents expecting a baby or with children age birth through 2 years: Perceived usefulness of topics

Topics applying to respondents with children age 3 to 5. Respondents with a child or children age 3 to 5 (*n=105*) were asked about six topics specific to that age range. The three topics with the highest percentage of respondents saying they had learned *a lot,* as displayed in Figure 5, were skills a child needs to be ready for kindergarten (90.29%), school expectations for preschool through kindergarten (88.46%), and child growth and development (88.35%).
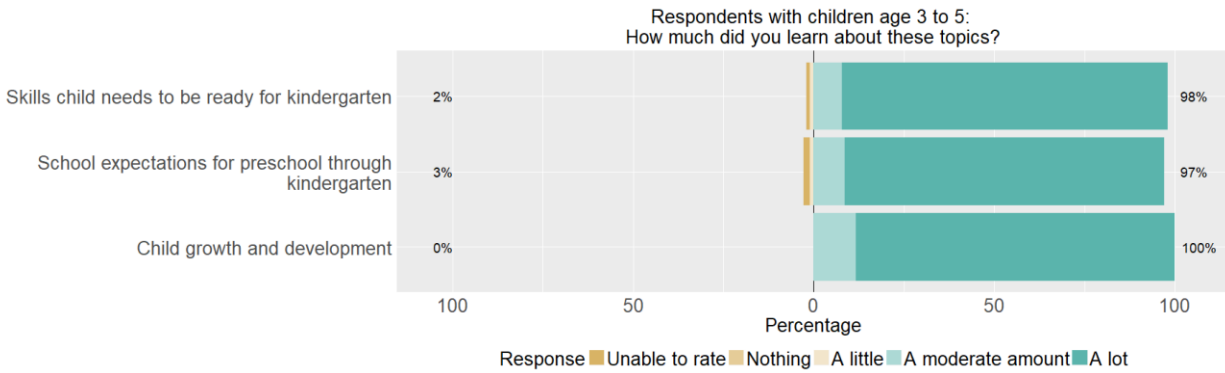
Figure 5. Respondents with children age 3 to 5: Perceived amount of learning about topics

Across all six topics included in this section, more than 80% of respondents reported finding the topics *very useful.* Over 90% of respondents said that the topic of skills child needs to be ready for kindergarten was *very useful* (90.91%), and 88.89% reported that the following topics were *very useful*: preschool referrals and enrollment, school expectations for preschool through kindergarten, and child growth and development (see Figure 6).
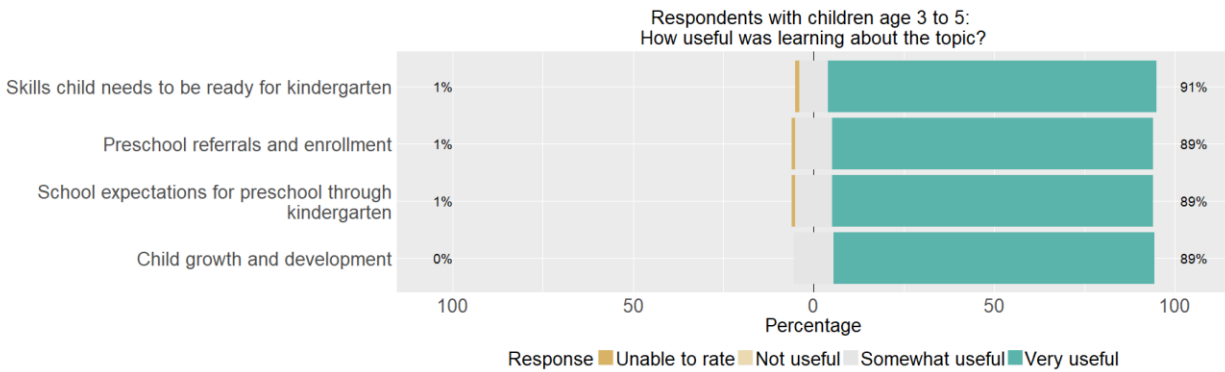
Figure 6. Respondents with children age 3 to 5: Perceived usefulness of topics

Topics applying to respondents with children in kindergarten through 3rd grade. For the six topics that applied to parents with a child or children in kindergarten through 3rd grade (*n*=98), the two with the highest percentage of respondents saying they had learned *a lot* were attending parent-teacher conferences (87.63%) and skills the child needs to be a successful reader (82.47%). These results are shown in Figure 7. There were three topics for which less than 80% of respondents said they had learned *a lot*: elementary school choice and enrollment (74.47%), parent engagement in school (75.26%), and dental care (76.84%).
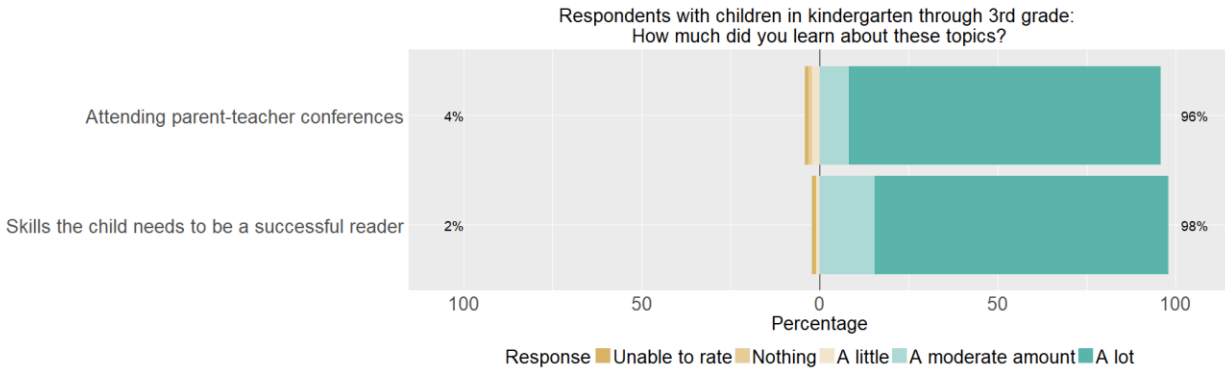
Figure 7. Respondents with children in kindergarten through 3rd grade: Perceived amount of learning about topics

The majority of kindergarten through 3rd grade parents found topics such as skills the child needs to be a successful reader (93.55%) and attending parent-teacher conferences (91.3%) to be *very useful,* as demonstrated in Figure 8. Nearly 20% of respondents found two other topics to be *somewhat useful*: parent engagement in school (18.09%) and elementary school choice (17.02%). One percent of respondents indicated the latter two topics were *not useful*, while few parents selected *unable to rate* for parent engagement in school (3%) or elementary school choice (1%).
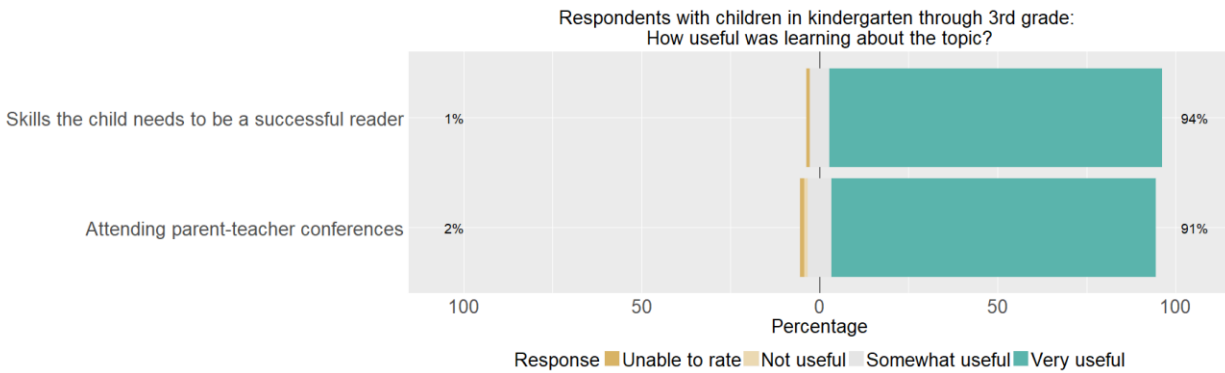
Figure 8. Respondents with children in kindergarten through 3rd grade: Perceived usefulness of topics

To further explore question 1, we compared the Life Skills Progression Scale data (LSP data) with the parent survey questions about amount learned and usefulness of topics, in order to explore whether triangulation of the two instruments was possible.[2] To do this we used the unique family IDs to match the data. We calculated Spearman correlations between the six LSP variables and the parent survey items. Table 1 reports the correlations that were at least moderate in size ($r > +/-.3$).

Use of Information (captures a parent's ability to use information) and *Use of Resources* were the LSP variables with the highest number of moderate correlations with parent survey items. Both of

---

[2] The point of triangulation was to see how aligned parents' perceptions were with family educators' when it came to the amount of learning and the value of the topics.

these LSP variables had moderate positive correlations with how much parents thought they learned about:

- Skills child needs to be ready for kindergarten,
- School expectations for preschool through kindergarten, and
- Parent engagement in child care setting or preschool (for example, volunteering, attending events, conferences, or field trips);

As well as how useful parents found learning about:

- Learning in the home (for example, activities, games, songs) and
- Immunizations

As such, for some of the topics included in the parent survey, parents' perceived learning and usefulness of knowledge were associated with higher LSP observation scores on *Use of Information* and *Use of Resources.*

A more puzzling finding was that the *Nurturing* LSP variable had moderate *negative* correlations with how useful parents found the topics of school expectations for kindergarten through third grade and attending parent-teacher conferences. The *Support of Development* LSP variable had moderate *negative* correlations with how much parents thought they had learned about healthy pregnancy/prenatal care and how useful they found the topic of caring for a newborn/infant to be.

Overall, most of the correlations between LSP variables and parent survey items were small (and subsequently not reported in Table 1). There are several potential explanations for why greater triangulation was not reached. For instance, the two instruments are very different from each other. One instrument is filled out by parents and is focused on their perceived learning, whereas the other instrument is filled out by the family educator based on their observations of the parents' skills. Furthermore, whereas the LSP measures very general constructs (e.g., nurturing, discipline), the parent survey asks about more specific topics (e.g., immunizations, school expectations for preschool through kindergarten).

Table 1

Spearman Correlations between LSP Variables and Parent Survey Items*

	LSP Variable					
	Nurturing	Discipline	Support of Development	Safety	Use of Information	Use of resources
How useful was learning about						
Q9_2: Learning in the home (for example . . .)	--	--	--	--	.39	.35
Q9_3: Reading aloud to the child	--	--	--	--	.31	--
Q9_7: Immunizations	--	--	--	--	.45	.38
Q15_2: Caring for newborn/infant	--	--	-.33	--	--	--
Q27_2: School expectations for kindergarten through 3rd grade	-.36	--	--	--	--	--
Q27_5: Attending parent-teacher conferences	-.32	--	--	--	--	--
How much did you learn about						
Q13_1: Healthy pregnancy/prenatal care	--	--	-.35	--	--	--
Q19_2: Skills child needs to be ready for kindergarten	--	--	--	--	.35	.33
Q19_3: School expectations for preschool through kindergarten	--	--	--	--	.35	.35
Q19_4: Parent engagement in child care setting or preschool (for example . . .)	--	--	--	--	.41	.39

*Only correlations that are at least moderate in size ($r > +/- .3$) are reported.

The next section presents survey results that help to answer the second implementation research question:

2. From the parents' perspective, what program delivery features of Great by Eight are most important for program effectiveness?

The survey asked respondents to rate six aspects of the program using the options *very unimportant, unimportant, important,* and *very important*. Respondents could also select *unable to rate*. For all program components included in the question, at least 70% of respondents selected *very important,* indicating a high overall level of perceived importance of the different aspects of the program. As shown in Figure 9, the program components with the highest percentage of respondents selecting *very important* were:

- The family educator understands and has respect for my culture (91.06%),
- The family educator comes to my home (87.5%), and
- The program provides year-round support over many years (86.99%).
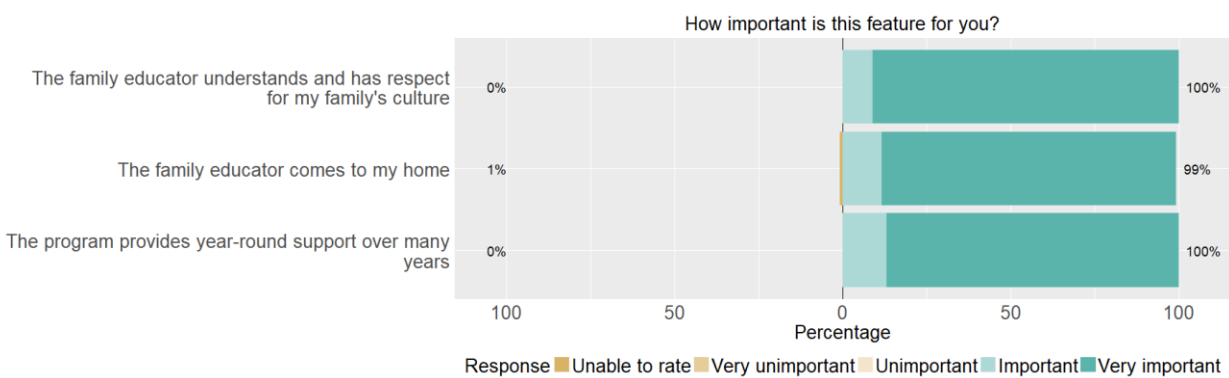
Figure 9. From the parents' perspective on the family educator: Reported level of importance of program features

The program components with the fewest respondents selecting *very important* were:

- The family educator is bilingual and can converse with me in my native language (70.73%), and
- The program helps me to meet my basic needs such as housing, food, and medical attention (71.54%)

It should be noted that 13 respondents (10.57%) selected *unable to rate* for the bilingual question, likely because they are native English speakers or fluent in English.

The next section presents survey results that help to answer the third implementation research question:

3. How well do family educators relate to participating parents and their families? What aspects of the parent-family educator relationship need improvement?

