ILLINOIS BIRTH THROUGH THREE WAIVER: DEVELOPMENTALLY INFORMED CHILD AND FAMILY INTERVENTIONS

IB3

EXECUTIVE SUMMARY:
INTERIM EVALUATION REPORT
REPORTING PERIOD: 7/1/2013 – 12/31/2015

PREPARED FOR THE ILLINOIS DEPARTMENT OF CHILDREN AND FAMILY SERVICES BY

SCHOOL OF SOCIAL WORK AT THE UNIVERSITY OF NORTH CAROLINA
HELEN BADER SCHOOL OF SOCIAL WELFARE AT THE UNIVERSITY OF WISCONSIN
CHAPIN HALL AT THE UNIVERSITY OF CHICAGO

SUBMITTED TO THE CHILDREN'S BUREAU BY:

THE ILLINOIS DEPARTMENT OF CHILDREN AND FAMILY SERVICES

MARCH 31, 2016

EXECUTIVE SUMMARY

The IB3 Title IV-E Waiver Demonstration focuses on a very vulnerable population of maltreated children: infants, toddlers, and preschoolers aged birth through three years old who have been removed from their parents' custody and placed into the protective custody of child welfare authorities. Advances in neuroscience confirm that the first three years of a child's life are an extremely sensitive period for social and emotional development. Not only does maltreatment have adverse effects on the developing brain, but the deprivation of consistent and responsive parenting can lead to changes that result in potentially long-lasting deficits in cognitive and behavioral functioning. If not appropriately addressed, these adverse childhood experiences can increase children's vulnerability to stress and predispose them to social, emotional, and health problems throughout their adult life.

The purpose of the IB3 waiver demonstration project is to support the adaptation of evidence-based, trauma-informed parenting programs and test their effectiveness in addressing the adverse effects of maltreatment and in promoting secure attachment relationships. The selected interventions of Child-Parent Psychotherapy (CPP) and Nurturing Parents Program (NPP) are intended to create a developmentally appropriate, responsive parenting environment that can facilitate timely family reunification or expedite alternative permanency arrangements when reunification cannot be attained. By offering families developmentally appropriate parent training and support, including child-parent therapeutic interventions when indicated, it is anticipated that children assigned to the intervention group will experience reduced trauma symptoms, increased permanence, and improved child well-being compared to children who receive services as usual.

This interim evaluation report presents preliminary evaluation findings from the first ten (10) quarters of the Illinois Birth through Three (IB3) Title IV-E waiver demonstration. The waiver supports the adaptation of evidence-based, trauma-informed parenting programs to the care and permanency planning for infants, toddlers, and preschoolers who are taken into the legal custody of the Illinois Department of Children and Family Services (IDCFS).

THE FRAMEWORK

The implementation and evaluation of the IB3 waiver demonstration is patterned after *A Framework to Design, Test, Spread, and Sustain Effective Practice in Child Welfare* ("Framework", Framework Group, 2014), which the U.S. Children's Bureau disseminated to support the implementation and evaluation of federally funded programs and innovations. The Framework conceptualizes the implementation and evaluation process as cycling through five phases of "increasingly generalizable studies" (Shadish, Cook & Campbell, 2002) prior to scaling-up the program for widespread dissemination.

During the Identify & Explore phase of waiver implementation and development, a group of Illinois officials, voluntary agency administrators, and university partners identified the exceptionally long lengths of stay in Illinois foster care as a special area of concern, particularly for children aged birth though three years old. After conducting a literature review, two evidence-based programs were selected as potentially well suited to address the developmental needs of children in foster care and to enhance the parenting competencies of the families with whom they are intended

to be reunified.

