NURTURES Phase II
Evaluation Report: Year 2
August 1, 2018 – July 31, 2019

Prepared by:

Acumen Research and Evaluation, LLC
June 28, 2019

On the behalf of:

THE UNIVERSITY OF TOLEDO 1872

National Science Foundation

Nurtures

Table of Contents

Executive Summary

This report evaluates the NURTURES II project from July 2018 through June 2019 (Year 2). The evaluation model remained consistent with Year 1 but the TerraNova achievement test was substituted with the Gates-MacGinitie Reading Tests (GMRT) to verify student group equivalency and the Galileo K-12 tests to compare student learning outcomes.

Cohort 1 included 34 teacher participants (19 professional development (PD) and 15 professional development plus Family Science (PDFS)) and 17 control teachers. The distribution by grade level was not balanced as expected due to some participant teachers requesting to postpone until Cohort 2 and the difficulty in obtaining control teachers. Recruitment for Cohort 2 has concluded resulting in a more balanced sample that, when combined with Cohort 1 will provide an overall adequate sample.

NURTURES II had a positive impact on teachers attitudes about teaching science in spite of the fact that teachers from all three groups scored high on the pre-test (Preschool Teacher Attitudes and Beliefs towards Science). The test measures teachers' perceptions of their comfort level teaching science, the benefits to the child to learn science in early years, and challenges to teaching science to young learners. Participant post-test scores were near the maximum with the greatest gains made in how teachers perceive and handle challenges in teaching science.

Video-recordings of a random sample of teachers were scored using the Electronic Quality of Inquiry Protocol (EQUIP)—a tool that measures inquiry-based instruction based upon Next Generation Science Standards. The sample included four PD, 3 PDFS, and six control teachers. Because all teachers improved overall between the pre-test and post-test, conclusions cannot be drawn as to how NURTURES II influenced teaching. However, the purpose of the EQUIP observations was to determine fidelity of treatment implementation (do the teachers implement their lessons with fidelity to the NURTURES II model?). In this case, the PD group achieved a level of proficiency on all scales for their spring lessons—maintaining a proficient score for instructional and improving from developmental compared with fall assessment. The PDFS group was proficient on the instructional and curriculum scales (increased from developing) but maintained a rank of developing on the assessment and discourse scales.

The curriculum and instruction scales on the EQUIP also provide insight into teacher science and engineering practice knowledge and integration into lessons. Both participant groups scored proficient on the post-observation for those scales.

Teachers in the PDFS group hosted two Family Science Events (Events) at their schools. The evaluation determined that when the teachers implemented the Events with fidelity to the NURTURES II guide, the Event was successful and implemented with minimal challenges. The major weaknesses of those who did not follow the guide included not recruiting enough facilitators to assist at each of the activities, not setting up in advance of the Event, and having a layout that may have caused confusion (e.g., either scheduling the Event when another school activity was scheduled, having the activities too close together, or not allowing enough space for certain activities like testing designs).

Feedback from families regarding Family Science Packs was low (about 4%) so generalizations cannot be made as to family fidelity of implementation of the Packs. However, those who did respond indicated that they engaged in the Pack as designed and found the experience to generate participation in other science and engineering related programs such as visiting a museum or park.

Students in the three groups measured equivalent on the GMRT. The post-test, the Galileo, was analyzed by the research team and findings can be found in the annual report. While there were suggestions that a statistically significant effect was realized by first grade students in the treatment groups, this assertion should be made with caution as there was only one control group classroom for comparison. The addition of the 2019-20 Cohort data will increase the sample sizes to an adequate level to fairly represent the target population.

NURTURES II leadership should examine how they might promote discourse and assessment in teaching (as measured by the EQUIP—see Table 4 in report) for Cohort 2. The necessity of preparing for the Family Events should be emphasized and perhaps Cohort 1 teachers could share their experiences with Cohort 2 PDFS teachers prior to hosting the 2019 fall events. Efforts to gather more feedback from families who use the Family Packs should be considered.

Overall, NURTURES II has been enacted with fidelity to its model. Teachers in general have realized gains in their confidence teaching science, have implemented lessons with fidelity to the model, and have exhibited knowledge of science and engineering practices. Student outcomes will be better assessed at the conclusion of Cohort 2 participation.