Respondents were asked to rate seven items related to their interactions with their family educator, using the options *never, seldom, usually, and always.* Respondents also had the choice of selecting *unable to rate.* Over 90% of respondents selected *always* for all seven aspects of the parent-family educator relationship, indicating a very high level of satisfaction in this area. The aspects with the highest percentage of respondents selecting *always,* as displayed in Figure 10, were:

- My family educator explains things using language that I can easily understand (95.12%),
- My family educator respects my family and my culture (94.35%), and
- My family educator answers my questions thoroughly (94.35%).
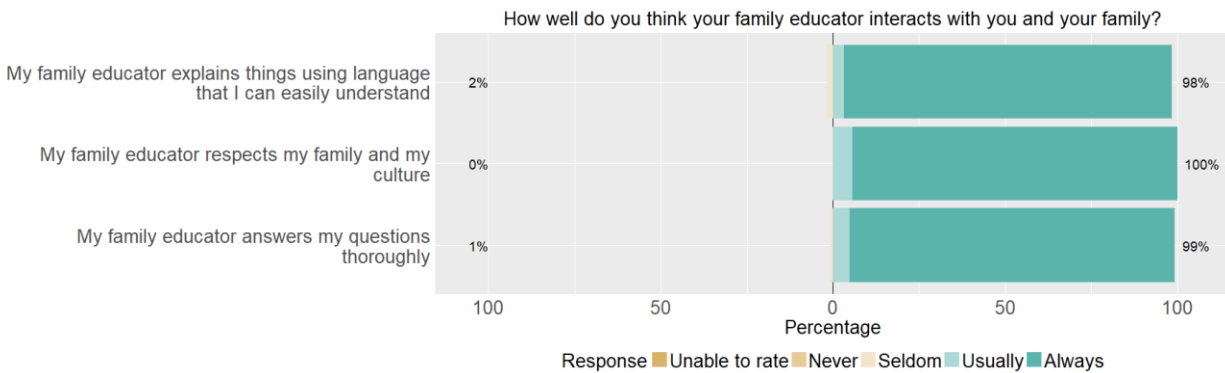
Figure 10. Parents' satisfaction with specific aspects of their interactions with family educators

The survey also asked respondents to rate the overall quality of their relationship with the family educator, using the options *unable to rate, poor, fair, good,* and *excellent.* Nearly 90% of respondents selected *excellent* (88.43%). Only one respondent (0.83%) selected *fair,* with no respondents selecting *poor* or *unable to rate* (see Figure 11).
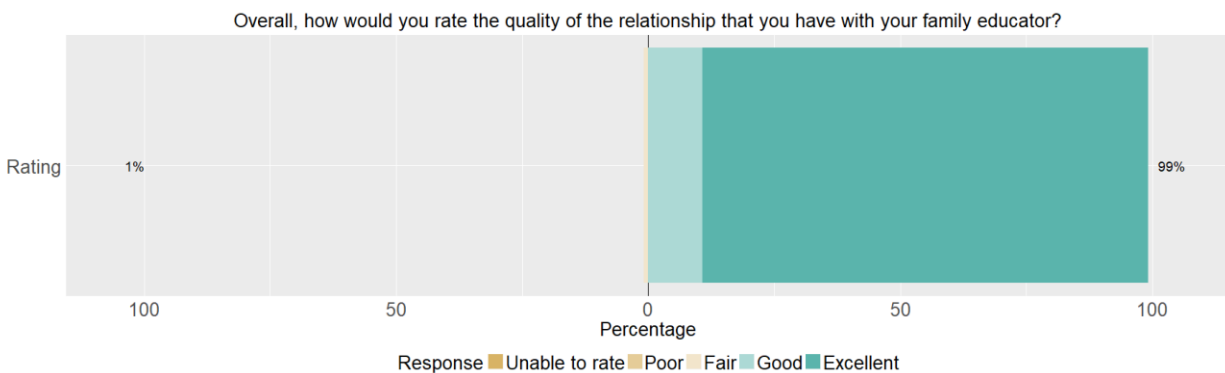
Figure 11. Parents' rating of overall quality of relationship with their family educator

In the revised implementation evaluation SEP, the data sources for answering the third implementation question were the parent survey *and* the Home Visit Rating Scales (HOVRS). Evaluators and Way to Grow staff proposed to triangulate survey responses that speak to question three with supervisor ratings of the family educator on the HOVRS. For example, ratings given by

supervisors to the aspect of "Home visitor relationship with family" would be compared to parents' quality ratings of their perceived relationship with their family educator. The HOVRS was a newly adapted standardized tool for evaluation to help validate parent engagement codes and collect data on home visitor interactions.

Because of the lack of HOVRS data available, we must depart from the revised SEP and rely on just the parent survey. Between mid-March 2014, and the end of June 2017, Way to Grow's database for the HOVRS shows only 47 observations by supervisors of family educators in parents' homes. Most of the 47 observations show incomplete ratings data, that is, only seven of the 47 have complete rating data across the six observation rating scales. Of the 47 observations, just 13 were made of parents who also had surveys. And of those 13, only two had complete data across the six observation rating scales (Facilitation of Parent Child Interaction, Responsiveness to Family, Relationship with Family, Non-intrusiveness, Parent-Child Interaction, and Parent-Engagement). The evaluators did not triangulate survey data with only two completions across the six observation scales.

The next section presents findings that help to answer the fourth implementation research question:

4. What is the home visit dosage (frequency, spacing, duration) currently received by parents participating in the Great by Eight program? How does the actual dosage compare to the dosage recommended in the program guidelines?

The data source for responding to this final implementation research question was the Way to Grow's programming longitudinal database. During the 5/1/16 to 12/31/17 timeframe of this final evaluation, family educators were extremely busy with 18,401 client/family contacts. Home visit dosage, however, refers only to the in person early learning visits ($n = 8,967$) and elementary program visits ($n = 4,468$), and not, for example telephone visits, email contacts, group based visits, and so on. And for reasons of continuity in responding to sampling designs, reporting implementation and impact results, as well as the practicality of handling data and analyses, evaluators chose to limit the analysis of frequency, duration and spacing to the 100 sampled families for the linear mixed-effects (LME) models analyses.

Frequency of home visits. Of the 100 sampled families for the LME analysis (discussed in the Impact Evaluation section of this report), 30 had no early learning and/or elementary program home visits recorded during this final implementation evaluation reporting period. Thus, the number of families in this implementation analysis for the time period between 5/1/16 to 12/15/17 was reduced to 70. The range of visits (frequencies) per family for this time period was from 1 to 98 visits for the sample during the reporting period. As shown in Table 2, the mean number of visits per family during the 84 weeks of the reporting period was 21; the median was 20, but the mode of families per frequency was 1.

Table 2

Frequency of Home Visits for Sampled Families (N = 70)

Total number of visits	1458
Mean visits per family	20.7
Standard deviation	18.8
Median frequency per family	20
Mode of families per frequency	1

Duration of home visits. The range of minutes per family home visit was from 5 to 200 for the sample. As shown in Table 3, the mean number of minutes per home visit was 48.5; the median was 45 minutes, and the mode was 60 minutes per visit.

Table 3

Duration in Minutes of Home Visits for Sampled Families (N = 70)

Total number of visits	1458
Mean minutes per visit	48.5
Standard deviation	21.8
Median minutes per visit	45
Mode of minutes per visit	60

Spacing between visits. The spacing between visits for this sample during this timeframe cannot be usefully captured in any meaningful number such as a mean number of days. We thus depart from the SEP and will not pursue it. For example, one family had 17 visits – all elementary program home visits – in the timeframe; the mean space between visits was 35 days, but the standard deviation was also 35 days with a range from 4 to 109 days. A second example: another family had 7 home visits, all from the early learning category. The mean space between visits was 11 days, but the standard deviation was 10 days with a range from 0 (two visits on the same day) to 28 days. Were we to analyze these two families we would find a mean of 27 days between visits and a standard deviation of 32 days with a range from 0 to 109.

Summary of findings and some limitations

Parent survey. For all topics concerning children's school-related and health-related outcomes that are presented by family educators during home visits, at least 70% of respondents reported having learned *a lot* and that the topics were *very useful,* indicating a high overall level of reported learning and perceived usefulness. To be sure, there was variation depending on the parent's child's age group: birth to 2 years old, three to five year olds, or kindergartners through third graders.

Evaluators also compared selected Life Skills Progression Scale (LSP) data with the parent survey questions about amount learned and usefulness of topics, in order to explore triangulation of the two instruments. We calculated Spearman correlations between the six LSP variables and the parent survey items. Overall, most of the correlations between LSP variables and parent survey items were small. We explored several potential explanations for why greater triangulation was not reached.

The survey asked respondents to rate the importance of six aspects of the Great by Eight program using the options *very unimportant, unimportant, important,* and *very important*. Respondents could also select *unable to rate.* For all program components included in the question, at least 70% of respondents selected *very important,* indicating a high overall level of perceived importance of the different aspects of the program.

Respondents were asked to rate several items related to their interactions with their family educator, using the options *never, seldom, usually, and always*. Respondents also had the choice of selecting *unable to rate.* Over 90% of respondents selected *always* for all seven aspects of the parent-family educator relationship, indicating a very high level of satisfaction in this area.

The survey also asked respondents to rate the overall *quality* of their relationship with the family educator, using the options *unable to rate, poor, fair, good,* and *excellent.* Nearly 90% of respondents selected *excellent*.

Home visit dosage. During the 5/1/16 to 12/15/17 timeframe of this final evaluation, family educators were extremely busy with 18,401 client/family contacts. In this report, home visit dosage refers only to the early learning visits ($n = 8,967$) and elementary program visits ($n = 4,468$), and not, for example telephone visits, e-mail contacts, group-based visits, or the like. And for reasons of continuity in reporting implementation and impact results as well as the practicality of handling data and analyses, evaluators chose to limit the analysis of frequency, duration and spacing to the 100 sampled families for the linear mixed-effects (LME) models analyses.

Frequency of home visits. Of the 100 sampled families, the number of families in this analysis with one or more visits was 70. The range of visits (frequencies) per family was from 1 to 98 visits for the sample during the reporting period with a mean of 21 visits; the median was 20, but the mode of families per frequency was 1.

Duration of home visits. The range of minutes per family home visit was from 5 to 200 for the sample. The mean number of minutes per home visit was 48.5; the median was 45 minutes, and the mode was 60 minutes per visit.

Spacing between visits. We discovered that the spacing between visits for this sample during this timeframe cannot be usefully captured in any meaningful number(s) such as a mean number of days.

Impact Evaluation

This final evaluation research included two impact studies as described in the SEP. Per the SEP, the **confirmatory research questions** were

1. Are Great by Eight children more prepared for kindergarten than non-program children?

2. Are Great by Eight children more prepared to test proficient on the third grade Minnesota Comprehensive Assessment (MCA) than non-program children?

3. Do parents show increased engagement with their children's development, learning, and schooling?

4. Does child development and growth progress achieve appropriate milestones and in conjunction with preparation and success for formal schooling?

5. Do Great by Eight health components (health referrals and education measured by dosage) improve and sustain health promoting behaviors?

6. Do Great by Eight parents and children access resources and navigate health and education systems?

Study One: Kindergarten Readiness and Third Grade Proficiency

Study One addresses the first two impact research questions. Great by Eight kindergartners and third graders were matched with similar children enrolled in the Minneapolis Public School (MPS) District to study the program's effectiveness on educational progress. Kindergarten children were compared on the Minneapolis Beginning Kindergarten Assessment (BKA), a standardized assessment of literacy skills administered in the fall of the kindergarten year by assessors. Third graders' academic achievement was measured using Minnesota Department of Education's standardized Minnesota Comprehensive Assessment (MCA) reading test. Descriptions including psychometric characteristics of the BKA and MCA are detailed in Appendix B.

Samples

All Great by Eight kindergartners who entered kindergarten in the fall of 2017 in Minneapolis Public Schools were included in the program evaluation ($n = 59$). All Great by Eight third graders who entered third grade in the fall of 2015 or the fall of 2016 in Minneapolis Public Schools were included in the program evaluation ($n = 72$): 29 from the 2015-16 school year and 43 from 2016-17 school year.

The Minneapolis Public School's Department of Research, Evaluation, Assessment, & Accountability (REAA) identified comparison kindergarten and third grade children for the analysis. Propensity score matching procedures, based on a set of covariates related to the study's outcome, were used to select children as similar as possible to the Great by Eight samples. The

matching process was conducted using the R Matching statistical package. The package provides functions for multivariate and propensity score matching and for finding optimal balance between intervention and non-intervention groups. In general, R functions determine a comparison group match and evaluate how good covariate balance is before and after matching. The GenMatch function finds optimal balance using multivariate matching with a genetic search algorithm that assigns weight to each covariate. Balance is determined by implementing a variety of standardized statistics including paired *t*-tests, univariate and multivariate Kolmogorov-Smirnov (KS) tests (Sekhon, 2011).

The evaluation team worked closely with the Way to Grow staff to identify key covariates that best represented program children's characteristics for comparison group matching. Additionally, the variables were thought to influence academic outcomes. The following list discusses these covariates and rationales for including them in comparison group development.

The following covariates were used to form comparison groups for both kindergartners and third graders.

Free or Reduced-Price Lunch Status (as an imperfect proxy for socio-economic status). Approximately 99% of Great by Eight families live below national poverty levels. An indicator of living below federal poverty guidelines for school children is receiving free or reduced lunch through the National School Lunch Program. Eligibility is based on household size and family income. Children in the intervention group were compared to children with similar free or reduced lunch status.

Ethnicity. Great by Eight program children come from diverse backgrounds. Approximately 25% of the children are Hispanic/Latino, 15% are Asian or Hmong, 24% are African American, 35% are Somali, and 1% is Native American. Matching techniques equated kindergarten and third grade children with similar ethnicity backgrounds.

Primary Language. Similar to ethnicity, primary language reflects cultural background and the language most often used in homes. English is a secondary language for a large percentage of program children. Great by Eight children were matched to comparison groups with similar primary language characteristics.

Gender. To accommodate any differences between groups on gender, this variable was included in the matching procedures.

Age. Children vary on age across any grade level. For example, some kindergartners enter school after just turning five while others are months ahead of them. Age is an indicator of social, cognitive, and physical development in early childhood and, therefore, was included as a covariate for matching.

School affiliation (as an imperfect proxy for geographical location). Approximately 90% of Great by Eight children are enrolled in Minneapolis Public Schools. Geographical location is often an indicator of socioeconomic status, access to support systems, and quality of life. Way to Grow children were matched with students that attended the same schools.

Special education status (sometimes called exceptional students education status). Special education status may influence academic progress and, therefore, was included in the matching procedure.

English language user. Some of the Great by Eight children are in special English language learner programs. These program children were matched with children enrolled in similar programs.

Final Sample Characteristics

Minneapolis Public School's Department of Research, Evaluation, Assessment, & Accountability (REAA) provided program evaluators with the propensity score matching analysis and results. A variety of univariate and multivariate metrics showed covariate balance had been obtained between intervention and identified comparison groups after the matching process. These statistics included paired *t*- tests and Kolmogorov-Smirnov (KS) tests. The tests showed no significant difference on key covariates between groups. A variety of descriptive statistics based on empirical-QQ plots were also provided. Program evaluators were given de-identified files on Great by Eight's kindergartners and third graders and their final matched counterparts. They subsequently performed statistical analyses to substantiate comparability among control and program children. The following discussion reports these descriptive findings.

Kindergartners. These children entered kindergarten in the fall of 2017. The mean age of the Great by Eight kindergarten sample was 5.56 years and the comparison group was 5.54. Age was the child's age on the first day of school. Percentages of males and females were reasonably equivalent between the two groups (Table 4).

Table 4

Distribution of Kindergarten Children's Gender (N = 118)

Gender	Great by Eight		Comparison Group	
	N	%	*N*	%
Female	30	51	34	58
Male	29	49	25	42
Total	59	100	59	100

The largest ethnic group among the kindergarten children in the Great by Eight program (44%) and in the comparison group (43%) was African American[3]. The second largest was Hispanic[4] children with 34% in the intervention group and 37% of their matched counterparts. Table 5 shows program and comparison kindergartners' ethnicity distribution.

---

[3] African, African American, and Black students comprise the African American category.
[4] Hispanic and Latino students comprise the Hispanic category

Table 5

Distribution of Kindergarten Children's Ethnicity (N = 118)

Ethnicity	Great by Eight		Comparison Group	
	N	%	*N*	%
African American	26	44	25	43
Hispanic	20	34	22	37
White	6	10	4	7
American Indian	5	9	6	10
Asian	2	3	2	3
Total	59	100	59	100

The largest primary home language group was English – 44% among the Great by Eight kindergartners and 41% among the comparison children. Spanish was the second largest home language and Somali third. Table 6 displays the kindergartner's primary home language distribution.

Table 6

Distribution of Kindergarten Children's Primary Home Language (N = 118)

	Great by Eight		Comparison Group	
Primary Home Language	*N*	%	*N*	%
English	26	44	24	41
Spanish	19	32	20	34
Somali	11	19	12	20
Hmong	2	3	2	3
Other/Unknown	1	2	1	2
Total	59	100	59	100

Table 7 presents three other key covariate distributions among intervention and nonintervention kindergarten children.

Table 7

Distribution of Kindergarten Key Covariates

	Great by Eight		Comparison Group	
Covariate	*N*	%	*N*	%
Eligible for Free or Reduced-Price Lunch	56	95	55	93
English Language Learner	23	39	24	41
Receiving Special Education	9	15	8	14

Table 7 shows that Great by Eight kindergartners and their matched comparison group had similar characteristics on the three key covariates. Family socio-economic status is reflected in children qualifying for free or reduced-price lunch in public school. More than 90% of program and comparison children were receiving this service. Thirty-nine percent of program and 41% of comparison children were identified as English language learners, and about 15% of both groups were receiving special education services.