Child Parent Psychotherapy (CCP) is a dyadic (caregiver and child) intervention for infants, toddler, and preschoolers who have experienced at least one traumatic event such as the sudden or traumatic death of someone close, a serious accident, sexual abuse, or exposure to domestic violence, among others. The primary goal of CPP is to support and strengthen the relationship between a child and his or her parent (caregiver) as the vehicle for imparting to the child a positive feeling of safety, permanence, and well-being. The Nurturing Parenting Program (NPP) is delivered as a group intervention for 7 to 8 caregivers and aims to enhance their parenting competencies with respect to the following: setting age-appropriate expectations, cultivating empathy for children's needs, using alternatives to physical discipline, establishing appropriate role responsibilities, and encouraging children's free expression of thoughts and opinions. It is aimed at modifying maladaptive beliefs that lead to abusive and neglectful parenting behaviors so that children can be safely and permanently reunified with their families.

IMPLEMENTATION INTEGRITY (PROCESS STUDY)

The IB3 demonstration is predicated on the formula that programmatic success is a product of intervention integrity (i.e. the fidelity with which the program is implemented) and intervention validity (i.e., the efficacy of the program in achieving its intended objectives). Shortfalls in achieving desired results may result from deficiencies in implementation integrity or lack of intervention validity, or both.

Implementation integrity is the focus of the process study. Before proceeding to full implementation and summative evaluation, it was important to ascertain whether formative implementation was achieving acceptable levels of coverage, exposure, participant responsiveness, program differentiation and adherence.

Program coverage refers to the extent to which study subjects and conditions are representative of the intended target population and service delivery settings. During the usability testing phase of the process study, the DCFS Early Childhood Project refined its enhanced developmental screening process for referring children to the CPP and NPP based on levels of risk and need. In addition to the Denver II Developmental Screening tool, the Ages and Stages Questionnaire, and the Ages and Stages: Social and Emotional instrument, the enhanced process incorporated the following additional tools: 1) Devereux Early Childhood Assessment for Infants and Toddlers [DECA]; 2) Infant Toddler Symptom Checklist [ITSC]; and 3) The Parenting Stress Inventory. The assessment tools were used to make two crucial determinations: 1) the category of risk assigned to the child (high, moderate or low); and 2) services, if needed, to address trauma and attachment issues.

The results of the process study of the enhanced developmental screening process showed exceptionally good coverage of the intended target population of children. Approximately 87% of all children under age 4 years old who entered foster care in Cook County during fiscal years 2014 and 2015 were assigned to the demonstration. Of the 964 assigned children, almost 90% received enhanced developmental screenings. Approximately 56% were determined to be at high risk, 31% at moderate risk, and the remainder at low risk or deferred for assessment. All of the children allocated (rotationally assigned) to comparison agencies received services as usual. High risk children

allocated to intervention agencies were slated to receive CPP as the primary treatment modality. Because CPP can be used with any caregiver, CPP therapists could exercise discretion in gearing the intervention towards foster caregivers, biological parents or both (either simultaneously or sequentially). NPP was the primary treatment modality for both parents and caregivers of children assessed at moderate risk.

During usability testing, it was determined that a far greater proportion of children than anticipated was being assessed at high or moderate risk. The original projection of provider capacity assumed that approximately 45% of screened children would be determined to be at moderate to high risk. In actuality, 87% of screened children received this designation. It was initially suspected that the ratings of children assigned to intervention agencies may have unwittily been inflated in order to qualify more children for intervention services; but thorough review of the process demonstrated little basis for biased ratings. When all assessments were tabulated, there were statistically equivalent proportions of high, moderate, and low risk cases distributed among the intervention and comparison agency clusters ($\chi^2 = 1.25$, $p < .741$).

Exposure refers to the extent of client participation and level of service dosages received by participants. Even though risk levels were evenly balanced between the two groups, the larger than anticipated proportion of high risk cases created a resource crunch that threatened to reduce service exposure below acceptable levels. During the 1st two quarters of implementation, nearly 50% of high risk cases were referred to CPP as their primary treatment modality. By the 3rd quarter, however, CPP wait lists had to be created. The caregivers of newly placed high risk children were instead referred to NPP programs until CPP vacancies became available. This pattern of sequential referrals accounted for approximately 25% of cases assigned during the 3rd and 4th quarters of 2014. By 2015, however, all new and sequential referrals to CPP essentially stopped due to the lengthening waiting list.