1) Evaluation Model

Minor changes were made to the evaluation model in the second year. Specifically, the TerraNova achievement test was substituted with two reliable instruments with evidence of validity. Baseline assessment of the two treatment and the control group students was measured with the Gates-MacGinitie Reading Tests (GMRT). The GMRT is aligned with national and state reading standards for grades K – 12 and it tests general levels of vocabulary and reading comprehension. Research has shown a strong positive relationship between reading comprehension and science and math achievement (e.g., Akbasli, Sahin, & Yaykiran, 2016). This instrument was administered in the fall 2018 to enforce the assumption of group equivalency based upon random assignment. The posttest—the Galileo K-12 (Assessment Technology Incorporated) measured science knowledge for grades 1 -3 and was administered in May 2019. Detailed explanations of the instruments and findings are provided in the NURTURES II annual report. All other evaluation measures were retained and a complete listing can be found in the 2018-19 Evaluation Report.

This report includes follow up findings from the 2018 Summer Institute (SI) participants and their students. While Cohort 2 (2019 SI) baseline participant data has been collected, details and findings of analyses will be included in the 2019-20 Evaluation Report.

2) Project Participants

Cohort 1: The 2018-19 participants (n = 34) were randomly selected using a stratified (grade) cluster sampling by school from a population of all applicants (n=57). The remaining 17 teachers served as the control group and were offered first option to attend the 2019 SI. As noted in the previous annual evaluation report, because grade level groups were unequal (see Table 1), some grade levels had a higher representation both within and across groups. This was due, in part, to the cluster assignment for treatment groups to avoid treatment contamination. The groups are PD only, PD plus family science (PDFS), and Control. All participants in Cohort 1 were female. Other demographic information (race, educational level, relevant PD other than NURTURES II) was collected fall 2018 and used to verify group equivalency. The groups were statistically equivalent on demographics. For example, when looking at the percent of teachers with a Masters degree, the control and PD had 26% and the PDFS had 21%.

Table 1. 2018 Summer Institute participant demographics

Grade	PD only	PD + FS	Control
PK	6	7	10
K	5	1	3
1	3	2	2
2	3	2	2
3	2	3	0
Total	19	15	17

A. Teacher Beliefs and Attitudes

P-TABS: NURTURES II teacher participants completed three administrations of the P-TABS instrument: before the start of the summer institute (July 2018), again five months later (end of November 2018), and, finally, ten months after the summer institute (May 2019). The P-

TABS, Preschool Teacher Attitudes and Beliefs towards Science (Maier, Greenfield, Bulotsky-Shearer, 2013), measures teachers' perceptions of their comfort level teaching science, the benefits to the child to learn science in early years, and challenges to teaching science to young learners. Specifically, P-TABS measures the constructs of comfort level teaching science (CT), perceived challenges (PC), and child benefits (CB).

P-TABS uses a four-point Likert agreement scale (Strongly Agree, Mildly Agree, Mildly Disagree, and Strongly Disagree). The CT construct consists of 15 items and the assumed item mean was 37.5 (i.e. a score of 15 would be the lowest possible score, indicating that the teacher strongly disagreed with every item and therefore would exhibit the lowest level of teacher comfort, whereas a score of 60 would indicate strong agreement with every item and a teacher possessing the highest level of comfort). Similarly, the 9 PC items produced an assumed item mean of 22.5 (range 9 – 36); and the 11 CB items resulted in an assumed item mean of 27.5 (range 11 – 44).

Table 2. Between-group comparisons of P-TABS constructs

Comfort teaching Science

	Group	Mean	S.D.	t-test
Pre-SI (July 2018)	PD	49.6	4.9	t = -1.180; p = .250
	PDFS	47.1	5.6	
Post-SI 1 (November 2018)	**PD**	**52.8**	4.4	F = 4.838; **p = .013**
	PDFS	**51.8**	5.1	
	Control	**47.1**	6.4	
Post-SI 2 (May 2019)	PD	54.5	2.2	t = -1.035; p = .311
	PDFS	53	4.8	

Perceived challenges

	Group	Mean	S.D.	t-test
Pre-SI (July 2018)	PD	23.4	4.5	t = -.051; p = .960
	PDFS	23.3	3.6	
Post-SI 1 (November 2018)	PD	28.1	4.1	F = 3.019; p = .060
	PDFS	27.6	4.4	
	Control	24.7	4.2	
Post-SI 2 (May 2019)	PD	29.5	4	t = 1.252; p = .311
	PDFS	31.4	3.3	