Given the high comparability between the two groups based on these demographic characteristics, we find that the Great by Eight kindergartners and their comparison group well-matched (covariate balance had been obtained) for statistical comparison purposes.

Third Graders. There were 72 Great by Eight third graders who participated in the study: 29 from the 2015-16 school year and 43 from 2016-17 school year. Propensity score matching procedures identified 72 comparable third graders (the comparison group), and again, 29 matched from the 2015-16 school year and 43 matched from 2016-17 school year. The mean age of the Great by Eight total third grade sample was 8.44 years, and the total third grade comparison group was 8.47. Age was the child's age on the first day of school. Percentages of males and females were reasonably equivalent between the two groups (Table 8).

Table 8

Distribution of Third Grade Children's Gender (N = 144)

Gender	Great by Eight		Comparison Group	
	N	*%*	*N*	*%*
Female	29	40	25	35
Male	43	60	47	65
Total	72	100	72	100

Fifty percent of the third graders in the Great by Eight program and the comparison group were African American. Hispanic children showed the second highest ethnic percentage. A small percentage of American Indian and Asian children were identified in each group. There were no children identified as White in either sample. Table 9 shows program and comparison third graders' ethnicity distribution.

Table 9

Distribution of Third Grade Children's Ethnicity (N = 144)

Ethnicity	Great by Eight		Comparison Group	
	N	*%*	*N*	*%*
African American	36	50	36	50
Hispanic	30	42	31	43
American Indian	3	4	3	4
Asian	3	4	2	3
Total	72	100	72	100

The largest home language group was Spanish. Somali was the second largest and English third. The table below displays primary language distributions among the children.

Table 10

Distribution of Third Grade Children's Primary Home Language (N = 144)

Primary Home Language	Great by Eight		Comparison Group	
	N	*%*	*N*	*%*
Spanish	29	40	30	42
Somali	18	25	16	22
English	17	24	18	25
Other/Unknown	5	7	6	8
Hmong	3	4	2	3
Total	72	100	72	100

Great by Eight third graders and their matched comparison group had similar characteristics on three other key covariates. One hundred percent of program and comparison children qualified for free or reduced-price lunch in public school. Sixty-three percent of program and 67% of comparison children were identified as English language learners. And 17% of program and 14% of comparison children were receiving special education services. Table 11 presents the three other key covariate distributions among intervention and nonintervention third grade children.

Table 11

Distribution of Third Grade Children's Key Covariates

Covariate	Great by Eight		Comparison Group	
	N	%	N	%
Eligible for Free or Reduced- Price Lunch	72	100	72	100
English Language Learner	45	63	48	67
Receiving Special Education	12	17	10	14

Given the high comparability between the two groups based on these demographic characteristics, we find that the Great by Eight third graders and their comparison group well-matched (covariate balance had been obtained) for statistical comparison purposes.

Measurement Instruments

The program evaluation compared kindergarten Great by Eight children and their matched counterparts on the total literacy score of the Minnesota Beginning of Kindergarten Assessment. Great by Eight third grade children were compared to similar third graders on the Minnesota Department of Education's standardized Minnesota Comprehensive Assessment reading test.

Descriptions including psychometric characteristics of the measurement instruments are detailed in Appendix B.

Procedures

The evaluation team received approval to conduct the study from the University of Minnesota's Institutional Review Board and MPS's REAA in February 2014, and the approvals have been renewed approximately every two years. The evaluation plan requested academic literacy measures for Great by Eight kindergartners (BKA) and third graders (MCA reading). Additionally, REAA would provide the service of identifying comparison groups for Great by Eight children through propensity score matching techniques. The same academic scores for comparison children would also be provided to the evaluation team.

Third grade MCA scores for the 2016-17 school year became available in late summer of 2017. The Way to Grow staff and evaluation team submitted rosters of Great by Eight's 2016-17 kindergartners and 2015-16 and 2016-17 third graders to REAA. REAA conducted matching procedures and identified comparison groups for the Great by Eight children. When the process was completed, de-identified files for both Great by Eight and comparison children were given to the evaluation team. BKA and MCA literacy scores were provided in the files along with some demographic and performance information in the children's records.

Analysis

The independent variable for data analysis was Great by Eight program involvement versus non-program participation. The dependent variable for kindergartners was the BKA total literacy scores and, for third graders, MCA reading scores. The first step in the analysis was to calculate test score descriptive statistics for the intervention and comparison groups. These statistics included means and standard deviations. Percentages of children meeting or exceeding Minneapolis Public School's assessment standards were also calculated.

An independent samples t-test was used to determine if there were significant differences among test scores between Great by Eight children and their comparison counterparts. The analysis was conducted using IBM SPSS Statistics 22 software. Assumptions that supported the use of independent t-tests to measure program effects were: 1) the technique compares the means of two independent samples; 2) due to intervention and comparison group matching techniques, the variance among the groups should be similar; and 3) samples sizes were similar. When appropriate, effect sizes (*Cohen's d*) were calculated to measure the strength of the relationship between independent and dependent variables.

Study One Findings

Kindergartners

The confirmatory impact research question was

Are Great by Eight children more prepared for kindergarten than non-program children?

Unfortunately, only 31 (53%) of the 59 Great by Eight kindergartners and 27 (46%) of the 59 kindergartners in the comparison group had BKA scores. Table 12 also displays the numbers and percentages of those with BKA scores who were categorized as "prepared for K," that is, their total BKA score was 50 or higher. The kindergartners without BKA scores were probably not "data missing at random" in that, and as reported above, in the fall of 2017, the BKA was not administered district-wide but was administered optionally as schools in the district chose. Further, MPS's REAA was not allowed by their legal department to share the names of the schools in which kindergartners were enrolled in kindergarten.

Table 12

BKA Total Score and Percent Categorized as Prepared for K*

Group	*N*	*M*	*SD*	*N* scored prepared	% scored prepared
Great by Eight	31	44.65	12.04	11	35.5
Comparison	27	40.96	14.52	6	22.2

*50 or higher on BKA total.

Although the Great by Eight group had a higher mean score, the result of an independent samples *t*-test comparing Great by Eight and matched group kindergartners' mean BKA scores was not statistically significant at an $\alpha \leq .05$: $t(56) = 1.06$, $p = .296$. Note further that the mean for each group was categorized as "not prepared for K." Further, the 13.3% difference between the percent of Great by Eight students rated prepared on the BKA compared with the matched students was not statistically significant (Chi-squared = 1.211, $df = 1$, $p = .271$).

Third Graders

The impact confirmatory research question was

Are Great by Eight children more prepared to test proficient on the third grade Minnesota Comprehensive Assessment (MCA) than non-program children?

Academic progress of the Great by Eight third graders was measured using Minnesota Department of Education's standardized Minnesota Comprehensive Assessment (MCA) reading assessment. This assessment is administered during the spring of the third-grade year. Table 13 provides means and standard deviation for program children and their comparison group on the reading measure.

Three (4%) Great by Eight third graders and nine (13%) comparison group third graders were missing MCA reading data.

Table 13

MCA Means and Standard Deviations for Third Grade Reading Assessment

	Great by Eight (N = 69)		Comparison Group (N = 63)	
	M	*SD*	*M*	*SD*
MCA Reading Assessment	333.94	19.50	326.59	20.60

Similar to kindergarten assessments, MPS has established standards on academic performance related to third-grade reading skills. A score of 350 and above indicates that children have met or exceeded the reading test benchmark. The average score for both Great by Eight and comparison group students did not meet this standard. Rather, Great by Eight children's mean reading score was categorized as "did not meet" the benchmark (339 and below) and the comparison group's average also "did not meet" the district's third grade reading expectation. Table 14 displays the percentages of children in both groups at each reading standard classification.

Table 14

Percentage of Great by Eight and Comparison Third Grade Children at Reading Standard Classifications

MCA Reading Assessment Standards	Great by Eight (N = 69)		Comparison Group (N = 63)	
	N	%	*N*	%
Meets or exceeds (350+)	14	20	10	16
Partially meets (340-349)	15	22	4	6
Does not meet (339 and below)	40	58	49	78

An independent samples *t*-test was applied to the data to determine if a significant difference between program and comparison group reading scores was present. Table 15 shows the results of the statistical analysis. The analysis shows that there was a statistically significant difference ($\alpha \leq .05$) between groups on reading assessment scores with Great by Eight children scoring higher than comparison children (Table 13). A *Cohen's d* effect size calculation found a small to moderate program intervention impact.

Table 15

Independent t-test Results and Effect Size Comparing Great by Eight and Matched Group Third Graders on Reading Assessment

	t	*df*	*p*	*d*
MCA Reading Assessment	2.107	62	.037	0.37

Study One Summary of Findings and Limitations

Study one was designed to analyze the impact of the Great by Eight home visiting program on kindergarten readiness and reading proficiency. One strength of the investigation was that comparison groups for Great by Eight children were formed through propensity score matching. The result of an independent samples *t*-test comparing Great by Eight and matched group kindergartners' mean BKA total scores was not statistically significant. For third grade reading proficiency, however, an independent samples *t*-test showed that there was a statistically significant difference between groups on reading assessment scores, with Great by Eight children scoring higher than comparison children. Further, a *Cohen's d* effect size calculation found a small to moderate program intervention positive impact. Finally, 42 percent of Great by Eight children compared with 22 percent of comparison group children *met* or *exceeded* or *partially met* the reading proficiency score.

The program evaluation had some limitations. First, the kindergartener sample sizes were small – 31 in Great by Eight and 27 comparison children with BKA scores. If the propensity subclasses become too small, statistical analysis of outcomes may lead to unstable estimations, which would influence the validity of findings (Fan & Nowell, 2011).

Another possible limitation in a quasi-experimental design using propensity score matching techniques is the omission of important covariates that could lead to bias for statistical comparisons. Although the evaluation team and the Way to Grow staff identified key covariates for comparison group matching, there may have been critical covariates that could not be identified and considered in the analysis. For example, the comparison group's history may have been a potential problem and a threat to internal validity of the study. Great by Eight children and their parents were receiving the program's home visiting services. The program also connects families to educational and community support. However, we were not able to determine interventions or support services that the comparison groups were or had been receiving.

Study Two: Changes and Progress in Children and Parents

Design

The program evaluation used linear mixed-effects models to examine changes in parents based on multiple measures in a longitudinal format. The evaluation was specifically interested in exploring and confirming changes based on home visitation interventions related to increased parent engagement with their children's development, learning, and schooling. Specific research questions are listed below.

Confirmatory Research Questions

1. Do parents show increased engagement with their children's development, learning, and schooling?

2. Does child development and growth progress achieve appropriate milestones and in conjunction with preparation and success for formal schooling?

3. Do Great by Eight health components (health referrals and education measured by dosage) improve and sustain health promoting behaviors?

4. Do Great by Eight parents and children access resources and navigate health and education systems?

Design

A linear mixed-effects (LME) model was used to investigate patterns of change (trajectories) among parent measurements. Mixed models use both fixed and random effects. Fixed effects represent, on average, how subjects change over time on a dependent variable while individual variation is generally considered and inputted as a random effect in the model. Mixed models provide a useful and flexible means for the study of longitudinal data (Fitzmaurice, Laird, & Ware, 2011; Singer & Willett, 2003), where continuous repeated measurements are taken on subjects on multiple occasions.

The LME model has several advantages over more conventional methods such as repeated measures analysis of variance, ANCOVA, and regression models. First, multiple measurements per subject generally result in correlated data which violates assumptions of standard (between subjects) ANCOVA and regression. A unique feature of mixed model analysis is that covariance structures can be partitioned into between-subject and within-subject components. The LME allows a wide variety of correlation patterns (or variance-covariance error structures) to be explicitly modeled in the data analysis (Maxwell & Delaney, 2004). Second, the LME methodological framework allows both population and individual patterns of change to be considered in the model (Blozis & Cudeck, 1999). Third, traditional repeated measure approaches discard results on any subject when there are missing measurements. LME models allow any data on a subject to be used as long as the missing data meets the missing-at-random (MAR) definition. Finally, and importantly, another advantage of mixed models is that they naturally handle uneven intervals or spacing of repeated measurements (Singer & Willett, 2003).

Sample

Per the SEP, the sample for the first, third and fourth impact research questions was the random sample of 100 families from the group of children aged three to eight years participating in the Great by Eight program and their parent(s) selected to participate in the linear mixed-effects (LME) evaluation plan. These are the ages of interest for the SIF-supported Way to Grow programming in which family educators are delivering home visits for early learning and elementary program curricular visits.

For the second research question – child development and growth toward school readiness – evaluators faced a challenge. We wanted to restrict a sample to the same timeframe of the other analyses and include children with at least two testing rounds in order to chart change. Unfortunately, this meant that among the 100 sampled families/children, just seven had data in the timeframe and with at least two testing rounds. Not wanting to abandon the research question, we compromised with *all* four-year olds in the timeframe with at least two assessments. This yielded 98 children for the analysis.

Outcome Measures

Parent engagement. The construct of parent engagement was measured with the Life Skills Progression (LSP) Scale, an observational tool completed by family educators that summarizes a parent's skills in parent-child relationships, education and employment, health and medical care, mental health and substance abuse, and basic essentials (Wollesen & Peifer, 2006). The LSP describes individual parent and children's progress using 43 types of life skills, which are grouped into five categories. Indicators in the "Relationship with Children" LSP category were used to assess parent engagement with their children over time. These scales included 1) Nurturing (Skills and Ability); 2) Discipline (Appropriateness); 3) Support of Child Development; and 4) Safety (protecting children from environmental harm). The Way to Grow staff indicated that their home visiting program emphasized three additional parental engagement characteristics that are measured through the LSP. These scales were Use of Resources, Child Care (quality of child care environment), and Cognitive Ability (cognitive understanding of children's needs). In total, seven LSP repeated measures were used to assess longitudinal progress of parent engagement with their children. Each scale is scored independently across a range of 0-5 points, using 0.50-point increments.

Descriptions including psychometric characteristics of measurement instruments used for parent engagement (LSP, Parent Engagement Code, Parent Involvement with Children's School programming data) are detailed in Appendix B.

Time. In the original data, the date each LSP assessment was completed served as the time variable. There was a total of 135 different dates. To simplify the analysis, the dates were recoded into "measurement time 1," "measurement time 2," and so on. Each family had an average of two assessments per year, which were spaced approximately six months apart. This average time between measurements was taken into account during the re-coding. For example, if a family's second rating was taken a year after the first one, it would be re-coded as "measurement time 3" instead of "measurement time 2." The number of measurement times for each family ranged from one to eight, with most families having three or four measurements recorded during the evaluation period. As mentioned above, one advantage of a linear mixed-effects model is that this range in number of measurements is not problematic for the analysis.

This variable was then re-coded again by subtracting 1, so that measurement time = 0 refers to the first measurement. This last step meant that the estimated intercepts could be interpreted as the score at the first measurement.

Covariates. Three covariates were included to see if they could account for differences between parents' LSP scores: number of home visits, primary home language, and ethnicity.

Number of home visits. This variable represents a count of how many home visits (early learning and/or elementary program) each family had received over the time they have been in the program. These visits, conducted by Way to Grow family educators, were the main program intervention. Among the 100 families that were randomly selected for this impact analysis, the number of visits in the time frame ranged from 2 to 201, with the average being 36 visits.[5]

Primary home language. For primary home language, the variable was dichotomized for those who speak English at home (coded 1) and those who did not (coded 0). In the random sample used in this analysis, 34 families spoke English at home and 65 had a different primary home language – data was missing for 1 family.

Ethnicity. Table 16 lists the ethnicity breakdown for the sample. The ethnic groups for parents were Hispanic or Latino, African, African American, Asian, American Indian/Native Alaskan, and White. For analysis purposes only, White parents was used as the reference group to compare observed differences for the other ethnicities that are representative of the programs' participants, as shown in the tables in this section.