Putting families on CPP wait lists did not mean that they stopped receiving IB3 interventions altogether. The limitation was addressed by increased NPP referrals. Two-thirds of cases that were enrolled in the demonstration during 2015 were referred to either the parent (PV) or caregiver (CV) version of NPP, and in one-fourth of cases to both. Completion rates were respectable with over one-half of birth parents completing all 16 weeks of NPP-PV and one-half of CPP participants completing or still attending the 52-week CPP program. A major short-fall that raised concern about participant responsiveness was the low completion rate for the NPP-CV. Only 22% of NPP-CV referrals completed the program. Interviews with a convenience sample of foster caregivers did identify logistical barriers, such as child care, transportation, and time of day as hindering their full participation. Others questioned the need for another training when they had already received many other trainings. Nonetheless, almost all foster parents, who completed NPP, said they very much enjoyed their experience and learned a great deal.

Program differentiation refers to the extent to which the subjects in each intervention condition received only the assigned treatment. The proportion of comparison-group cross-overs from services-as-usual to enrollment in one or more of the IB3 intervention services was negligible. Ninety-nine percent of both high and moderate risk cases in the comparison group had no record of referral to either the CPP or NPP programs. Of course, it can't be ruled out definitively that the usual services available to comparison cases, such as parenting coaches and regular parent training, offered content similar to NPP and CPP. Based on observations of the programs available to

families, however, there is ample reason to believe that the content of IB3 interventions is substantively different from the content offered by services-as-usual. There is little reason to suspect that the potential impact of the IB3 interventions on permanency and well-being outcomes is being diluted by the lack of program differentiation between services available to intervention subjects and those usual services available to comparison subjects.

Challenges related to the assessment of program adherence made it difficult to determine the extent to which specified program components were delivered as prescribed by training, certification, or program manuals. Fidelity data for CPP adherence is lacking because the program is less standardized for use with very young children. Essential components are documented in the supervision checklist, but it was not feasible to document specific elements of each interaction without extensive video-recording and coding by observers proficient in CPP. Ultimately, it was decided to forego collecting CPP fidelity data and instead rely on the credentialing connected to the recruitment of CPP therapists. During the early phases of implementation, CPP therapists were recruited from existing capacity within Cook County, which had been created through an eighteen-month Learning Collaborative that had trained professionals in the CPP model. Fidelity to the model, however, was potentially threatened by the turnover of 55% of CPP clinicians across provider agencies. The staff has largely been replaced by qualified personnel. The IB3 demonstration requires that all CPP therapists possess a master's degree in a mental health discipline and log the necessary time in conducting CPP with supervision for the requisite length of time dictated by the model.

In contrast, NPP is a manualized approach to parent education that requires strict adherence to the curriculum for each lesson. Rather than collect fidelity data, the NPP purveyor recommended that regular posttests be administered at the end of each lesson in order to determine that the targeted competency had been mastered by participants. In line with this approach to fidelity assessment, the outcomes study tracks changes in parenting competences from baseline to program completion using version 2 of the Adult and Adolescent Parenting Inventory (AAPI-2). The inventory assesses parenting and child rearing competencies in five areas that are predictive of nurturing and responsive caregiving. The results of paired sample t-tests of changes in competence based on 171 completed pre and posttests showed moderate to strong improvements in four of the five areas. These findings provide credible indirect evidence of the fidelity of NPP implementation.