Benefits for the children

	Group	Mean	S.D.	t-test
Pre-SI (July 2018)	PD	40.5	3.4	t = .752; p = .692
	PDFS	41	2.7	
Post-SI 1 (November 2018)	PD	43.4	1.2	F = 2.977; p = .062
	PDFS	42.6	2.1	
	Control	41.5	2.7	
Post-SI 2 (May 2019)	PD	43.6	1.1	t = -.999; p = .328
	PDFS	42.9	2.2	

Teachers who completed the survey at all three occasions were included: 14 (out of 19—74%) PD teachers and 11 (out of 15—73%) PDFS teachers, for a total of 25 (out of 34—73.5%) teachers. The 18 control teachers who completed the instrument around the same time as the first Post-SI data collection period were included in a between-groups analysis to provide a fuller frame of reference—to not only examine growth but to provide a comparison between treatment and non-treatment. Table 2 reports the independent t-test comparisons of the three constructs measured by P-TABS, as measured for each group prior over the year. Higher measure scores imply more favorable participant attitudes. Every group scored above the assumed mean average on all of the measures. These scores indicate a group of confident teachers who are able to

address any perceived challenges and understand the benefits of science education for their students.

The PD and PDFS groups are not statistically different on any of the three measures across any of the three time periods. Small sample sizes may introduce measurement error that could mask actual effects—Type II error—although the intervention is basically the same for both groups with the exception of offering family science. Both PD and PDFS groups are statistically higher than the control group on the CT, or global measure of comfort teaching science. Although there is a statistically significant difference, a fuller analysis and consultation with the program staff will determine whether or not this has meaningful significance.

Investigating the results of the dependent t-tests shown in Table 3, highlights that there are notable within-group developments over the course of the year for both the PD and the PDFS groups. Specifically, both groups made significant gains across each of the three measures. The findings indicate that ***the teachers grew from a high level to a nearly-maximal level of understanding the benefits of teaching science to their students***. Scores indicated that the greatest gains were made in how teachers perceive and handle challenges in teaching science, moving from the "moderate" teacher attitude range to a more competent, resourceful range on average.

Table 3. Between-group comparisons of P-TABS constructs

PD within-group comparison

	CT	PC	CB
Pre-SI	49.6	23.4	40.5
Post-SI 1	**52.8***	**28.1****	**43.4***
Post-SI 2	**54.5****	**29.5***	**43.6***

*Significant at $p < .01$ compared to pre-SI
**Significant at $p < .001$ compared to pre-SI

PDFS within-group comparison

	CT	PC	CB
Pre-SI	47.1	23.3	41
Post-SI 1	**51.8***	**27.6****	**42.6***
Post-SI 2	**53****	**31.4****	**42.9***

*Significant at $p < .05$ compared to pre-SI
**Significant at $p < .01$ compared to pre-SI

These findings will enable a conversation with program staff to determine how the statistically significant gains expressed by these measures translate into actual pedagogy. Additionally, these findings alert program staff to the situation where additional attention and time could be spent on helping both the PD and PDFS groups of teachers address and overcome perceived challenges, as both groups are already exhibiting particularly high levels of the other two constructs.

B. Teaching Practice

Pre- and post-participation videos of teachers teaching science in their classrooms examined the effect participation in NURTURES II may have on teaching practice. A stratified random sample of teachers represented the various grade levels and school types (rural vs urban). The resulting observation sample of 7 treatment teachers included the following:

PD only: 1 PK 1 K 1 1st 1 2nd
PD + FS: 1 1st 1 2nd 1 3rd

The sampling plan was to include 9 PD and 9 PDFS teachers with 3 teachers representing PK, K-1, and 2-3 per group. The original sample was randomly selected by Acumen. However, due to practical issues including teachers requesting to participate during a different year, the original sample had to be modified. Efforts to improve the distribution were made for recruiting Cohort 2.

To determine the fidelity of teacher implementation of instructional strategies learned as a result of participation in NURTURES 2 (i.e., the use of inquiry-based strategies and Next Generation Science Standards—NGSS), pre- (spring 2018) and post-participation (fall/winter 2018) video-recorded classroom observations were scored using the Electronic Quality of Inquiry Protocol (EQUIP) (Inquiry In Motion, Clemson University).