Table 16

Parents' Ethnicity

Ethnicity	N
Hispanic or Latino	29
African	25
African American	23
Asian	11
American Indian/Native Alaskan	7
White	5
Total	100

Analysis

The LME analysis was conducted using the lme4 and nlme packages in the R statistical software package. The analysis consisted of examining multiple repeated measures on parent engagement with their children gathered over time.

In building a mixed-effects model for each outcome variable, the first step was to determine how many fixed effects and random effects to model. As with regression, estimates for both the

---

[5] Note, number of home visits reported for the impact analysis include all data for families over time, whereas descriptive data reported for implementation analysis include only fourth year data per the SEP.

intercept and slope are produced. However, with mixed-effects models the researcher must determine whether the intercept and slope are fixed or random. A fixed effect means that the estimate is the same for all individuals in the population and is analogous to the type of parameter estimated in regression. A random effect means that there is enough variation across the population that its estimation must include both a fixed component and a variance component. The number of random effects can be determined by comparing the fit indices of a random intercept model (random intercept, fixed slope), a random slope model (fixed intercept, random slope), and a random coefficient model (random intercept, random slope).

For the Nurturing, Discipline, Child Care, and Cognitive Ability LSP outcome variables, a random coefficients model was most appropriate, meaning that the analysis accounted for individual variation around both the intercept and slope. In other words, individuals varied in terms of their score at the first measurement time (intercept), as well as in their rate of change in score over time (slope). For the Support of Development, Safety, and Use of Resources LSP outcome variables, a random intercepts model was most appropriate, meaning that the analysis accounted for individual variation in the intercept but not the slope. In other words, individuals varied in terms of their score at the first measurement time but not in their rate of change over time.

The next step in the analysis was to determine which error structure to use for the individual time-specific residuals. The researchers opted to use an autoregressive error structure, which is commonly used when measurements are ordered and adjacent measurements are more highly correlated than distant measures (Maxwell & Delaney, 2004).

The final step was to add in covariates to reduce between-subject variability in scores. Because there was much more variation in the intercepts (i.e., parents' scores at the first measurement time) compared to the slopes (i.e., rate of change in scores over time) for all seven outcome variables, interactions between the covariates and the time variable were not included. One exception was the number of visits variable; for each outcome variable, the evaluators tested whether to include an interaction between time and number of home visits, but the coefficients for the interaction were not statistically significant. In other words, parents with a greater number of home visits did not demonstrate a higher rate of change in their rating over the time they were in the program. The covariates were thus used only to explain differences in intercepts and not slopes.

Study Two Findings

Parent Engagement

The impact confirmatory research question was

1. Do parents show increased engagement with their children's development, learning, and schooling?

Change in parent engagement over time. The mean value for each LSP variable at each time point is reported in Table 17. At all time points, including the first measurement time, the mean value for each variable except for Child Care was between 4 and 5. Note well that because parents were rated so highly from the first time they were assessed, there was little room for the average

score to increase over time. With the exception of Child Care, all means were higher at measurement time 7 (the last measurement time with more than 1 score recorded) compared to measurement time 1, but the increases were very small in magnitude. The magnitude of change for Child Care was slightly higher between measurement times 1 and 6.

Table 17

Means for Life Skills Progression Scales over Time

	Time 1 (*n = 100)*	Time 2 (*n = 76)*	Time 3 (*n = 58)*	Time 4 (*n = 50)*	Time 5 (*n = 41)*	Time 6 (*n = 27)*	Time 7 (*n = 21)*	Time 8 (*n = 1)*
Nurturing	4.48 (.75)	4.53 (.78)	4.57 (.61)	4.50 (.63)	4.56 (.54)	4.67 (.46)	4.69 (.43)	4.00 (NA)
Discipline	4.35 (.87)	4.58 (.55)	4.39 (.61)	4.43 (.61)	4.39 (.62)	4.59 (.61)	4.48 (.72)	4.00 (NA)
Support of Development	4.23 (.87)	4.31 (.90)	4.26 (.91)	4.31 (.83)	4.30 (.81)	4.52 (.71)	4.45 (.86)	4.00 (NA)
Safety	4.58 (.75)	4.79 (.44)	4.72 (.49)	4.77 (.48)	4.72 (.50)	4.81 (.48)	4.67 (.48)	5.00 (NA)
Use of Resources	4.41 (.76)	4.60 (.69)	4.27 (.79)	4.49 (.83)	4.41 (.89)	4.63 (.64)	4.67 (.66)	5.00 (NA)
Child Care	2.96 (1.79)	3.12 (1.75)	3.29 (1.72)	3.44 (1.79)	3.33 (1.70)	3.39 (1.59)	2.55 (1.71)	1.00 (NA)
Cognitive Ability	4.80 (.46)	4.85 (.64)	4.85 (.32)	4.89 (.27)	4.88 (.33)	4.91 (.28)	4.95 (.22)	5.00 (NA)

Note: Means are reported with standard deviations in parentheses. Each scale is scored across a range of 0-5 points using 0.5-point increments.

Effect of covariates. Tables 18 and 19 report the parameter estimates for each model. Table 18 includes the four random coefficient models (random intercept, random slope), and Table 19 includes the three random intercept models (random intercept, fixed slope). Because the research questions are focused on aggregate patterns, the fixed effects are the main variables of interest for interpretation. The variance of the random effects, which contains information about individual differences, is also reported.

Interpretation of fixed effects. The intercept estimate can be interpreted as the predicted value for an observation for which all values on all variables equal zero.[6] The measurement time coefficient, which represents the slope, is the estimated amount that the variable would change from one measurement time to the next. For each variable, the effect of time on the outcome is very small and not statistically significant. This mirrors the descriptive statistics reported above in Table 17, which show very little change over time in the outcome variables.

---

[6] i.e. A participant who identifies as White, non-English speaking, at the first observation.

The number of visits coefficient represents the effect of how many visits the family had received on their score at measurement time 1. The coefficient for number of visits was 0 or close to 0 for all LSP variables.

Table 18

Parameter Estimates for the Linear Random Coefficient Models

Parameter	Nurturing	Discipline	Child Care	Cognitive Ability
Fixed effects				
Intercept	5.05***	4.15***	2.02*	4.72***
	(.32)	(.36)	(.85)	(.22)
Measurement time	.03	.00	.10	.01
	(.02)	(.02)	(.05)	(.01)
Number of visits	.00	.00	.00	.00
	(.00)	(.00)	(.00)	(.00)
Ethnicity				
African	-.83*	.17	1.43	.03
	(.33)	(.37)	(.87)	(.23)
African American	-.36	.30	.67	-.37
	(.30)	(.33)	(.77)	(.20)
American Indian/Native	.19	.76*	1.63	-.17
Alaskan	(.34)	(.38)	(.90)	(.24)
Asian	-.90**	-.28	-.58	.06
	(.34)	(.39)	(.93)	(.23)
Hispanic or Latino	-.31	.39	.98	.04
	(.32)	(.36)	(.86)	(.22)
Primary language at home (1 =	-.39	-.25	.45	.32*
English is primary language)	(.21)	(.25)	(.60)	(.14)
Random effects variance				
Intercept	.58	.64	1.49	.40
Measurement time	.10	.11	.31	.04
Residual	.40	.37	.90	.31

Note. standard errors are in parentheses. Full maximum likelihood was used.
*** = $p < .001$; ** = $p < .01$; * = $p < .05$

The coefficient for each ethnic group represents the average difference on the outcome for a family of that ethnicity compared to the reference group (i.e., White parents). There were several ethnicity coefficients that were statistically significant. African families on average had
.83 points lower in their Nurturing score for the first measurement. The average scores at measurement time 1 for African American families were .78 points lower for the Support of Development scale and .68 points lower for the Use of Resources scale. American Indian/Native Alaskan families scored an average of .76 points higher on the Discipline scale at time 1. Finally, Asian parents on average scored .9 points lower for Nurturing, 1.32 points lower for Support of Development, and .81 points lower for Use of Resources.

There were two scales for which the primary language covariate had a statistically significant coefficient. Parents who spoke English as their primary language at home had an average of .32 points higher on their Cognitive Ability score at time 1, but .38 points lower on the Safety score.

Table 19

Parameter Estimates for the Linear Random Intercept Models

Parameter	Support of Development	Safety	Use of Resources
Fixed effects			
Intercept	4.73***	4.64***	4.56***
	(.37)	(.25)	(.36)
Measurement time	.03	.01	.01
	(.02)	(.01)	(.02)
Number of visits	.00*	.00*	.00*
	(.00)	(.00)	(.00)
Ethnicity			
African	-.61	.01	-.18
	(.38)	(.26)	(.37)
African American	-.78*	.13	-.68*
	(.34)	(.22)	(32)
American Indian/Native Alaskan	-.26	.29	-.27
	(.40)	(.26)	(.38)
Asian	-1.32**	-.45	-.81*
	(.41)	(.27)	(.40)
Hispanic or Latino	-.54	.11	-.19
	(.38)	(.25)	(.37)
Primary language at home (1 = English is primary language)	.00	-.38*	.11
	(.27)	(.18)	(.27)
Random effects variance			
Intercept	.49	.36	.52
Residual	.61	.37	.48

Note. standard errors are in parentheses. Full maximum likelihood was used.
*** = $p < .001$; ** = $p < .01$; * = $p < .05$

Change in between-subject variability. In longitudinal linear mixed-effects models, researchers can partition variability into within-subject variability (i.e., change in individuals over time) and between-subject variability (i.e., differences between individuals). The goal of adding covariates to the model is to account for differences in the outcome variable between individuals. It is therefore useful to calculate the reduction in between-subject variability between the null model that contains no covariates and the final model. Table 20 reports this change for each outcome variable. Safety and Support of Development had the greatest amount of reduction in between-subject variability (9.26% and 10%, respectively), whereas Discipline and Cognitive Ability had the least amount of change (1.72% and 1.85%, respectively).

Table 20

Reduction in Between-Subject Variability for each LSP Outcome Variable

	Between-subject variability: Null model with no covariates	Between-subject variability: Final model with covariates	Percent reduction in between-subject variability
Nurturing	.56	.54	3.57%
Discipline	.58	.57	1.72%
Support of Development	.50	.45	10%
Safety	.54	.49	9.26%
Use of Resources	.56	.52	7.14%
Child Care	.57	.55	3.51%
Cognitive Ability	.54	.53	1.85%

Limitations

Assumptions violation. When using a linear mixed-effects approach, it is important to check the following assumptions of the final models:

1. Each random effect has a normal distribution
2. The errors are normally distributed with a mean of zero
3. The errors are independent of each other

Some of the models had random effects and errors with a skew that was greater than +/-1, therefore demonstrating evidence of non-normality and violating assumptions 1 and 2. However, re-running the models with extreme outliers removed largely corrected this problem without significantly affecting the parameter estimates. One exception was the Discipline model, for which the version with outliers removed produced substantially different parameter estimates. The results for the Discipline model should thus be interpreted with caution.

Outcome variable. As noted above, the LSP ratings that parents received were consistently high, even during the earliest observations. This meant that the LSP instrument could not capture substantial change over time. The lack of statistically significant covariate coefficients – or, in the

case of the *number of visits* variable, coefficients that did not differ significantly from zero – may be due to the very small amount of variation in the outcome variables that could be explained in the first place. In the future, the program may want to consider using a different instrument that can capture a greater amount of change over time.

Parent Engagement Code. The Parent Engagement Code is an observational tool developed by Way to Grow and used by family educators to measure the quality of parent engagement during home visits. The tool has a five-point scale that is used to evaluate parent and child interactions with one being the lowest score and five being the highest. Family educators rate parent engagement during each home visit to monitor a parent's progress. These repeated observation scores on parents were provided in the longitudinal data base. Descriptive statistics indicated that the average score over time on parents was 4.51 with the standard deviation equal to 0.79.

Because this scale lacks the .5 increments that the LSP variables have, the evaluators determined that the Parent Engagement Code could not be treated as a continuous variable. Consequently, a linear mixed-effects model was not possible. Because the engagement ratings were heavily skewed towards a score of 5, the evaluators dichotomized the variable so that a score of 5 = 1 (n = 2,323) and a score of less than 5 = 0 (n = 1,279). A logistic mixed-effects model was used, which is similar to LME but uses a binary outcome variable.[7] A random coefficients model (random intercept and random slope) was determined to be most appropriate.

The time measure was again recoded from the date of the visit to "measurement time 1," "measurement time 2," and so on. Because the Parent Engagement Code was recorded at each visit, there was a far greater range of visit numbers compared to the LSP variables. The number of visits ranged from 2 to 195. The measurement time variable showed evidence of extreme non-normality, which can be problematic for logistic mixed-effects procedures; the evaluators thus performed a square root transformation of the variable. The other covariates used in this analysis were ethnicity and primary home language.

Table 21 presents the Parent Engagement Code mixed model analysis results. The coefficients that are produced by logistic mixed-effects models are reported in log-odds. In order to interpret the coefficients, they can be converted to odds ratios by taking the antilog of the coefficient (i.e., e^x, where x is the coefficient estimate; Cohen et al., 2003). It represents the amount by which the odds in favor of the outcome are multiplied for a one-unit increase in the predictor. The study showed statistically significant results for ethnicity variables for two parent groups. Asian and Hispanic/Latino parents were 99 percent less likely to receive a rating of five when compared to the reference group (i.e., White parents) (Odds Ratios = .001 and .003, respectively).

---

[7] The evaluators also ran a cumulative logit mixed-effects model that uses an ordinal outcome variable (the Parent Engagement Code was kept on its original 1-5 scale). The results of the cumulative logit mixed-effects model were similar to the logistic model and did not reveal any additional information. Because the results were similar, the results of the logistic mixed-effects model are reported because they are easier to interpret.

Table 21

Parameter Estimates for the Logistic Mixed-Effects Parent Engagement Code Model

Parameter	Parent Engagement Code
Fixed effects	
Intercept	6.16*
	(2.45)
Measurement time (square root transformation)	.15
	(.10)
Ethnicity	
African	-2.06
	(2.42)
African American	-.74
	(2.22)
American Indian/Native Alaskan	-2.25
	(2.60)
Asian	-6.83**
	(2.64)
Hispanic or Latino	-.5.63*
	(2.46)
Primary language at home (1 = English is	-.1.77
primary language)	(1.65)
Random effects variance	
Intercept	16.96
Measurement time	.19

Note. standard errors are in parentheses. Maximum likelihood with Laplace Approximation was used.
Reported coefficients are in log odds. To convert coefficients to probabilities, calculate: $\exp(x)/(1 + \exp(x))$
*** = $p < .001$; ** = $p < .01$; * = $p < .05$

Parent Involvement with Children's School. Way to Grow staff tracks program parents' involvement with their children's schooling. The program's data base contained dates and times that parents of elementary school children in grades K-3 attended parent-teacher conferences, school events, or volunteered at their child's school. Frequencies for each of these three categories were totaled for analysis of parent engagement with schools. Table 22 displays descriptive statistics on the parental school involvement for K-3 families (*n*=64) who were part of the 100 sampled families/children during the May 1, 2016 through December 15, 2017. Forty-five percent of the K-3 parents attended an average of two parent-teacher conferences, 16 percent attended less than two school events, and less than one percent volunteered at school.

Table 22

Descriptive Statistics on Parent School Involvement

	N of parents	*N* of events	Minimum	Maximum	Mean	*SD*
Attending parent-teacher conferences	29	58	1	6	2	1.2
Volunteering at child's school	1*	1				
Attending school events	10	15	1	2	< 2	0.5

* Prior to May 1, 2016, seven parents are recorded as having volunteered at child's school once each.

Child Development and Growth in Preparation for and Success in Formal Schooling

The impact confirmatory research question was

2. Does child development and growth progress achieve appropriate milestones and in conjunction with preparation and success for formal schooling?