Intervention Validity (Outcomes Study)

Intervention validity is the focus of the proximal outcomes study. It also cycled through successive phases of evidence-building beginning with statistical conclusion validity and continuing with building confidence in the internal, external, and construct validity of the findings. Intervention validity is enhanced by the IB3 sampling plan that assigned 87% of all children aged birth through three years old who entered foster care in Cook County during fiscal years 2014 and 2015. Intervention validity is also supported by utilizing a two-stage unbiased allocation procedure that assigned children and families to the intervention and comparison groups. The first-tier allocation involved the randomization of three DCFS offices and 19 voluntary child welfare agencies to comparison and intervention clusters. The second-tier involved the rotational assignment of child cases to different agencies in accordance with rules that the IDCFS developed for performance contracting.

Once a child is assigned to the intervention cluster of case management agencies and DCFS offices, the child retains the assignment for the duration of the waiver demonstration even if the child changes case management agencies or offices. The same holds true for children assigned to the comparison cluster. Therefore all analyses compare outcomes based on children's original assignment group. As of September 2015, 85% of child cases were managed by the same child welfare agency or DCFS office to which they were originally assigned. The proportions are statistically equivalent for both intervention (84%) and comparison (87%) clusters (chi square = 1.81; 1 *df*; $p < .178$).

The study assessed the comparability of agency clusters using the Local Agency Director Questionnaire (LADQ) that was administered during the months of February and March of 2013. All but one agency (from the comparison cluster) completed the pencil-and-paper form. Because there were 138 comparisons that were tested between intervention and comparison groups, it may be expected that some significant differences would arise purely by chance. Based on these data, it can be assumed that the comparison cluster provides a suitable "counterfactual" for estimating the effects of IB3 programs on the outcomes observed for children and family served by agencies in the intervention cluster.

Statistical conclusion validity refers to the extent to which there is a statistically significant association between the interventions and the desired outcomes. In light of the exceptionally long lengths of stay of foster children in Cook County (less than 10% have exited state custody since the start of the demonstration), only three types of proximal permanency outcomes could be reliably compared: return home rates regardless of whether state custody was relinquished (i.e., includes trial home visits); reunification rates with case closure; and permanency rates which encompass reunification, adoption, and legal guardianships. Only the return home rate showed a marginally significant association ($p < .10$) with assignment to the intervention cluster of agencies in the expected direction of improved permanence. The other two proximal outcomes were also in the expected direction but the observed difference was not large enough to rule out chance error.

After statistical conclusion validity is established, the next phase in building credible evidence of intervention validity is to rule out threats to the internal validity of the findings. Internal validity refers to the extent to which there is confidence that the observed statistical association results from a causal relationship between the interventions and the outcomes. The outcomes study utilized a two-tier unbiased allocation procedure for assigning cases to intervention and comparison conditions: 1) randomization of DCFS offices and voluntary agencies to treatment clusters; and 2) rotational assignment of child cases to intervention and comparison agencies. With unbiased allocations, systematic differences between the two groups should occur only by chance and, if the number of cases is sufficiently large, with a very low probability. If the assumption of statistical equivalence between groups holds, the observed differences in outcomes between intervention and comparison cases can be confidently attributed to the casual effect of the assignment rather than to any preexisting differences at baseline (selection bias), changes that would have occurred in any event (maturation bias), or happenings that unfold over time (history bias).

The test of the statistical equivalence of the intervention and comparison cases yielded mixed results. On the one hand, the two groups were well balanced on those conditions that brought families to the attention of IDCFS, including type of maltreatment, reasons for case opening, and risk level at developmental screenings. On the other hand, there were systematic differences related

to the local ecologies of the cases managed by comparison and intervention agencies, such as children of Hispanic origins, initial placement into kinship foster care, and case management by DCFS office. In addition, changes to the allocation formula for rotational assignment midway through the study resulted in more recent cases being assigned to comparison agencies. These imbalances necessitated the inclusion of indicators of these differences as statistical controls into data analyses in order to minimize the threats of selection and history bias to the internal validity of the findings.