The EQUIP rubric measures four factors associated with inquiry instruction and based upon NGSS—instruction factors, discourse factors, assessment factors, and curriculum factors. Within these four factors are 19 indicators. Scores on pre- and post-participation observations were compared first to determine if there were patterns of proficiency among the time-based observations and second to determine if there were areas of improvement between the observations. The rubric included four levels—pre-inquiry, developing, proficient, and exemplary—and therefore the data is ordinal. To analyze, median scores were examined and are provided in Table 4. Pre-observation scores placed all groups roughly at the developing level except for the Instructional category where treatment teachers were scored proficient.

Table 4. EQUIP median score

		Pre-Observation		
Group	***Instructional***	***Discourse***	***Assessment***	***Curriculum***
PD	3	2.25	2	2.5
PDFS	3	2	2	2
		Post-Observation		
Group	***Instructional***	***Discourse***	***Assessment***	***Curriculum***
PD	3	3	3	3
PDFS	3	2	2	3

Post-observation scores showed improvement for the PD teachers who moved to proficient in all categories indicating that their teaching reflected NGSS best practices. Two PD teachers scored exemplary in the instructional category. The PDFS group improved to proficient in the curriculum and instructional categories and remained at developing in the other two. PD and PDFS teachers' scores in instructional and curriculum categories at the post-observation verified that fidelity of implementation was maintained for those categories and the PD group reached an acceptable measure of fidelity on all four categories. This information can be triangulated with student outcome measures to verify effects of the NURTURES intervention.

C. Teacher Science and Engineering Practice (SEP)

The instruction and curriculum scales on the EQUIP are geared towards examining teacher understanding and implementation of SEP. The instruction scale examines inquiry-based instructional strategies that lead to an experiential student learning experience. The curriculum

scale examines the relationship between content and investigation and how it is linked to student-centered learning. As noted in the previous section on the Teaching Practice, both PD and PDFS scored at a proficient level on these two scales indicating an acceptable level of SEP understanding.

3) Family Science

The added effects of family science were examined several ways. First, because the Family Science Events were hosted and implemented by the teachers and their schools, fidelity of implementation was verified by direct observation. Acumen staff attended the Family events and scored the sessions using a fidelity of implementation rubric based upon NURTURES II Family event guidelines. Family Packs were distributed in the PDFS classrooms. These self-contained family science activities provided parents with a guide and all needed materials to conduct a mini-science experiment with their child. To examine whether the Packs were used as intended and to gain insight into the family experience, Family Pack worksheets were collected and reviewed. In addition, formative evaluation of Family Pack use included a parent survey.

A. Family Events (Events)

Fall 2018: During the fall 2018, seven NURTURES II Events were observed. Table 5 provides the basic demographics for each of the observed Events. Teachers were given materials for three family science activities—Build a Boat, Terrarium, and Pizza Box Motion Maze. The goal for Build a Boat was for the family to construct a boat from aluminum foil that would float and would also carry a payload (washers). Families were encouraged to experiment with different designs and to gradually add washers to see how many their boat would hold. The Terrarium activity involved discussions about seeds and what is needed to make them grow. Families used an empty plastic bottle, soil, water, and seeds to create their terrariums. The Pizza Box Maze activity challenged families to design and create a maze within a pizza box using pieces of cardboard. The maze was to include a pathway for a small ball to roll from a starting to an ending point. Children were to explore how tilting the box in various directions would cause the ball to roll. Each activity included written Guides for parents that encouraged discourse and open-ended questioning of their child as they engaged in the activity.

Table 5: NURTURES 2 Fall Family event attendance

School	*Time*	*Attendance*	*Activities*	*Grade*	*Facilitators*
Sterling	Weekday	14 students	All three	PreK	1 teacher and 6 adults*
St. Mary	Evening	6 families	Boat; Terrarium	PreK	1 teacher
Fostoria	Evening	7 families	All three	PreK	1 teacher
Norwalk	Evening	7 families	Boat; Maze	3rd	1 teacher
Otsego	Evening	18 families	All 3	K-3	5 teachers
St. Joseph	Evening	35 families	All 3	2nd	2 teachers
St. Pious	Evening	7 families	All 3	1st	1 teacher

*Because the Sterling Event was held during class time, adults were recruited from the families of students to assist. All students in the class participated.