IGDI and DIBELS. The Individual Growth and Development Indicators (IGDI) and Dynamic Indicators of Basic Literacy Skills (DIBELS) are administered to Great by Eight four-year olds to monitor children's literacy growth prior to kindergarten. These assessments were administered up to three times during the year before kindergarten. Results on the IGDI picture naming, rhyming and alliteration scales and the DIBELS initial sounds and letter naming indicators were reported in the program's data base. Table 23 provides descriptive statistics on these measures.

Table 23

Descriptive Statistics on DIBELS and IGDI Results by Time

	IGDI			DIBELS		
	M	*SD*	*N*	*M*	*SD*	*N*
Fall	26.61	13.11	75	15.09	16.69	75
Winter	34.45	15.22	82	26.86	25.15	90
Spring	46.13	16.42	78	38.16	25.85	88

Table 23 shows the average IGDI and DIBELS scores for four year olds at each time point. Mean scores increased on the IGDI from 26.61 in the fall to 46.13 in the spring and on the DIBELS from 15.09 to 38.16 the during the same time period.

The Minneapolis school district considers preschool children on a positive trajectory for kindergarten when they score at least 10 when combining the IGDI rhyming and alliteration subscales and seven when adding the DIBELS initial sounds and letter naming assessments. Table 24 shows the percent of Great by Eight children meeting those expectations at each time point.

Table 24

Percent of Children Meeting Kindergarten Literacy Expectations

	Fall %	Winter %	Spring %
IGDI	34.7	54.9	80.8
DIBELS	50.7	74.4	88.6
Both	25.3	44.4	69.3

Table 24 indicates that approximately 25% of Great by Eight children were on track for kindergarten at their initial literacy screening in the fall. The spring assessment identified approximately 69% of children meeting the district's expectations.

The evaluators conducted a paired Wilcoxan Signed Rank *t*-test[8] to statistically analyze children's progress from one testing period to another. The analysis was completed to determine any significant differences in the children's IGDI and DIBELS literacy scores over time. The combined rhyming and alliteration IGDI scores were used in the statistical procedure. Table 25 presents the IGDI *t*-test results.

Table 25

Paired Wilcoxan Signed Rank t-tests for Change in IDGI Scores Over Time

	z	*p*-value
Fall to Winter	-4.73	<0.001
Fall to Spring	-6.61	<0.001
Winter to Spring	-6.03	<0.001

Table 25 shows that literacy skills significantly ($p < .001$) improved as children moved from one IGDI assessment period to another. When comparing the initial screening (fall) to the last measurement (spring), children, on average, gained approximately 14 points on the combined rhyming and alliteration scales.

The children's initial sounds and letter naming scores on the DIBELS were totaled to assess change over time. Table 26 presents the paired *t*-test results on these indicators.

---

[8] The Wilcoxan Signed Rank *t*-test is a type of *t*-test that is used on data that is not normally distributed. The IGDI and DIBELS scales both had non-normal distributions.

Table 26

Paired Wilcoxan Signed Rank t-tests for Change in DIBELS Scores Over Time

	z	p-value
Fall to Winter	-6.22	<0.001
Fall to Spring	-6.96	<0.001
Winter to Spring	-6.76	<0.001

Table 26 shows significant ($p < .001$) positive score differences on the DIBELS when assessment time periods were compared. Children gained, on average, approximately 23 points on combined DIBELS literacy scales.

Parents' Access to Resources and Navigation of Health and Education

The impact confirmatory research questions were

3. Do Great by Eight health components (health referrals and education measured by dosage) improve and sustain health promoting behaviors?

4. Do Great by Eight parents and children access resources and navigate health and education systems?

A major goal of Way to Grow's Great by Eight program is to connect program parents and their children to supportive community resources. Family educators work collaboratively with community and school agencies to help program participants receive appropriate health and educational services. Families are monitored and tracked and their progress is recorded in the program's data base. The database showed for the 100 sampled families during the date range from May 1, 2016 through December 15, 2017, 50 out of the 100 program parents were referred to one or more community services. Parents were often referred to the same resource more than once.

Table 27 shows that 279 community resource referrals were made to the 50 families with at least one referral during the timeframe. The table also shows that among the 50 families, 118 community services were distributed as shown among the families. Twenty-five parents had referrals to an early learning preschool program for their children who were 3-5 years of age, and 56 percent of parents had referrals related to family medical care.

Table 27

Referrals and Connections to Community Resources (N of families =50)

	Number of Referrals	Number of families referred
Early Education	54	25
Other	49	14
Medical Provider	47	28
Parent/Household	32	10
Financial	26	11
Housing	21	6
Education	20	9
Food	15	10
Baby	9	2
Legal	5	2
Health Insurance	1	1
Total	279	

Life Skills Progression (LSP) Scale. Way to Grow staff uses four indicators on the Life Skills Progression (LSP) assessment to rate parental behavior related to family health over time. Two scales assess whether parents have obtained appropriate medical care (immunizations and dental) for their children. The third scale rates parents on accessing health/medical insurance, an outcome that family educators stress during home visits. The last indicator evaluates general parental child wellness care.

Linear mixed-effects statistical procedures were used to identify any predictor variables on the four repeated LSP measures. For all four outcome variables, a random coefficients model (random intercept, random slope) was used. The same independent variables that were included in the parent *engagement* LSP variables were used for this analysis: measurement time, number of home visits, primary home language, and ethnicity.

Change in health variables over time. Table 28 presents the mean score on each scale at each measurement time. With the exception of the Medical Health Insurance scale, the mean value at each measurement time (including time 1) was between 4 and 5, which meant that there was little room for the scores to increase over time. For all scales except for Medical Health Insurance, the means were higher at measurement time 7 (the last measurement time with more than 1 score recorded) compared to measurement time 1, but the increases were small in magnitude. For the Medical Health Insurance variable, there was a slight decrease in the average score over time.

Table 28

Means for Life Skills Progression Scales over Time

	Time 1 (n = 100)	Time 2 (n = 76)	Time 3 (n = 58)	Time 4 (n = 50)	Time 5 (n = 41)	Time 6 (n = 27)	Time 7 (n = 21)	Time 8 (n = 1)
Child Immunization	4.57 (.95)	4.71 (.74)	4.56 (.85)	4.82 (.43)	4.93 (.24)	4.91 (.28)	4.86 (.36)	5.00 (NA)
Child dental care	4.13 (1.36)	4.26 (1.38)	4.10 (1.24)	4.38 (1.26)	4.67 (.47)	4.72 (.45)	4.76 (.44)	5.00 (NA)
Medical health insurance	3.60 (1.16)	3.59 (1.20)	3.28 (1.17)	3.52 (1.10)	3.54 (1.13)	3.41 (1.07)	3.05 (1.15)	4.00 (NA)
Child wellness care	4.51 (.96)	4.75 (.96)	4.66 (.64)	4.74 (.49)	4.76 (.42)	4.85 (.46)	4.86 (.36)	5.00 (NA)

Note: Means are reported with standard deviations in parentheses. Each scale is scored across a range of 0-5 points using 0.5-point increments.

Effect of covariates. Table 29 reports the parameter estimates for each model. Because the research questions are focused on aggregate patterns, the fixed effects are the main variables of interest for interpretation. The variance of the random effects, which contains information about individual differences, is also reported.

Interpretation of fixed effects. The intercept estimate can be interpreted as the predicted value for an observation for which all values on all variables equal zero.[9] The measurement time coefficient, which represents the slope, is the estimated amount that the variable would change from one measurement time to the next. For three of the four health scales, the measurement time variable had a statistically significant coefficient. On average, between one measurement time to the next, there was a .06-point increase in Child Immunization, a .11-point increase in Child Dental Care, and a .04-point increase in Child Wellness Care.

The number of visits coefficient represents the effect of how many visits the family had received on their score at measurement time 1. The coefficient for number of visits was 0 for all health variables.

---

[9] i.e. A participant who identifies as White, non-English speaking, at the first observation.

The coefficient for each ethnic group represents the average difference on the outcome for a family of that ethnicity compared to the reference group (i.e., White parents).

The results in Table 29 show there were several ethnicity coefficients that were statistically significant on the four health scales. The average Child Wellness Care score at measurement time 1 for African American parents was .6 points higher than the average score for White parents. American Indian/Native Alaskan families scored an average of .71 points higher on the Child Immunization scale and .83 points higher on the Child Wellness Care scale at time 1. Finally, Hispanic or Latino parents on average scored .67 points higher for Child Immunization. Parents who spoke English as their primary language at home had an average of .51 points lower on their Child Wellness Care score at time 1.

Table 29

Parameter Estimates for the LSP Health Scales (Linear Random Coefficient Models)

Parameter	Child Immunization	Child Dental Care	Medical Health Insurance	Child Wellness Care
Fixed effects				
Intercept	4.07***	3.87***	3.51***	4.23***
	(.30)	(.58)	(.49)	(.29)
Measurement time	.06**	.11***	-.01	.04*
	(.02)	(.03)	(.03)	(.02)
Number of visits	.00	.00	.00	.00
	(.00)	(.00)	(.00)	(.00)
Ethnicity				
African	.50	.28	.39	.36
	(.30)	(.60)	(.51)	(.30)
African American	.53	.23	.19	.60*
	(.29)	(.56)	(.44)	(.27)
American	.71*	.70	.29	.83**
Indian/Native Alaskan	(.32)	(.63)	(.52)	(.31)
Asian	.38	.00	.46	.30
	(.31)	(.60)	(.55)	(.30)
Hispanic or Latino	.67*	.14	-.81	.44
	(.30)	(.59)	(.50)	(.29)
Primary language at home (1 =	-.13	-.18	.39	-.51**
English is primary language)	(.16)	(.57)	(.35)	(.17)
Random effects variance				
Intercept	.54	1.24	.80	.55
Measurement time	.11	.20	.14	.11
Residual	.58	.62	.56	.48

Note. standard errors are in parentheses. Full maximum likelihood was used.
*** = $p < .001$; ** = $p < .01$; * = $p < .05$

Change in between-subject variability. Table 30 reports each model's reduction in between-subject variability as a result of adding covariates. Medical Health Insurance had the greatest amount of reduction in between-subject variability (10.17%), whereas Child Dental Care had the least amount of change (1.64%).

Table 30

Reduction in Between-Subject Variability for each LSP Health Scale

	Between-subject variability: Null model with no covariates	Between-subject variability: Final model with covariates	Percent reduction in between-subject variability
Child Immunization	.46	.44	4.35%
Child dental care	.61	.60	1.64%
Medical health insurance	.59	.53	10.17%
Child wellness care	.51	.48	5.89%

Limitations

Assumptions violation. When using a linear mixed-effects approach, it is important to check the following assumptions of the final models:

1. Each random effect has a normal distribution
2. The errors are normally distributed with a mean of zero
3. The errors are independent of each other

All of the models had some random effects and/or errors with a skew that was greater than +/-1, therefore demonstrating evidence of non-normality and violating assumptions 1 and 2. Re-running the models with extreme outliers removed only corrected this problem without significantly affecting parameter estimates for the Child Dental Care scale. The results for Child Immunization, Medical Health Insurance, and Child Wellness Care should thus be interpreted with caution. It is likely that these models were mis-specified due to important covariates missing from the data. In the future, the program could consider including additional variables that may help to explain differences in health outcomes between families.

Outcome variable. As noted above, most of the health LSP ratings that parents received were consistently high, even during the earliest observations. This meant that there was only a small amount of variation in the outcome variables that could be explained in the first place. In the future, the program may want to consider using a different instrument that can capture a greater amount of change over time.

Study Two Summary of Findings and Limitations

Parent engagement with their children's development, learning, and schooling

Study two was designed to explore and confirm changes based on home visitation interventions related to increased parent engagement.

LSP scale. We found very little change over time in the parent engagement outcome variables, the seven LSP scales speaking to parent engagement. The LSP ratings that parents received were consistently high, even during the earliest observations. Because parents were rated so highly from the first time they were assessed, there was little room for the average score to increase over time. This meant that the LSP instrument *could not* capture substantial change over time. The lack of statistically significant covariate coefficients – or, in the case of the *number of visits* variable, coefficients that did not differ significantly from zero – may be due to the very small amount of variation in the outcome variables that could be explained in the first place.

Parent Engagement Code. The Parent Engagement Code is a zero to 5 rating made by the family educator at each home visit on the *quality* of parent engagement. Our analysis of the Parent Engagement Code revealed significant differences between parents of different ethnicities. Specifically, we found that Asian and Hispanic/Latino parents were 99 percent less likely to receive a rating of five when compared to the reference group (i.e., White parents).

Parent Involvement with Children's School. We found that most parents were involved at their child's school through parent teacher conferences. Forty-five percent of the K-3 parents (*n*=64) in the 100 sampled families during the May 1 through December 15, 2017 timeframe, attended an average of two parent-teacher conferences, 16 percent of these parents attended less than two school events, and less than one percent volunteered at school.

Child development and growth in preparation for and success in formal schooling

IGDI and DIBELS. We found that the Great by Eight four year olds were on a positive trajectory for meeting expectations at each time point for kindergarten readiness. Literacy skills improved significantly as children moved from one IGDI and DIBELS assessment period to another.

Parents' access to resources and navigation of health and education

Way to Grow's programming data showed that for the 100 sampled families during the date range from May 1, 2016 through December 15, 2017, 50 out of the 100 program parents were referred to one or more community services, and parents were often referred to the same resource more than once. For the 50 parents, 118 community service referrals were made. Twenty-five parents had referrals to an early learning preschool program for their children who were 3-5 years of age, and 56 percent of parents had referrals related to family medical care.

Life Skills Progression (LSP) Scale. Two scales assess whether parents have obtained appropriate medical care (immunizations and dental care) for their children. The third scale rates parents on accessing health/medical insurance, an outcome that family educators stress during home visits.

The last indicator evaluates general parental child wellness care. The same independent variables that were included in the parent engagement LSP variables were used for this analysis: measurement time, number of home visits, primary home language, and ethnicity.

We found that, with the exception of the Medical Health Insurance scale, the mean value at each measurement time (including time 1) was between 4 and 5, which meant that there was again little room for the scores to increase over time. For all scales except for Medical Health Insurance, the means were higher at measurement time 7 (the last measurement time with more than 1 score recorded) compared to measurement time 1, but the increases were small in magnitude. For the Medical Health Insurance variable, there was a slight decrease in the average score over time.

On average, between one measurement time to the next, there was a small but significant point increase in Child Immunization, Child Dental Care, and Child Wellness Care. There were several ethnicity coefficients that were statistically significant.

A Brief Comparison of Results with Interim Evaluations

Two interim Great by Eight SIF evaluation reports are available. The first was *Great by eight home visiting program evaluation: interim report* (Daugherty & Edwards, 2015*).* Daugherty & Edwards was submitted to and reviewed by Greater Twin Cities United Way and reviewed by CNCS. The second report was *Year three great by eight evaluation: interim report* (Dretzke, Schultz, & Desjardins, 2016). Dretzke et al. was submitted to and reviewed by CNCS (review date was November 21, 2016).[10]

Daugherty & Edwards followed the original SEP and reported on an implementation study and two impact studies. They had Great by Eight parent and children's data logged from September 2013 through April 2015. Dretzke et al. reported on an implementation study and an impact study reporting on data logged from May 1, 2015 through April 30, 2016.

For Impact Study One, Daugherty & Edwards reported "non-significant differences on literacy measures when Great by Eight children were compared to similar children in the Minneapolis public school system." This was the case for both ready for kindergarten and third grade reading proficiency. In our final evaluation, we did find that Great by Eight third graders scored significantly higher on reading proficiency than comparison children. Dretzke et al. did not undertake Study One.