Data analysis also addressed the issue of external validity. This concept refers to the extent which a causal relationship is generalizable across variations in different populations and settings. Because of the systematic differences between groups with respect to Hispanic origins, kinship foster care, and DCFS management, interactions between assignment group and these variables were tested for significance. Neither the main nor interaction effect was significant for Hispanic origins; but there were significant interactions for the other two control variables. Predictive margins of assignment group interactions showed that a statistically significant intervention effect was confined to children initially placed in non-kinship family settings under the case management of voluntary agencies. Even though children initially placed with kin demonstrated higher return home rates than children placed with non-kin, the advantage was the same whether they were assigned to intervention or comparison agencies. Further, higher return home rates were registered among children assigned to DCFS offices in the comparison group (Cook County South region) than offices in the intervention group (Cook County Central and North regions). This difference resulted in a sign reversal of the intervention effect for children under DCFS case management.

The final phase of data analysis was restricted to children initially placed with kin under voluntary agency management because this was the only segment with a significant intervention effect. The focus was on construct validity, which refers to the extent to which the causal relationship corresponds to its higher-order theory of change as specified in the IB3 Logic Model. Predictive margins of exposure of the restricted population to the IB3 interventions showed an increased return home rate as expected for cases that completed or were still attending the NPP and CPP programs compared to drop-outs, no shows, or children assigned to comparison agencies. The highest return home rate was registered among non-referred cases, which make sense considering the non-referrals included cases that reunified quickly before a referral could be made as well as cases rated as low risk.

The total costs of services to the IB3 intervention groups from July 1, 2013 to December 31, 2015 amounted to $3,637,952 in foster care subsidies and $7,845,320 in administrative costs for a total cost of $11,483,272. On a per child basis, an average of $23,152 was spent on the care and case management of intervention cases. If these same children had been assigned to the comparison, it is estimated that each child would have averaged $22,231. The average difference of $921 per child reflects the additional costs of providing the IB3 interventions and associated case management expenditures. The intervention costs were lower than projected because of contractual challenges concerning CPP. Only 29% of fiscal 2015 obligated funds were invoiced by the five CPP providers. Further these claims covered only 45% of their actual costs.

As the demonstration progresses, it is anticipated that the higher intervention costs will eventually be recovered and potentially yield savings as the foster care and administrative costs for intervention cases decline more rapidly than average spending on comparison cases. The results

from hazards regression analysis of duration-specific return home and reunification rates presage the potential for a strengthening of the intervention effect as the demonstration completes its third year of implementation. A graph of smoothed hazards rates showed flat levels after two years in foster care for cases assigned to comparison agencies but sharply rising rates for children assigned to intervention agencies. If this pattern continues into year three of the demonstration, it is very likely that the intervention effect on reunification rates will strengthen during this critical period of judicial oversight when decisions are made about alternative permanency plans for the children.

WELL BEING OUTCOMES

The assessment tools support critical decisions for all children under the age of 4 in Cook County. The tools are used to make two crucial determinations: 1) the category of risk assigned to the child (high, moderate or low); 2) services, if any, needed to address the trauma and attachment issues. Both the ITSC and the DECA serve to assess the child's trauma symptoms. The CANS is used to assess trauma experiences and history. The parent and caregiver's level of stress is determined by the PSI.

ASSESSMENT FINDINGS

The IB3 enhanced screening assessment process administers a variety of age-specific standardized instruments to arrive at a risk determination for each child. The Devereux Early Childhood Assessment (DECA) for infants and toddlers provides a standardized, norm-referenced assessment of "within-child" protective factors for children. There are separate instruments for infants, ages 1 month up to 18 months, and for toddlers, ages 18 months up to 36 months. There is also a separate preschooler version for children ages 3 through 5 years old.

Two-thirds of the children who are enrolled in the IB3 waiver demonstration are assessed as infants. For this age group, the DECA assesses the infant's ability to take *initiative*, that is to use independent thought and actions to try new things, imitate the actions of others, and look to a familiar adult when exploring surroundings. Another set of items assesses the child's *attachment* to significant adults; that is showing affection for a familiar adult, smiling, and seeking comfort from familiar adults. A Total Protective Factors scale is constructed from these two constructs, which provides an overall indication of the infant's protective factors.