Fidelity of Implementation

Previous evaluations of the original NURTURES project provided evidence that, when implemented as designed, the Event activities encouraged parents to engage their children in balanced interaction (where both parent and child contributed to the discussion and activity nearly equally). The discourse between parent and child included question/short answer, "why" questions, and other open-ended questions. Parents also made frequent use of the written parent Guides to get ideas of questions to ask their children. The original design for NURTURES Events included facilitators at each activity who were familiar with the activity and were comfortable prompting parents in (or modeling) questioning techniques to engage their children in experiential learning.

The first step to determining the positive effects of NURTURES II Events was to verify the degree to which they adhered to the original, successful design. Referring to Table 5, only two Events had at least one facilitator for each of the activities (Sterling and Otsego). It is not known if the parents and grandparents assisting the Sterling Event were provided instruction as to how to facilitate the activities; however, their role was not so much as facilitator for a family but rather as surrogate parents to work directly with the children because the Event was held during the day when many parents were working and could not attend. As a result, the Sterling Event did not address the essential goal of NURTURES II—to improve family engagement—because in most cases families were not present.

While the Events were similar in that they included NURTURES activities, materials, and Guides, there were differences in implementation that affected how smoothly the Event ran as well as the degree to which family discourse was encouraged and enacted. Events that had the most success with getting parents to use the Guides (including asking the types of questions suggested) were those that introduced the purpose of the Event to the parents prior to engaging in the activities. In most cases this occurred with individual families as they arrived. Sometimes, even when provided with the purpose of the Event, parents either viewed the Guide only as far as understanding how to complete the activity (so no questioning or discourse occurred) or totally ignored the Guide. When families were not formally greeted as they entered, parents and children milled about and it was difficult for the teacher to address the entire group. *The role of the facilitator was vital in this situation as it was observed that when a facilitator was present and modelled questioning strategies, parents began to follow suit by asking higher order, open-ended questions.*

At every school observed, teachers served as facilitators. When there were fewer teachers than activities, parent/child interaction at "unmanned" activities was not balanced and many families simply began without referencing the Guide. To be fair, most teachers stated that they were surprised at the attendance so perhaps did not anticipate that more facilitators would be needed. However, regardless of the turnout, the chance that families would engage in an activity without a facilitator was high when there were more activities than facilitators. To address this,

one teacher moved families through the activities as a group. That caused quite a bit of confusion, however, because each activity was crowded. In other situations, the teachers were busy making sure there were enough supplies at the tables, were harnessing children who were running about, or were talking with parents about school-related subjects but not the Event activities. Parents seized the opportunity to discuss their child's progress in school which infringed on the teacher's role as activity facilitator. One school's Event drew several large families. In this situation, the number of siblings (older and younger than the target ages) caused quite a bit of chaos.

Observations of the three activities revealed several common themes:

Build a Boat: This was by far the most popular activity regardless of the child's age. Of all the activities, this one showed the most balanced overall interaction between parent and child. Most parents read through the Guide prior to beginning the activity but did not complete the design phase (draw your boat prior to building it). Rather, they jumped directly to building. With the youngest children, the parents built the boat and asked the child if it looked like what the child wanted. Once the boat was built, most of the children took over testing it in water and adding weights. Only a few families overall tried to redesign their boat to see if it could carry more weight. For one Event, the containers of water were not quite large enough for some of the boats. This is something that teachers should consider as they plan. The use of the graph to log the trials was mixed. One school had an excellent graphing system that encouraged families to post their results while most either left graphing to the family or didn't do it at all. One teacher did mention that more information would have been helpful regarding that aspect of the Boat activity.

Terrarium: Family interaction for this activity was dependent upon the age of the child. Younger children were observed helping with the measuring and older children completed the activity as the parent read instructions. In many cases, however, the actual construction of the terrarium was completed by the parent and the child's interest waxed and waned. There was a suggestion from a parent that a terrarium with plants growing in it would help children visualize between parents and children were mostly questions about what a plant needs to grow (so one-word responses).

Pizza Box Motion Maze: This activity was the least balanced between parent and child. While families designed and constructed mazes, emphasis of the activity was on experimenting with motion. Designing and constructing the maze was too advanced for most of the children so parents completed this. Many of the children were not engaged. In fact, the most enthusiasm was by older siblings who made their own mazes. Once the mazes were constructed, some younger children were observed

playing with them. This activity did not result in much discourse—the parent built the maze; the child rolled a ball through it.