For Study Two, Daugherty & Edwards undertook an extensive impact analysis but deviated from the SEP in an important way: they did not randomly sample the 100 families/children for the

---

[10] Over the SIF-funded five years, there have been three different lead external evaluators and several different graduate research assistants involved in the evaluation. All have been CAREI personnel. Although we have not tried to gauge the costs of reduced continuity in the Great by Eight evaluation(s), we do not doubt that there have been costs. Throughout the evaluation, however, Way to Grow's Chief Executive Officer and the lead Education and Evaluation staff member have been on board and worked closely with the changing CAREI evaluation staff.

analysis but rather included all children's and parents' data.[11] Also they did not discuss the fact that they did not/could not conduct an interrupted time series analysis as we could not in our final report. Dretzke et al. also undertook many aspects of what would be a Study Two analysis but, following Daugherty & Edwards lead, did not randomly sample the 100 families/children for the analysis, per the SEP. In any case, here are the highlights from their conclusions.

Importantly, Daugherty & Edwards found that the longitudinal study results showed that the *frequency* of elementary home visits was significantly related to higher ratings on five out of seven LSP scales: nurturing; discipline; support of development; use of resources; and parent cognitive ability (cognitive understanding of children's needs).

Number of elementary home visits was also a significant positive predictor of parents attending parent-teacher conferences and school events as well as volunteering at their children's schools. Recall that in our final analysis of parent involvement with their children's school we found that most parents were involved at their child's school through parent teacher conferences.

Importantly and interestingly, Daugherty & Edwards found that *length of time* in the program was not related to any parent engagement measures.

As expected, Daugherty & Edwards study's results showed some ethnic differences among parent-child interactions. In our final report, we too found differential effects on LSP scales, although the differences found do not map one-to-one with Daugherty & Edwards. Dretzke et al. also found that on all of the scales except for the cognitive abilities scale, there were differences associated with differing ethnicities.

With regard to child development and growth in preparation for success in formal schooling, in both Daugherty & Edwards and this final report, evaluators found statistically significant increases in IGDI and DIBELS on fall to winter and winter to spring assessments. Dretzke et al. also found the same significant gains on DIBELS from assessment to assessment. For some reason, they did not examine IGDIs. Daugherty & Edwards found that children whose primary home language was *not* English scored significantly higher on the DIBELS when compared to their English-speaking counterparts. That trend was not apparent in the IGDI analysis. They also found that elementary home visits significantly predicted higher scores on the IGDI literacy scales whereas program *duration* was strongly associated with higher DIBELS scores. Finally, their results showed some variability related to child characteristics, including ethnicity and special needs status.

Turning to parents' access to resources and navigation of health and education, all three analyses found somewhat similar access to health and education resources, Daugherty & Edwards, for example, reported that approximately 60% of their parents were connected to family medical and health resources and early education programs for their children. Dretzke et al. reported 59%. In the final report the percentage was closer to 50%. Daugherty & Edwards also found that early learning home visits were significantly associated with parents obtaining appropriate immunizations for their children, and the number of elementary home visits was a significant

---

[11] In the SEP, Daugherty and Way to Grow staff wrote: "Approximately 500 children ages 3 to 8 and 475 parents are enrolled in the Great by Eight program. A random sample of 100 children (aged three to eight years) participating in the Great by Eight program and their parent(s) will be selected to participate in the interrupted time series [Study Two] evaluation plan." (26)

predictor of parents acquiring medical insurance and dental care for their children. Their analysis did show some ethnicity differences across the four LSP scales.

Interpretations and Limitations

Impact

From the three program evaluations over the nearly five years of the SIF award, we find that Great by Eight home visiting has a positive impact on assisting children at risk for low academic achievement to be better-prepared or prepared for kindergarten and formal schooling. Determining the size of the impact has been limited by uneven BKA assessments across the years from fall 2013 through fall 2017 and limitation the Minnesota MCA reading assessments. IGDI and DIBELS assessments are used to inform practice the of family educators to bolster positive impact. Also – although strange to call it a limitation – for parents of children in kindergarten through third grade, we cannot disentangle Great by Eight programming from children's classroom learning and other schools supports, such as tutoring.

It is clear that the home visiting dosage, including frequency and type of home visit, has a positive effect on higher ratings on parents' interactions/engagement with their children in support of their physical and educational development over time. Limitations to clearly documenting this outcome have to do with limitations of the measurement instruments and likely to family educator positive bias and/or rating inflation. The reliability and validity estimates of the Parent Engagement Code are not known. And, while we have reliability and validity estimates of the LSP scales, they still might be an inadequate choice of outcome measure for the efficacy of the home visiting program. Or, while we have reliability and validity estimates of the IGDI and DIBELS assessment tools, we do not have quality control estimates of outside assessor's reliability in assessments.

It is also clear that the home visiting dosage has a positive effect on parents' accessing health, medical, and educational resources, all of which has a positive impact on children's physical, emotional, and cognitive development.

Implementation

From the three implementation evaluations, it is clear that Great by Eight has provided an intensive home visiting program and has, over the SIF-funded period, impressively scaled up its programming. Parents are very satisfied with the Great by Eight program in terms of what they are learning and how things are improving in their lives. Parents found all topics concerning their children's school-related and health-related outcomes presented by family educators during home visits *very useful* and they reported learning *a lot*. Most parents rated all program aspects *very important,* indicating a high overall level of perceived importance of the different aspects of the program.

Nine out of 10 parents responded *always* for all seven aspects of the parent-family educator relationship, indicating a very high level of satisfaction in this area. And 9 out of 10 parents rated the overall *quality* of their relationship with the family educator as *excellent.*

A main limitation of implementation outcomes based on the parent survey data is that we have no reliability data. Although the survey content was derived from responses given by focus groups

conducted with parents, program administrators, and family educators in the implementation evaluation of Daugherty & Edwards, reliability studies have not been conducted.

Across the three implementation evaluations, *the* limitation to the evaluations and thus to the impact evaluations is the almost complete lack of fidelity of implementation information at the point of dosage administration, namely, during home visits. While quality control during family educator training and supervisor support at Way to Grow might be exemplary, we do not know in important ways how family educators are conducting home visits. For example, because of the lack of HOVRS data available for the sample of families who completed a survey, we were not able to study fidelity of implementation of home visits.

Conclusion and Recommendations

The purpose of this evaluation was to examine the implementation and impact of the Great by Eight program spanning activities from May 2016 through the end of 2017. Using a quasi-experimental matched-groups design, the impact evaluation was intended to fulfill the SIF requirement to achieve at least a moderate level of evidence for Way to Grow's funded project. As discussed in *Interpretations and Limitations* above, the results indicate that Great by Eight does improve at-risk children being ready for kindergarten and third grade reading proficiency to some degree.

Recommendations

First, Way to Grow should look closely at undertaking future implementation evaluation that studies quality control in the Great by Eight program. From the standpoint of this evaluation, we believe that it is essential they increase systematic review efforts documenting family educators' early learning and elementary program home visits. The Home Visit Rating Scales (HOVRS) may or may not be the best instrument for supervisors to assess family educator fidelity of implementation efforts during home visits. In addition to parent surveys and staff discussions, frequent in-home observation of educator-family interactions are necessary.

Second, we recommend that Way to Grow assess how heavily they should rely on public school data for their in-house evaluation of Great by Eight program impacts. The Great by Eight program should explore assessments appropriate for the children and families it serves to more reliably assess home visiting, early learning and elementary programing.

In the same way, they should be cautious in undertaking comparisons with groups of non-participant children on any metric. A main issue in doing so is not being able to document the developmental histories of the comparison group children.

Third, Way to Grow should consider tracking participant children and families over a longer timeframe to the degree that this is feasible. If feasible, this could help demonstrate more distal program impacts.

Fourth, we recommend that Way to Grow assess the appropriateness of their current outcome measurement instruments. For example, the LSP scales may or may not be reliable indicators for

assessing Way to Grow parents, and the Parent Engagement Code may not provide useful formative information for family educator support efforts for parents.

Finally, while the focus of this impact evaluation was on educational/academic outcomes, Way to Grow may benefit greatly from future evaluations around non-academic outcomes, such as social, emotional and [other] outcomes.

References

Albritton, S., Klotz, J., & Roberson, T. (2004). The effects of participating in a Parents as Teachers program on parental involvement in the learning process at school and home. *E-Journal of Teaching and Learning in Diverse Settings, 1*, 188-208.

Azzi-Lessing, L. (2011). Home visitation programs: Critical Issues and future directions. Early Childhood Research Quarterly, 26, 387-398.

Biglan, A., Ary, D., & Wagenaar, A. C. (2000). The value of interrupted time-series experiments for community intervention research. *Prevention Science*, *1*(1), 31–49.

Blozis, S. A., & Cudeck, R. (1999). Conditionally linear mixed-effects models with latent variable covariates. *Journal of Educational and Behavioral Statistics, 24*, 245-270.

Bronfenbrenner, U. & Morris, P. (1998).The ecology of developmental processes. In W. Damon (Series Ed.) & R. M. Lerner (Vol. Ed.), *Handbook of child psychology: Vol. 1: Theoretical models of human development* (pp. 993-1028).New York: Wiley.

Caldera, D., Burrell, L., Rodriguez, K., Crowne, S. S., Rohde, C., & Duggan, A. (2007). Impact of a statewide home visiting program on parenting and on child health and development. *Child Abuse & Neglect, 31*(8), 829–852.

Children's Defense Fund (2013). *Home visiting programs*. Retrieved from: http://www.childrensdefense.org/policy-priorities/early-childhood-education-care/home-visiting-programs.html

Cohen, J., Cohen, P., West, S., & Aiken, L. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences.* New York, NY: Taylor & Francis.

Daughtery, M. and Edwards, K. *Great by Eight Home Visiting Program Evaluation: Interim Report.* Available on request to Naomi Zuk-Fisher, Greater Twin Cities United Way or Delia Kundin, Center for Applied Research and Evaluation, College of Education and Human Development, University of Minnesota.

Design Options for Home Visiting Evaluation (DOHVE). (October 2011*). Life skills progression brief: Information and guidelines for use in meeting MIECHV benchmarks.*

Dretzke, B. J., Schultz, A. R., & Desjardins, C. (2016, August). *Year Three Great by Eight Evaluation: Interim Report*. Saint Paul, MN: University of Minnesota, College of Education and Human Development, Center for Applied Research and Educational Improvement.

Fan, X. & Nowell, D. L. (2011). Using propensity score matching in educational research. *Gifted Child Quarterly,* 55, 74-79.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). Applied longitudinal analysis (2nd ed.). Hoboken, N. J. Wiley.

Good R. & Kaminski, R. (2006). *Dynamic Indicators of Basic Early Literacy Skills, University of Oregon.* Retrieved August 28, 2013 from http://dibels.uoregon.edu/.

Kirkland, K. & Mitchell-Herzfeld, S. (2012). Evaluating the effectiveness of home visiting services in promoting children's adjustment to school. New York State Office of Child and Family Services, Healthy Families Final Evaluation Report (Grant # 2010-0588) Retrieved from:
http://www.pewstates.org/uploadedFiles/PCS_Assets/2013/School_Readiness_report.pdf

Maxwell, S.D. & Delaney, H. D., (2004). *Designing experiments and analyzing data.* Lawrence Erlbaum Associates.

McConnell, S. R. (1990). Best practices in evaluating educational programs. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (2nd. ed.) (pp. 353-370). Washington, D. C.: National Association of School Psychologists.

Michlin, M. L. & Schultz, A. R. (2016). *Summer Youth Employment Programs Effect Size Review.* Saint Paul, MN: University of Minnesota, College of Education and Human Development, Center for Applied Research and Educational Improvement.

Mitchell-Herzfeld, S., DuMont, K., Lee, E., & Spera, C. (2005). *New York Healthy Families Program randomized trial: Findings from the 2 year follow-up*. Presented at the Prevent Child Abuse New York Conference, Albany, NY.

Pfannenstiel, J. C., Seitz, V., & Zigler, E. (2002). Promoting school readiness: The role of the Parents as Teachers program. *NHSA Dialog: A Research-to-Practice Journal for the Early Intervention Field, 6,* 71-86.

Priest, J., Davis, K., McConnell, S., McEvoy, M., & Shin, J. (1999, December). Individual growth and development indicators of preschoolers' "expressing meaning" skills: Follow that trajectory! Paper presented at the annual meeting of the Division for Early Childhood, Council for Exceptional Children, Washington, D.C.

Rodriguez, M., DuMont, K., Mitchell-Herzfeld, S., Walden, N., Greene, R. (2010). Effects of Healthy Families New York on the promotion of maternal parenting competencies and the prevention of harsh parenting. *Child Abuse & Neglect, 34*, 711-723.

Sekhon, R. J. (2011). Multivariate and propensity score matching software with automated balance optimization: The matching package R. *Journal of Statistical Software*, *42*, 1-52.

Singer, J. D. & Willett, J. (2003). *Applied longitudinal data analysis: Modeling change and even occurrence.* New York: Oxford University Press.

Wollesen, L. & Peifer, K. (2006). An outcome and intervention planning instrument for use with families at-risk. Paul H. Brooks Publishing Co.

Zigler, E., Pfannenstiel, J.C., & Seitz, V. (2008). The Parents as Teachers program and school success: A replication and extension. *Journal of Primary Prevention, 29,* 103-120.

Appendix A
Great by Eight Logic Model

Resources	Outputs (Activities)	Short Term Outcomes	Impact
▪ Participating families ▪ Stable and engaged funders & board ▪ Qualified, multilingual & multicultural staff ▪ Research-based curriculum, tools, training, assessments ▪ Relationships with community & partner agencies ▪ Facilities, equipment, technology & data systems	**liver home visits to at-risk families with children ages three to eight, in which family educators:** ▪ Provide parents with information about children's learning and development ▪ Use research based early childhood curricula and materials to teach children literacy and math skills ▪ Demonstrate and model for parents age appropriate teaching strategies ▪ Support activities and attitudes that promote learning and teaching in the home and school ▪ Scaffold and reinforce positive parenting practices ▪ Assess children's growth and development and teach parents specific intervention strategies to respond to child data results ▪ Teach parents how to use progress monitoring to support continuous growth and change interventions ▪ Teach parents how to advocate and be engaged in their children's education ▪ Educate parents and children on topics specific to health and safety (i.e. immunization, well child checks and screenings) ▪ Assist parents in the development of preventative and proactive health and safety habits (i.e. develop a safety plan)	▪ Parents engage in high quality interactions with their child ▪ Improved home environment that supports children's learning ▪ Increased and sustained parental engagement ▪ Children are prepared for school ▪ Children meet or exceed standards for proficiency on the MCA ▪ Parents and children access health and education services	▪ Parents have the skills to support their child's learning and development ▪ Children are prepared for school ▪ Children meet or exceed standards for proficiency on the MCA ▪ Children engage in school and develop skills that support learning ▪ Parents and children have improved health and demonstrate health promoting behavior

Resources	Outputs (Activities)	Short Term Outcomes	Impact
▪ Systemic & systematic policies	**ovide referral services to families that include:** 　▪ Transportation support 　▪ Medical and dental services 　▪ Child care assistance 　▪ Parent education classes 　▪ High quality early learning programs 　▪ Financial literacy classes 　▪ Health and nutrition resources and classes 　▪ Basic need support with housing, clothing, food, and employment **intain community partnerships and community-based visits that:** 　▪ Establish a Great by Eight presence in Minneapolis public elementary schools to work with teachers, parents, and children. 　▪ Facilitate parent engagement in early intervention and special education services		▪ Parents access resources and navigate health and education systems

Outcomes and Impacts should be SMART: • **S**pecific, • **M**easurable,- • **A**ction-oriented • **R**ealistic • **T**ime

Appendix B
Measurement Instruments

Impact Evaluation Study One Instruments

Beginning Kindergarten Assessment (BKA) The Minneapolis Beginning Kindergarten Assessment (BKA) is a standardized assessment of reading and numerical skills and, in the past, was administered districtwide in the fall of each year. During the 15-minute one-on-one assessment, children demonstrate their knowledge and skills in vocabulary, listening comprehension, phonological awareness, alphabetic understanding, print concepts, counting and number concepts. A total literacy score is calculated in a weighted equation that includes measures of concepts in print, letter names and sounds, vocabulary, and rhyming in the formula. MPS's REAA provided the evaluation data, providing only the total literacy score but not the scores from the individual scales.