For the DECA, a T score describes the level of a child's rating in comparison to all other children. It is a standard score that has a mean of 50 and a standard deviation of 10. Because T scores have the same meaning throughout the normal distribution, T scores are used instead of raw scores when comparing scores earned on various rating scales.

T scores can be classified as Area of Need (40 or below), Typical (41-59), or Strength (60 and above). For the U.S. standardization sample, typically 16% of children had T scores in the Area of Need, 68% had scores in the typical range, and 16% had scores in the Strength range. Table 1 displays these T score ranges on attachment for infants with known risk determinations. Two observations are in order: 1) there are no differences between assignment groups, which indicates the two groups are well balanced on this within-child protective factor; and 2) both groups profile similarly to the U.S. standardization sample, in which two-thirds fall in the typical range.

Table 1. T Score Ranges on Attachment for Infants by Assignment Group				
T Score Range		**Group**		**Total**
		Comparison	Intervention	
Area of Need (40 or below)	Count	25	24	49
	Col.%	12%	12%	12%
Typical (41-59)	Count	135	140	275
	Col.%	67%	68%	33%
Strength (60 and above)	Count	41	43	84
	Col.%	20%	21%	21%
TOTAL	Count	201	201	408
		100%	100%	100%
				Chi-square = .071, p = .965

The ranges for toddlers and preschoolers are in marked contrast to the ranges for infants. More than twice as many children fall into the Area of Need range than is the norm for the standardization sample. Again there are no differences between the intervention and comparison group. Children who enter foster care at 19 months or older are significantly more disadvantaged in their attachment relationships as compared to infants who are removed at an earlier age. The same general patterns hold for the initiative construct and total protective factors.

Table 2. T Score Ranges on Attachment for Toddlers and Preschoolers by Assignment Group				
T Score Range		**Group**		**Total**
		Comparison	Intervention	
Area of Need (40 or below)	Count	34	44	78
	Col.%	42%	37%	39%
Typical (41-59)	Count	44	69	113
	Col.%	54%	58%	57%
Strength (60 and above)	Count	3	6	9
	Col.%	4%	5%	5%
TOTAL	Count	81	119	200
		100%	100%	100%
				Chi-square = .615, p = .735

Construct Validity: Is Exposure to the IB3 Interventions Improving Parenting Competencies?

The IB3 theory of change is predicated on the assumption that improvements in parenting competencies will enhance early brain development and provide the responsive parenting environment that will allow children to be returned to parental custody. One of the mechanisms that is critical to responsive parenting is empathy with the normal developmental needs of children. This

can be particularly challenging when caring for pre-verbal children who express their needs by crying or signaling through non-verbal cues. Fortunately as a species humans are innately equipped to respond appropriately, but sometimes signals get crossed. Personal trauma experiences, insecure attachments relationships in one's own childhood, and antiquated child-rearing advice that is no longer valid can interfere with the proper protection, care, and discipline of children. Both CPP and NPP are evidence-based interventions that attempt to improve caregivers' abilities to interpret, value, and respond sensitively to the normal developmental needs of children. The IB3 demonstration relies on the Adolescent and Adult Parenting Inventory (AAPI-2) to measure the degree to which such goals are being achieved.

Table 3. AAPI-2 Subscales

Subscale	Construct	Description
A	Expectations of Children	High risk involves inappropriate expectations that exceed the normal developmental capabilities of children. Tends to be demanding and controlling.
B	Empathy toward Children's Needs	High risk involves low levels of empathy in which the caregiver does not understand or value children's normal developmental needs. Children must act right and not be spoiled.
C	Use of Corporal Punishment	High risk sanctions hitting, spanking, and slapping of children as appropriate and required. A strong disciplinarian who lacks understanding of alternatives to corporal punishment is considered to be high risk.
D	Parent-Child Role Responsibilities	High risk tends to use children to meet self-needs. They expect children to make life better by providing them love, assurance, and comfort.
E	Children's Power and Independence	High risk tends to view children with power as threatening. They tend to view independent thinking as disrespectful.