Teacher Post-Event Interviews: At the conclusion of the Event, teachers were asked to reflect upon how well it went and what they might do differently in the future. All the teachers were pleased with the turnout with many having more families than anticipated. This being said, many noted that keeping families on task was an issue whether it be through more facilitators, better activity layout (enough tables per activity to accommodate a larger group), smaller rooms (to keep children from running about), educating parents as to the goal or purpose of the Event prior to hosting it, or having other activities available for older or younger children.

School support was an important element whether it be through the school administration or through other teachers. The Events that teachers felt the best about were those that involved a group effort, where a lone teacher was not running the Event solo. Of course, the Events with the largest turnouts invited children from more than one classroom so there was a larger population from which to draw and these were also the Events that had multiple teachers planning and implementing them. One school had so much support for the Event that all other extracurricular activities were cancelled that night.

Fall 2018 Conclusions

Teachers should consider a few key elements prior to implementing future events. First teachers should reflect upon the age-appropriateness of family activities prior to deciding which to include. While NURTURES II targets PreK through grade 3, some activities have varying degrees of effectiveness depending upon age..

Teachers should also consider adding facilitators to the Event. These could be other teachers in the school or volunteers such as the parents club or even local high school students who are often required to do public service. Prior to hosting the event, these facilitators should gain knowledge as to their roles as facilitators.

Finally, more specific details from the NURTURES project regarding set up and implementation of the Events would be helpful. Considerations like room size, supplies, how to introduce the Event to parents so they understand its purpose, and examples of ways to engage older and younger siblings should be considered. These findings were presented to NURTURES II leadership January 2019. Examination of the spring 2019 Events illustrate improvements that were made.

Spring 2019: Eight events were observed in the spring. Cardinal Stritch was not observed fall 2018 as this teacher did both Events in the spring. Spring Events provided teachers with four activities: Foam Rockets, Where is Froggy?, Pour-A-Pond, and Paper Tower. Table 6 provides the demographic information for each event.

Table 6. Spring 2019 Event attendance

School	*Time*	*Attendance*	*Activities*	*Grade*	*Facilitators*
Cardinal Stritch	After school	26 children; 5 parents	Rocket, Frog, Tower	3	1 teacher and 1 graduate student
Sterling	During school	13 children, 8 parents	Rocket, Pond, Tower	PreK	1 teacher and 2 volunteers

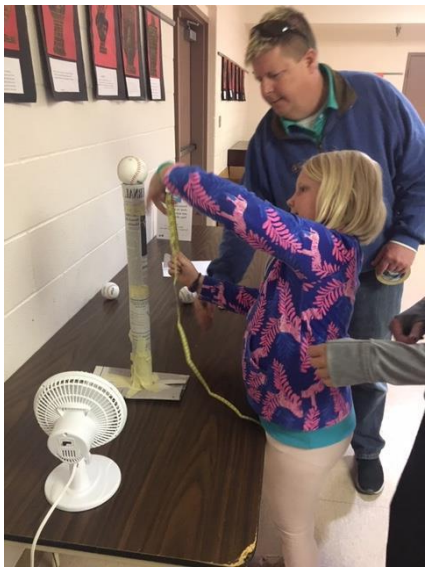
St. Mary	Evening	Evening	Rocket, Pond, Tower	PreK	1 teacher
Fostoria	Evening	7 families	Rocket, Frog, Tower	PreK	2 teachers
Norwalk	Evening	14 families	Rocket, Frog, Tower	2-3	1 teacher, 3 volunteers
Otsego	Evening	20 families	Rocket, Frog, Tower	K-3	5 teachers
St. Joseph	Evening	28 families	Rocket, Frog, Tower	2nd	2 teachers and 1 aid
St. Pious	Evening	9 families	Rocket, Pond, Tower	1st	1 teacher

Improvements to fidelity of implementation: In the fall, there were fewer teachers/facilitators than activities at six of the seven observations (86%). In spring, an improvement was observed with four of eight having fewer facilitators than activities (50%). Teachers who were able to increase the number of facilitators were creative in recruiting adult volunteers and older students to assist. When the teacher's attention was split, family discourse and implementation of the activities as designed were not consistent with intentions. Some Events were successful in spite of having fewer facilitators than activities because families in general had attended the fall Event and were familiar with the process. One teacher was able to facilitate efficiently by moving the entire attendance through each activity as a group. It should be noted that this event had moderate attendance.