The BKA was developed in Minneapolis schools in conjunction with the University of Minnesota, and in the past, it was a district-wide test. The sample utilized for test validation and norming computation were students enrolled in MPS's kindergarten classes during the 2003-2004 school year. The total sample size was 3,174 students. According to the Minneapolis Public School system, the purpose of the BKA is to measure students' status on literacy and numerical skills at the beginning of kindergarten. They also claim that the instrument has "acceptable reliability statistics including test/retest and internal consistency coefficients and acceptable content and construct validity statistics" (http://rea.mpls.k12.mn.us/assessment_overview, 2013).

Internal consistency reliability of the assessment was calculated using the Gilmer-Feldt coefficient. Test-retest reliability was conducted on 88 randomly selected students from the original sample. The retest was done four weeks after initial testing at the beginning of kindergarten. The results indicate that the reliability coefficient between Early Literacy Composite scores on the two testing sessions was high ($r = 0.92$) and domain coefficients ranged from 0.46 to 0.89. This information suggests a high level of stability of scores over time.

Content validity refers to the extent to which the test questions represent the skills in the specified subject areas. This type of validity shows how well the test samples the subject matter and the degree to which the test content is tied to the instructional domain that it is meant to measure. The BKA was constructed from extensive teacher involvement and reflects their collective judgment about important skills students should know when entering kindergarten. It is also based on a substantial amount of empirical research of early literacy skills necessary for student achievement (Kindergarten Assessments Technical Manual, 2008).

Predictive validity of later academic achievement was also established on the test. To evaluate the BKA's predictive validity, the sample's BKA scores were related to their End of Kindergarten Assessment (EKA) scores. EKA and the BKA are exactly the same except for the addition of two test-lets in the quantitative domain. The students were then followed during their first and second grades. Their EKA scores were compared to scores on oral reading and reading comprehension in the first grade and to second grade reading and math scores on a nationally normed test, the Northwest Achievement Levels Test (NALT). The correlation between the BKA

and EKA Early Literacy composite scores was 0.74 ($p < 0.001$). This finding indicated that the scores on the two measures were highly related and BKA scores are useful in predicting end of year EKA scores. The correlations between the BKA and the EKA subtests were significant ($p < 0.01$ for all). Additionally, correlations between the major components of the verbal literacy domain were significant ($p < 0.01$ for all). The sample's EKA scores were also significantly related ($p < 0.01$) to their first-grade oral reading and comprehension scores and their second grade NALT reading and math scores ($p < 0.01$).

The validity studies concluded that the BKA provides a good estimate of the skill level of students coming into their kindergarten year of school. This measure is meant to provide a baseline for students. It also provides vital information on a student's strengths and weaknesses, which in turn helps provide information to teachers about instructional effectiveness.

That said, MPS did not administer the BKA in the fall of 2015, and in 2016 its use was optional in the district kindergarten classes. The district has since moved to using FAST Bridges assessments.

Minnesota Comprehensive Assessment (MCA) Reading Assessment The MCA Reading Assessment is administered to Minneapolis Public School third graders in the spring of the third-grade year. Assessment scores for both program and comparison group third graders were received from MPS's REAA.

The purpose of MCA Reading Assessment is to evaluate Minnesota third grade students' achievement measured against Minnesota academic standards. Assessment results are used to inform curriculum decisions at the district and school level, inform instruction at the classroom level, and demonstrate student academic progress (http://education.state.mn.us/mdeprod/idcplg). Each student receives a score that falls in one of four achievement levels: Does Not Meet the Standards, Partially Meets the Standards, Meets the Standards and Exceeds the Standards. Strands of the Reading Assessment include: Vocabulary Expansion (use of a variety of strategies to expand reading, listening, and speaking vocabularies); Comprehension (demonstration of literal, interpretive, and evaluative reading comprehension); and Literature (engagement with, understanding of, and appreciation of a wide variety of fiction, poetic, and nonfiction texts).

Content validity was established on the standardized assessment in a 2006 academic standards alignment analysis study (http://education.state.mn.us/mdeprod/idcplg). Reading content experts, Minnesota public school teachers, researchers, and school administrators found the MCA Reading Assessment to align strongly with "range of knowledge, cognitive consistency, and balance of representation" relative to academic standards.

The Minnesota assessments, such as the Minnesota Comprehensive Assessments-Series III (MCA-III), the Minnesota Comprehensive Assessments-Series II (MCA-II) and the Minnesota Comprehensive Assessments-Modified (MCA-Modified), are standards-based assessments. The tests are constructed to adhere rigorously to content standards defined by the Minnesota Department of Education and Minnesota educators. For each subject and grade level, the content standards specify the subject matter the students should know and the skills they should be able

to perform. In addition, performance standards are defined to specify how much of the content standards students need to demonstrate mastery of in order to achieve proficiency.

Impact Evaluation Study Two Instruments

A range of analysis strategies, including linear and logistic mixed-effects models investigated changes based on home visiting interventions related to: 1) Increased parent engagement with their children's development, learning, and schooling; 2) Child development and growth progress achieving appropriate milestones and in preparation for and success in formal schooling; and, 3) Children and parents accessing high quality health and educational services. The following measurement tools were used to gather data:

1) Parent Engagement. The construct of parent engagement was measured with multiple instruments:

Life Skills Progression Scale (LSP). The LSP is an observational tool completed by family educators that summarizes a parent's skills in parent-child relationships, education and employment, health and medical care, mental health and substance abuse, and basic essentials (Wollesen & Peifer, 2006). The LSP describes individual parent and children's progress using 43 types of life skills, which are grouped into five categories. Way to Grow family educators assess parents with the LSP at program entry. The instrument is then used every few months to measure a parent's progress.

Four LSP scales in the Relationship with Children category were used to assess parent engagement with their children. These scales were: 1) Nurturing (skills and ability), 2) Discipline (appropriateness), 3) Support of Child Development, and 4) Safety (protecting children from environmental harm). The Way to Grow staff indicated that their home visiting program emphasized three additional parental engagement characteristics that are measured by other LSP scales. These scales were: 5) Use of Resources, 6) Child Care (quality of child care environment), and 7) Cognitive Ability (cognitive understanding of children's needs). In total, seven LSP scales were included in the impact analyses.

Each scale is scored independently across a range of 0 to 5 points using 0.50-point increments. In a publication of LSP use for program evaluation (Design Options for Home Visiting Evaluation, 2011), reliability and validity of the instrument were reported. With training, inter-rater reliability for the LSP ranged from 78% to 90% (acceptable to very good). Test-retest reliability from one pilot study resulted in an average inter-item correlation score of 0.90 (very good). The internal consistency reliability coefficient for the LSP from a second pilot was 0.74 (acceptable).

Construct validity (alpha scores) was calculated based on results from two pilot programs and ranged from 0.64 to 0.99 (acceptable to excellent). These studies were carried out in community-based home visitation programs in one county in California. Content validity was determined in 2003 by an expert advisory review panel of five ZERO TO THREE Fellows (Design Options for Home Visiting Evaluation, 2011).

Parent Engagement Code. The Parent Engagement Code is a five-point scale developed by Way to Grow. It is used on the Great by Eight Contact Tracking Form and completed by the family educators. They record dosage information on the form and then rate parent engagement for each home visit and other points of contact. The codes are defined using language that describes behaviors along a continuum of parent-child involvement, interactions and enjoyment, and discourse with one being the lowest level of engagement and five being the highest level of engagement.

Parent Involvement with Children's School. The Way to Grow staff tracks program parents' involvement with their children's schooling. Specifically, dates and times that parents attended parent-teacher conferences and school events, or volunteered at the school, were entered into the program's longitudinal data base. The staff uses the data to track parents' engagement with their children's educational experience. The data provided indicators of parental progress with school involvement.

2) Children's Development and Academic Progress. The Way to Grow staff monitors program children's development and academic progress. Measurement results guide home visitation interventions to assist with children's school readiness and academic progress. Two measures were used:

Individual Growth and Development Indicators (IGDIs) IGDIs are used to monitor the language and literacy development of preschoolers, three to five years of age. The IGDIs test is comprised of three key measures: Picture Naming (an assessment of expressive language development), Rhyming, and Alliteration (assessments of phonological awareness). Preschool IGDIs are used to monitor development over time. When performance on one or repeated assessments indicates that children are not making desired rates of progress toward a long-term goal, intervention can be designed and implemented.

The IGDI's are standardized and take approximately 10-12 minutes to complete on a child. The observer sets up a structured observation and rates the child's performance. For example, the rhyming activity asks children to identify pictures of items that rhyme and takes approximately two minutes to complete. Learning to score IGDI's is straightforward for most observers, based on strong results from tests of inter-observer reliability.

Validity studies on the IGDI have shown strong concurrent relationships between the Picture Naming measure and norm-referenced measures of preschoolers' language skills, including the Peabody Picture Vocabulary Test (PPVT) and the Preschool Language Scale with correlation coefficients ranging from $r = .47$ to $.69$ (Priest, Davis, McConnell, McEvoy, & Shin, 1999). One-month alternate form reliability coefficients for the observational measure range from $r = .25$ to $.79$. Additionally, studies have yielded strong correlations to measures of language (with PPVT-3, $r = .57$) and early literacy, including Clay's Concepts About Print ($r = .55$), letter identification ($r = .74$), and the Test of Phonological Awareness ($r = .75$) (McConnell, 1990).

Dynamic Indicators of Basic Early Literacy Skills The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) are a set of procedures and measures for assessing the acquisition of early literacy skills from kindergarten through sixth grade. They are designed to be short (one

minute) fluency measures used to regularly monitor the development of early literacy and early reading skills.

DIBELS are comprised of seven measures to function as indicators of phonemic awareness, alphabetic principles, accuracy and fluency with connected text, reading comprehension, and vocabulary. DIBELS were designed for use in identifying children experiencing difficulty in acquisition of basic early literacy skills in order to provide support early and prevent the occurrence of later reading difficulties. Alternative-form reliability has been determined at .93. Concurrent validity with the Woodcock-Johnson Psycho-Educational Battery (WJPEB) readiness score was .70 in kindergarten (Good & Kaminski, 2006).

Implementation Evaluation Instruments

Parent Survey

Per the revised and approved SEP for the implementation evaluation (2016), the evaluators and Way to Grow staff developed a survey to be administered to parents to assess their perceptions regarding three main areas

- How much they learned about school-related and health-related topics during home visits and the usefulness of the information
- The importance of the delivery features of the Great by Eight program[12]
- The quality of the parent-family educator relationship

The evaluators proposed administering the survey to a random sample of approximately 300 parent participants selected in equal numbers from the three program components defined by the age/grade categories of the participating children: (birth to 2 years, ages 3 to 5 years, grades K to 3). To ensure that sampled parents have had sufficient experience with the program to accurately answer the survey questions, they proposed only sampling parents who had participated in the Great by Eight program for at least six months. The survey would be administered online and could be taken on a home computer or cellphone. Computers would also be available for survey completion during parent events scheduled at the Way to Grow office, and Way to Grow staff would be present during these events to assist parents when literacy skills were a concern.

The school- and health-related topics to be included were presented in sections on the survey that pertain to the participating child's age.

School-related topics. The parent survey's school-related topics were:

All age/grade categories:
- Parent's role as child's first teacher
- Learning in the home (e.g., activities, games, songs)

---

[12] The delivery features that are hypothesized to be important were derived from responses given by focus groups conducted with program administrators and family educators in the implementation evaluation conducted in years one and two (2013/14 and 2014/15).

- Reading aloud to the child
- Parenting skills (e.g., nurturing, discipline, playing with the child)
- Advocating for the child
- Connections to resources and information

Prenatal to 2 years:
- (No additional school-related topics)

3 to 5 years (not yet in Kindergarten):
- Preschool referrals and enrollment
- Skills child needs to be ready for kindergarten
- School expectations for preschool through kindergarten
- Parent engagement in child care setting or preschool (e.g., volunteering, attending events, field trips)
- Participation in assessments and other testing

Kindergarten through 3rd grade:
- Skills the child needs to be a successful reader
- School expectations for kindergarten through 3rd grade
- Elementary school choice and enrollment
- Parent engagement in school (e.g., volunteering and attending events)
- Attending parent-teacher conferences

Health-related topics. The parent survey's health-related topics were:

All age/grade categories:
- Physical health and wellbeing (e.g., well child check-ups, safety)
- Nutrition

Prenatal to 2 years:
- Healthy pregnancy/prenatal care
- Caring for newborn/infant
- Infant/toddler growth and development
- Early childhood screenings

3 to 5 years (not yet in Kindergarten):
- Child growth and development

Kindergarten through 3rd grade:
- Dental care

Program delivery features. Research questions for the implementation evaluation of years one and two (2013/14 and 2014/15) did not address the program delivery features of the Great by Eight model. However, for years three and four, Great by Eight program staff felt it was important to identify the essential program delivery features in order to help ensure that

organizations wishing to replicate the program will have successful outcomes. The delivery features of the Great by Eight program to be assessed in the evaluation of years three and four (2015/16 and 2016/17) were identified by examining the results of the focus groups conducted for the implementation evaluation of years one and two (2013/14 and 2014/15). During the focus group conducted with program administrators, the researchers asked the administrators, if Great by Eight were to be replicated, what are the most essential qualities and components that needed to be included. The administrators participating in the focus group identified features related to the program's holistic approach, in-home visits, multi-lingual support staff, and year-round, multi-year support. Other essential features were identified in the focus groups with family educators, including understanding of and respect for the family's culture and connecting families with resources. Based on these focus group results, the evaluators hypothesized that the six following features are the essential delivery features of the Great by Eight model.

- **Holistic**: Great by Eight addresses the needs of the whole child and the whole family through a broad range of wrap-around services.
- **Home visits**: Family educators conduct home visits so families do not need to arrange for transportation and child care.
- **Year-round, multi-year support**: Great by Eight services are available to participating families year-round and throughout pregnancy and a child's early years (birth to age 8).
- **Connection to resources**: Great by Eight staff help families establish a stable, positive home environment by connecting them with internal and external community resources.
- **Multi-language support**: Family educators are multi-lingual and can converse with families in their native language.
- **Cultural sensitivity**: Family educators exhibit cultural sensitivity and have an understanding of and respect for each family's culture.

On the parent survey, respondents are asked to rate the importance of each feature based on their personal experiences with the program. The features are presented on the survey using wording that is easily understood by respondents. For example, the holistic feature is presented as "The program provides many different services and helps me connect with many community resources."

Family educator-parent relationship. The important aspects of the family educator-parent relationship were identified by examining the results of the family educator focus groups conducted for the implementation evaluation of years one and two (2013/14 and 2014/15) and by referring to literature on the qualities of an effective professional-client relationship. The qualities leading to a strong relationship between the parent and family educator were hypothesized to include some of following:

- The family educator really cares about the parent and his/her family.
- The family educator respects the family and the family's culture.
- The family educator is someone the parent can rely on to give assistance when it is needed.
- The family educator explains things using language that the parent can easily understand.

- The family educator answers the parent's questions thoroughly.
- The family educator gives the parent useful information when it is needed.
- The family educator helps the parent connect with useful community resources.

On the survey, parents are asked to rate the extent to which they feel their family educator demonstrates each of these qualities. In addition, the parents are asked to rate the overall quality of the relationship they have with their family educator.