The AAPI-2 is the revised and re-normed version of the original AAPI first developed in 1979. The inventory is designed to assess the parenting and child rearing attitudes of adult and adolescent parenting and pre-parenting populations. The AAPI-2, like its predecessor, is a validated and reliable inventory that is predictive of abusive parenting. Responses to the inventory discriminate between the parenting behaviors of known abusive parents and the behaviors of non-abusive parents. The AAPI-2 is used by NPP providers to assess changes in the parenting and child rearing attitudes of programs participants. Responses to the AAPI provide an index of risk assessment in five specific parenting and child rearing behaviors scored from 1 (highest risk) to 10 (lowest risk) as described in Table 3 above.

At the end of the reporting period, 201 parents and 74 substitute caregivers had completed baseline assessments of their parenting and child rearing attitudes. Of the parents, 70% were mothers and 30% were dads. Of the caregivers, 75% were females and 25% were males. The distribution of subscale responses for both parents and caregivers indicate a generally higher level of risk compared to general population norms. Approximately 16% of the general population scores in the high risk range compared to 25% of parents and caregivers in the IB3 sample. There is one

exception to the rule, however. Fewer parents and caregivers in the IB3 sample score in the high risk range (less than 10%) compared to general population norms on attitudes toward corporal punishment. Very few parents and caregivers endorse hitting, spanking, and slapping of children as appropriate ways of disciplining children.

Table 4. AAPI-2 Pair Sample Statistics

Group	Construct	Test	Mean	N	SD	Correlation	Effect Size
Birth Parents	Expectations	Posttest	5.58	118	1.914	0.400	0.216
		Pretest	5.12	118	1.953		
	Empathy	Posttest	6.53	118	2.434	0.554	0.796
		Pretest	4.80	118	2.186		
	Punishment	Posttest	7.11	118	1.876	0.491	0.559
		Pretest	6.06	118	1.850		
	Roles	Posttest	5.99	118	2.479	0.608	0.566
		Pretest	4.81	118	2.222		
	Power	Posttest	6.35	118	2.197	0.372	0.501
		Pretest	5.11	118	2.210		
Foster Parents	Expectations	Posttest	5.42	53	2.080	0.495	0.385
		Pretest	4.62	53	2.021		
	Empathy	Posttest	6.53	53	2.127	0.748	1.232
		Pretest	4.60	53	2.273		
	Punishment	Posttest	6.60	53	1.843	0.582	0.331
		Pretest	6.09	53	1.522		
	Roles	Posttest	6.09	53	2.177	0.574	0.592
		Pretest	4.89	53	2.242		
	Power	Posttest	6.25	53	2.425	0.639	0.406
		Pretest	5.42	53	2.389		

There are no significant differences between the attitudes expressed by parents and by caregivers. The sample sizes are too small to conduct separate comparisons between kinship caregivers and foster parents. Table 4 shows the posttest and pretest score for the 171 parents and caregivers who completed the NPP program. The posttest subscales indicate a substantial improvement in parenting competencies. An effect size indicates the standardized difference between posttest and pretest means. Effects sizes greater than .5 are considered medium and those greater than .8 are considered large. Changes in level of empathy fell into this latter category for both birth parents and foster parents.

According to the *AAPI OnLine Development Handbook* (Bavolek & Keene, 2010), empathy is the ability to be aware of another person's needs and feelings and to place the needs of another as a priority:

> "Parents lacking sufficient levels of empathy find children's needs and wants as irritating and overwhelming. Everyday normal demands are perceived as unrealistic, resulting in increased levels of stress. The needs of the child come into direct conflict with the needs of the parent, which are often similar in magnitude. Lacking an empathic home life, children

often fail to develop a solid moral code of conduct. Right and wrong, cooperation, and kindness are not important because they are not recognized as important values. Others are devalued as "self" takes center stage. The impact of one's negative actions on another is muted as the ability to care about the needs or feelings of another is not important" (Bavolek & Greene, 2010, pp.3-4).