Teachers at six of the eight events (75%) welcomed families with explanations about the Event, its purpose, and how to engage in the activities. The same six also had adequate set up of activities that promoted a positive experience. The activities themselves, engaged families to different degrees mostly based upon the age of the children and upon whether the activity was clearly explained prior to engagement:

Foam Rockets: Families built a foam rocket following instructions and then tested it by shooting with a rubber band across an open area. Families then measured how far their rockets went. When a facilitator was present, this activity was completed smoothly although parents often took the lead in the actual building, particularly when the children were in PreK and K. The guides were referenced by parents during this activity and parents encouraged their children to make predictions such as, "How far do you think your rocket will go?" and "Do you think it will go farther if you point the nose towards the ceiling?"

Paper Tower: This activity challenged families to design a tower built of newspaper that would hold a baseball and would withstand a breeze generated by a fan. As occurred in the original NURTURES project, parents often skipped the design portion of the activity and went straight to building. The design phase was observed during a few Events; however, often the design was abandoned once building began. This should not be viewed as a failing, however, as building often reveals design flaws. Success of this activity was highly dependent upon whether a facilitator was present to provide an overview and make suggestions because without this guidance parents were often at a loss for what to do. As with the Rocket activity, younger children (PreK and K) were less likely to be engaged in the activity or in building the tower. At some Events families resorted to just rolling up newspaper (see picture) but at a few, interesting designs were utilized.



Where is Froggy?: This life cycle activity used a mat with several habitats printed on it and plastic versions of the stages of a frog. Parents were encouraged to ask children questions about which phase of frog would be found or survive in which habitat. Suggested questions from the guide required that the children explain why they felt the particular stage of frog could be found in the habitat. Engagement in this activity varied with some parents actively questioning children and others just allowing children to play with the animals.

When a facilitator was not present, there was more variance in the quality of the experience. In some cases, children were also not accompanied by a parent (in the case where the Event took place during school and another Event where families attended but children younger than the target age went to another location in the school to do crafts) and when that occurred, children lost interest easily.

Pour-A- Pond: Teachers provided water in tubs from two or three different sources (e.g., tap water and pond water) for families to explore. Families were engaged in exploring the differences in what they observed from the water samples using scientific tools. The guide provided a space for them to record what they found. This activity was generally implemented as designed.

Teacher Post Event Interviews: All of the teachers felt that the families were engaged and enjoyed their experience. They were pleased at how well parents worked with their children and thought the activities were conducive to that. A few remarked that they had hoped for a better turnout but admitted that they did not recruit as

strongly as they could have. A few teachers were enthusiastic about using the rest of their materials for other family activities or in the classroom. All felt the Events were worth repeating and most planned on doing so next year.

Spring 2019 Conclusions

The teachers who learned from their fall experience exhibited greater Family Event fidelity of implementation than those who were less prepared. Teachers who provided an explanation of the Event process and purpose, who had adequate facilitators to assist at each activity, and who had carefully thought about how the activities should be set up including how to move families from one activity to another (such as how to set up areas for testing things like flying rockets and splashing in pond water) provided their families with a quality experience that not only encouraged science exploration but also engaged families in higher level discourse and, frankly, just a good time together.

B. Family Science Packs

Family Science Packs were distributed to students in classrooms of the PDFS teachers. The role of evaluation is to determine the nature of how families engaged in the packs. The packs themselves include instructions for a family science activity, suggestions to parents as to what types of questions to ask while engaged in the activity, and all materials needed to complete the activity. To gather that information, a Family Pack survey was included in each pack. Of the 15 classrooms, only 10 family surveys were returned—not a representative sample size to draw conclusions about the experience. However, some insight might be gained by examining those that were returned. Table 7 provides the date and grade level of the child in each family that responded. The majority of responding families had children in the second grade. Grandparents participated in two of the instances and siblings joined in another two instances. On average, it took families 20 to 30 minutes to complete the activity. Three families indicated that they engaged in science related family activities weekly, one confessed they never engaged in science related activities, and the remainder responded that it was a monthly occurrence. Most parents said the experience deepened their understanding of how to work with their child on science related activities and inspired new questions about science and engineering content. Most used the guide's suggested questions as they completed the activity and eight of the 10 families indicated they were planning to follow up the activity with a visit to an informal science location such as a Science Museum or a local garden.