General questions

Table C1. Years of participation in Way to Grow

Number of years	
Less than 1 year	7 (5.51%)
1 to 2 years	32 (25.2%)
3 to 4 years	30 (23.62%)
5 years or more	58 (45.67%)

Table C2. Number of children participating in Way to Grow

Number of children	
1	41 (32.03%)
2	45 (35.16%)
3	26 (20.31%)
4	9 (7.03%)
5	4 (3.13%)
6	2 (1.56%)
7	1 (.78%)

Table C3. Age/grade range of respondents' children during participation in Way to Grow in last 12 months

Age/grade range	
All ranges (birth through 3rd grade)	122 (95.31%)
Expecting a baby or ages birth to 2 years	81 (63.28%)
3 to 5 years old (not yet in kindergarten)	105 (82.68%)
Kindergarten through 3rd grade	98 (77.78%)

Questions for respondents with children age birth through 3rd grade while participating in Way to Grow in last 12 months

Table C4. Respondents with children age birth through 3rd grade: How much did you learn about these topics from your family educator during home visits?

	Unable to rate	Nothing	A little	A moderate amount	A lot
Parent's role as the child's first teacher (*n = 121*)	1 (0.83%)	--	2 (1.65%)	13 (10.74%)	105 (86.78%)
Learning in the home (for example, activities, games, songs) (*n = 122*)	--	1 (0.82%)	2 (1.64%)	15 (12.30%)	104 (85.25%)
Reading aloud to the child (*n = 122*)	1 (0.82%)	--	2 (1.64%)	15 (12.30%)	104 (85.25%)
Parenting skills (for example, nurturing, discipline, playing with the child) (*n = 122*)	1 (0.82%)	--	4 (3.28%)	20 (16.39%)	97 (79.51%)
Advocating for the child (*n = 122*)	1 (0.82%)	--	6 (4.92%)	17 (13.93%)	98 (80.33%)
Connections to resources and information (*n = 122*)	--	1 (0.82%)	3 (2.46%)	30 (24.59%)	88 (72.13%)
Immunizations (*n = 119*)	1 (0.84%)	2 (1.68%)	1 (0.84%)	22 (18.49%)	93 (78.15%)
Physical health and well being (for example, well child check-ups, safety) (*n = 122*)	1 (0.82%)	1 (0.82%)	5 (4.10%)	16 (13.11%)	99 (81.15%)
Nutrition (*n = 120*)	--	--	8 (6.67%)	17 (14.17%)	95 (79.17%)

Table C5. Respondents with children age birth through 3rd grade: If you selected a little or nothing for any of these home visit topics, please indicate why you chose that rating[13]

Response option	
I learned about the topic somewhere else, not during home visits	4 (4%)
The family educator shared information on the topic, but it was confusing or incomplete	2 (2%)
The family educator did not share information on the topic	2 (2%)
Other	9 (9%)
Not Applicable (I selected A lot, A moderate amount, or Unable to rate)	84 (84%)

---

[13] Respondents could select more than one option, so percentages may add up to more than 100%

Table C6. Respondents with children age birth through 3rd grade: Please indicate how useful you found each topic to be.

	Unable to rate	Not useful	Somewhat useful	Very useful
Parent's role as the child's first teacher (*n* = 115)	--	--	9 (7.83%)	106 (92.17%)
Learning in the home (for example, activities, games, songs) (*n* = 115)	--	--	13 (11.30%)	102 (88.70%)
Reading aloud to the child (*n* = 116)	1 (0.86%)	--	11 (9.48%)	104 (89.66%)
Parenting skills (for example, nurturing, discipline, playing with the child) (*n* = 116)	--	--	11 (9.48%)	105 (90.52%)
Advocating for the child (*n* = 114)	--	1 (0.88%)	15 (13.16%)	98 (85.96%)
Connections to resources and information (*n* = 116)	1 (0.86%)	1 (0.86%)	15 (12.93%)	99 (85.34%)
Immunizations (*n* = 116)	2 (1.72%)	2 (1.72%)	16 (13.79%)	96 (82.76%)
Physical health and well being (for example, well child check-ups, safety) (*n* = 116)	2 (1.72%)	2 (1.72%)	12 (10.34%)	100 (86.21%)
Nutrition (*n* = 115)	1 (0.87%)	--	19 (16.52%)	95 (82.61%)

Table C7. Respondents with children age birth through 3rd grade: If you selected not useful for any of these topics, please indicate why you chose that rating.[14]

Response option	
I don't remember receiving information on the topic	--
I did not understand the information or what was recommended	--
I understood the information but I was unable to do what was recommended	1 (1.08%)
I did not think the information applied to me or to my child	--
Other	3 (3.23%)
Not applicable (I selected Very useful, Somewhat useful, or Unable to rate)	89 (95.7%)

---

[14] Respondents could select more than one option, so percentages may add up to more than 100%

Questions for respondents expecting a baby or with children age birth through 2 years while participating in Way to Grow in last 12 months

Table C8. Respondents expecting a baby or with children age birth through 2 years: How much did you learn about these topics from your family educator during home visits?

	Unable to rate	Nothing	A little	A moderate amount	A lot
Healthy pregnancy/prenatal care (*n* = 80)	3 (3.75%)	1 (1.25%)	6 (7.50%)	14 (17.50%)	56 (70.00%)
Caring for newborn/infant (*n* = 80)	3 (3.75%)	2 (2.50%)	5 (6.25%)	11 (13.75%)	59 (73.75%)
Infant/toddler growth and development (*n* = 80)	1 (1.25%)	--	4 (5.00%)	8 (10.00%)	67 (83.75%)
Early childhood screenings (*n* = 80)	--	--	--	13 (16.25%)	67 (83.75%)

Table C9. Respondents expecting a baby or with children age birth through 2 years: If you selected a little or nothing for any of these home visit topics, please indicate why you chose that rating[15]

Response option	
I learned about the topic somewhere else, not during home visits	3 (4.76%)
The family educator shared information on the topic, but it was confusing or incomplete	2 (3.17%)
The family educator did not share information on the topic	--
Other	3 (4.76%)
Not Applicable (I selected A lot, A moderate amount, or Unable to rate)	37 (58.73%)

Table C10. Respondents expecting a baby or with children age birth through 2 years: Please indicate how useful you found each topic to be.

	Unable to rate	Not useful	Somewhat useful	Very useful
Healthy pregnancy/prenatal care (*n* = 77)	3 (3.90%)	3 (3.90%)	10 (12.99%)	61 (79.22%)
Caring for newborn/infant (*n* = 77)	3 (3.90%)	1 (1.30%)	15 (19.48%)	58 (75.32%)
Infant/toddler growth and development (*n* = 78)	1 (1.28%)	--	12 (15.38%)	65 (83.33%)
Early childhood screenings (*n* = 78)	--	--	8 (10.26%)	70 (89.74%)

---

[15] Respondents could select more than one option, so percentages may add up to more than 100%

Table C11. Respondents expecting a baby or with children age birth through 2 years: If you selected not useful for any of these topics, please indicate why you chose that rating.[16]

Response option	
I don't remember receiving information on the topic	3 (4.55%)
I did not understand the information or what was recommended	--
I understood the information but I was unable to do what was recommended	3 (4.55%)
I did not think the information applied to me or to my child	--
Other	1 (1.52%)
Not applicable (I selected Very useful, Somewhat useful, or Unable to rate)	60 (90.91%)

Questions for respondents with children ages 3 to 5 (pre-kindergarten) while participating in Way to Grow in last 12 months

Table C12. Respondents with children age 3 to 5: How much did you learn about these topics from your family educator during home visits?

	Unable to rate	Nothing	A little	A moderate amount	A lot
Preschool referrals and enrollment (*n* = 104)	2 (1.92%)	--	1 (0.96%)	19 (18.27%)	82 (78.85%)
Skills child needs to be ready for kindergarten (*n* = 103)	1 (0.97%)	--	1 (0.97%)	8 (7.77%)	93 (90.29%)
School expectations for preschool through kindergarten (*n* = 104)	2 (1.92%)	--	1 (0.96%)	9 (8.65%)	92 (88.46%)
Parent engagement in child care setting or preschool (for example, volunteering, attending events, conferences, or field trips) (*n* = 104)	1 (0.96%)	1 (0.96%)	2 (1.92%)	19 (18.27%)	81 (77.88%)
Participation in assessments and other testing (*n* = 104)	1 (0.96%)	2 (1.92%)	2 (1.92%)	13 (12.50%)	86 (82.69%)
Child growth and development (*n* = 103)	--	--	--	12 (11.65%)	91 (88.35%)

---

[16] Respondents could select more than one option, so percentages may add up to more than 100%

Table C13. Respondents with children age 3 to 5: If you selected a little or nothing for any of these home visit topics, please indicate why you chose that rating[17]

Response option	
I learned about the topic somewhere else, not during home visits	1 (1.3%)
The family educator shared information on the topic, but it was confusing or incomplete	1 (1.3%)
The family educator did not share information on the topic	--
Other	3 (3.9%)
Not Applicable (I selected A lot, A moderate amount, or Unable to rate)	72 (93.51%)

Table C14. Respondents with children age 3 to 5: Please indicate how useful you found each topic to be.

	Unable to rate	Not useful	Somewhat useful	Very useful
Preschool referrals and enrollment (*n* = 99	1 (1.01%)	--	10 (10.10%)	88 (88.89%)
Skills child needs to be ready for kindergarten (*n* = 99)	1 (1.01%)	--	8 (8.08%)	90 (90.91%)
School expectations for preschool through kindergarten (*n* = 99)	1 (1.01%)	--	10 (10.10%)	88 (88.89%)
Parent engagement in child care setting or preschool (for example, volunteering, attending events, conferences, or field trips) (*n* = 99)	1 (1.01%)	1 (1.01%)	13 (13.13%)	84 (84.85%)
Participation in assessments and other testing (*n* = 99)	1 (1.01%)	3 (3.03%)	12 (12.12%)	83 (83.84%)
Child growth and development (*n* = 99)	--	--	11 (11.11%)	88 (88.89%)

Table C15. Respondents with children age 3 to 5: If you selected not useful for any of these topics, please indicate why you chose that rating.[18]

Response option	
I don't remember receiving information on the topic	1 (1.22%)
I did not understand the information or what was recommended	--
I understood the information but I was unable to do what was recommended	2 (2.44%)
I did not think the information applied to me or to my child	--
Other	2 (2.44%)
Not applicable (I selected Very useful, Somewhat useful, or Unable to rate)	77 (93.9%)

---

[17] Respondents could select more than one option, so percentages may add up to more than 100%

[18] Respondents could select more than one option, so percentages may add up to more than 100%

Questions for respondents with children in kindergarten through 3rd grade while participating in Way to Grow in last 12 months

Table C16. Respondents with children in kindergarten through 3rd grade: How much did you learn about these topics from your family educator during home visits?

	Unable to rate	Nothing	A little	A moderate amount	A lot
Skills the child needs to be a successful reader (*n* = 97)	1 (1.03%)	--	1 (1.03%)	15 (15.46%)	80 (82.47%)
School expectations for kindergarten through 3rd grade (*n* = 97)	1 (1.03%)	1 (1.03%)	--	17 (17.53%)	78 (80.41%)
Elementary school choice and enrollment (*n* = 94)	2 (2.13%)	2 (2.13%)	3 (3.19%)	17 (18.09%)	70 (74.47%)
Parent engagement in school (for example, volunteering and attending events) (*n* = 97)	1 (1.03%)	1 (1.03%)	3 (3.09%)	19 (19.59%)	73 (75.26%)
Attending parent-teacher conferences (*n* = 97)	1 (1.03%)	1 (1.03%)	2 (2.06%)	8 (8.25%)	85 (87.63%)
Dental care (*n* = 95)	2 (2.11%)	1 (1.05%)	2 (2.11%)	17 (17.89%)	73 (76.84%)

Table C17. Respondents with children in kindergarten through 3rd grade: If you selected a little or nothing for any of these home visit topics, please indicate why you chose that rating[19]

Response option	
I learned about the topic somewhere else, not during home visits	3 (3.9%)
The family educator shared information on the topic, but it was confusing or incomplete	--
The family educator did not share information on the topic	--
Other	2 (2.6%)
Not Applicable (I selected A lot, A moderate amount, or Unable to rate)	73 (94.81%)

Table C18. Respondents with children in kindergarten through 3rd grade: Please indicate how useful you found each topic to be.

	Unable to rate	Not useful	Somewhat useful	Very useful
Skills the child needs to be a successful reader (*n* = 93)	1 (1.08%)	--	5 (5.38%)	87 (93.55%)
School expectations for kindergarten through 3rd grade (*n* = 93)	1 (1.08%)	--	14 (15.05%)	78 (83.87%)
Elementary school choice and enrollment (*n* = 94)	3 (3.19%)	1 (1.06%)	16 (17.02%)	74 (78.72%)
Parent engagement in school (for example, volunteering and attending events) (*n* = 94)	1 (1.06%)	1 (1.06%)	17 (18.09%)	75 (79.79%)
Attending parent-teacher conferences (*n* = 92)	1 (1.09%)	1 (1.09%)	6 (6.52%)	84 (91.30%)
Dental care (*n* = 92)	4 (4.35%)	1 (1.09%)	11 (11.96%)	76 (82.61%)

---

[19] Respondents could select more than one option, so percentages may add up to more than 100%

Table C19. Respondents with children in kindergarten through 3rd grade: If you selected not useful for any of these topics, please indicate why you chose that rating.[20]

Response option	
I don't remember receiving information on the topic	--
I did not understand the information or what was recommended	--
I understood the information but I was unable to do what was recommended	--
I did not think the information applied to me or to my child	--
Other	2 (2.63%)
Not applicable (I selected Very useful, Somewhat useful, or Unable to rate)	74 (97.37%)

Questions about family educator

Table C20. How well do you think your family educator interacts with you and your family?

My family educator…	Unable to rate	Never	Seldom	Usually	Always
Makes me feel that s/he really cares about me and my family (*n* = 123)	--	--	2 (1.63%)	8 (6.50%)	113 (91.87%)
Respects my family and my culture (*n* = 124)	--	--	--	7 (5.65%)	117 (94.35%)
Is someone I can rely on to give me assistance when I need it (*n* = 124)	--	--	1 (0.81%)	9 (7.26%)	114 (91.94%)
Explains things using language that I can easily understand (*n* = 123)	--	--	2 (1.63%)	4 (3.25%)	117 (95.12%)
Answers my questions thoroughly (*n* = 124)	--	--	1 (0.81%)	6 (4.84%)	117 (94.35%)
Gives me useful information when I need it (*n* = 124)	--	--	2 (1.61%)	7 (5.65%)	115 (92.74%)
Helps me connect with useful community resources (*n* = 124)	--	--	2 (1.61%)	10 (8.06%)	112 (90.32%)

Table C21. Overall, how would you rate the quality of the relationship that you have with your family educator?

Response option	
Unable to rate	--
Poor	--
Fair	1 (.83%)
Good	13 (10.74%)
Excellent	107 (88.43%)

---

[20] Respondents could select more than one option, so percentages may add up to more than 100%

Important features of the program

Table C22. How important is this feature for you?

	Unable to rate	Very unimportant	Unimportant	Important	Very important
The program provides many different services and helps me connect with many community resources (_n_ = 123)	--	--	1 (0.81%)	19 (15.45%)	103 (83.74%)
The family educator comes to my home (_n_ = 120)	1 (0.83%)	--	--	14 (11.67%)	105 (87.50%)
The program provides year-round support over many years (_n_ = 123)	--	--	--	16 (13.01%)	107 (86.99%)
The program helps me to meet my basic needs such as housing, food, and medical attention (_n_ = 123)	5 (4.07%)	--	2 (1.63%)	28 (22.76%)	88 (71.54%)
The family educator is bilingual and can converse with me in my native language (_n_ = 123)	13 (10.57%)	1 (0.81%)	8 (6.50%)	14 (11.38%)	87 (70.73%)
The family educator understands and has respect for my family's culture (_n_ = 123)	--	--	--	11 (8.94%)	112 (91.06%)