Table 7. Family Pack Respondents

Date	*Grade*
10/24/2018	PreK
11/5/2018	2nd
11/5/2018	2nd
11/5/2018	3rd
11/11/2018	1st
11/13/2018	2nd
11/13/2018	2nd
11/14/2018	2nd
1/23/2019	2nd
2/15/2019	2nd

The responses from the families indicate that the Family Packs were used as designed. Families enjoyed the experience and parents felt that it enhanced their child's understanding of science as well as their own.

4) Student Outcomes

Group Equivalency: The Gates-MacGinitie Reading Test (GMRT) was given to students in classes of teachers who comprised the three groups. The purpose of this test was to determine pre-intervention group equivalency of students. This test is not available for preschool and kindergarten children so only grades 1, 2, and 3 were tested. Specific details of the test and analysis can be found in the 2019 NURTURES II Annual Report. Overall, in each comparison, there was no statistical evidence to indicate the groups were not equivalent at the onset of the school year (and prior to teacher introducing NURTURES science lessons and Family Science). However, results should be viewed with caution due to the unequal distribution of sample sizes within each of the three groups. Project leadership was aware of the unequal distribution and has made an effort to gather more equal sample sizes for the 2019-20 cohort.

Student Science Proficiency: The NURTURES II research team measured student science proficiency using three instruments: the Galileo, the Kindergarten Early Learning Scale, and the Early Learning Scale (PreK). Details of the assessments and analyses can be found in the project's annual report. Current findings are inconclusive due to the small and unequal sample sizes. However, with the addition of the 2019-20 Cohort data, the sample sizes should be adequate to fairly represent the target population.

5) Conclusions/Recommendations

Teacher attitudes towards teaching science improved statistically significantly after participating in NURTURES II (P-TABS instrument). PD and PDFS experienced similar gains in their confidence in teaching science. While the teachers in all three groups scored high on the pre-test, post-test scores showed that the PD and PDFS teachers' confidence levels were elevated to near maximum scores on the scales.

Examination of teaching practice showed that teachers implemented their science lessons to a proficient degree in the areas of instruction and curriculum. The PDFS group, however, fell below expectations in assessment and discourse. NURTURES II leadership should investigate how implementation of assessment and discourse factors might be improved for Cohort 2. Examination of the videos as well as a comparison to their research data could provide insight as to exactly where teachers are falling short.

Family Event observations both fall 2018 and spring 2019 verify that when teachers follow the Event protocols provided by NURTURES II, the Events are successful. Some teachers did not recruit sufficient numbers of facilitators to assist at each of the activity centers resulting in confusion. In addition, attempting to combine the Event with other school functions proved unmanageable. It is recommended that NURTURES II ask teachers from Cohort 1 to share their experiences with Cohort 2 to better prepare them for hosting the Events. In some cases, it appeared that teachers did not review the event or prepare in advance thus detracting from the family experience. In addition, NURTURES II leadership could spend extra time with the PDFS group to review the importance of preparation and advanced set up prior to the commencement of the fall Events 2019.

It was difficult to make generalizations about the Family Pack experience as the number of families providing feedback was small. Considering that there were 15 teachers and estimating conservatively that the classrooms each enrolled 15 students, the 10 responses represented only 4% of the potential sample (15 x 15 = 225). Responses received verified that those families engaged in the activities as designed (fidelity of treatment implementation). This indicates that the instructions and materials were sufficient for families to follow and use. To gather a larger sample for Cohort 2, leadership might discuss with teachers the low response rate and ask for suggestions.

Students within each of the groups scored equivalent on the GMRT. However, there was a discrepancy in balance between grade levels as noted earlier. Analysis by year does not seem prudent due to the imbalance; therefore, it is recommended that assertions as to effects on student science ability be held until data from Cohort 2 students has been collected and analyzed.

6) References

Akbasli, S., Sahin, M., & Yaykiran, Z. (2016). The effect of reading comprehension on the performance in science and mathematics. *Journal of Education and Practice, 7*(16), 108-121.

Maier, M. F., Greenfield, D. B., & Bulotsky-Shearer, R. J. (2013). Development and validation of a preschool teachers' attitudes and beliefs toward science teaching questionnaire. *Early Childhood Research Quarterly, 28*(2), 366-378.