At pretest, one-third of parents and caregivers expressed attitudes and beliefs suggesting insufficient levels of empathy, which put them at high risk for neglectful parenting. After completing the NPP trainings, less than 10% answered with non-empathetic responses. Even though effect sizes were less pronounced for the other constructs of expectations, punishment, roles and power, the IB3 sample was above national population norms at post-test with respect to appropriate parenting beliefs and attitudes.

Even though there is a slight association between returning home and scoring higher on most parenting competences at post-test, the relationship is weakest for parents with higher empathic responses and strongest for parents with higher score on parent-child role responsibilities and children's power and independence. Still the probabilities of returning home are still quite low. Of the 119 children whose caregivers completed NPP trainings, only 1 out of 10 children returned home to parents who scored low risk at post-test for each of competency.

SUMMARY, LESSONS LEARNED, AND NEXT STEPS

This interim evaluation report presents preliminary evaluation findings from the first ten (10) quarters of the Illinois Birth through Three (IB3) Title IV-E waiver demonstration. The results of the Process Study indicated that the IB3 Demonstration achieved adequate levels of implementation integrity with respect to program coverage, differentiation, exposure, adherence, and participant responsiveness. Rotational assignment resulted in a fairly balanced allocation of the assigned cases to intervention agencies (51%) and comparison agencies (49%). Even though higher than expected ratings of high risk children resulted in a waiting list for intensive dyadic (parent-child) interventions, referrals to small-group, nurturing parenting programs accommodated much of the need for services.

Data analysis included appropriate statistical controls for imbalances in post-removal case characteristics of children assigned to the comparison and intervention agencies. Linear, logistic, and hazards regression models yielded a similar set of findings. The intervention effect was confined to children initially placed in non-kinship homes under the case management of voluntary agencies. Predictive margins of program exposure within this restricted sample showed an increased return home rate for cases that completed or were still attending IB3 programs compared to drop-outs, no shows, or children assigned to comparison agencies.

The total costs of services to the IB3 intervention group from July 1, 2013 to December 31, 2015 amounted to $3,637,952 in foster care subsidies and $7,845,320 in administrative costs for a total cost of $11,483,272. An average of $729 additional dollars was spent per child in the intervention group compared to the amount that would have been spent if they had received services as usual. As the demonstration progresses, it is anticipated that the higher intervention costs will eventually be recovered and potentially yield savings as the foster care and administrative costs for intervention cases decline more rapidly than average spending on comparison cases. The results

from hazards regression analysis of duration-specific return home and reunification rates suggest the potential for a strengthening of the intervention effect as the demonstration completes its third year of implementation. A graph of smoothed hazards rates shows that after two years in foster care the likelihood of reunification levels off for cases assigned to comparison agencies whereas they sharply rise for children assigned to intervention agencies. If this pattern continues into year three of the demonstration, it is very likely that the intervention effect on reunification rates will strengthen during this critical period of judicial oversight when decisions are made about alternative permanency plans for the children.

The major lesson learned from the first 10 quarters of IB implementation and evaluation is that tremendous effort is required to shift the standard expectations of the entire system from a traditional belief set about the resilience of infants and toddlers to a trauma-informed understanding of the adverse effects of maltreatment on early brain development and the need for prompt permanency decisions that promote secure attachments relationships. It is becoming abundantly evident that child well-being can best be assured within the context of stable, permanent parent-child relationships. This knowledge should lead to a reorientation in the delivery of child-parent interventions so that reunification is not delayed until services are completed. Instead, when sufficient progress is evident and in meaningful collaboration with clinical providers, interventions that support the parent-child relationship may continue for an appropriate period after the child is restored to parental custody. These are lessons that should be embraced by practitioners and decision makers, including caseworkers, foster parents, administrators and judges. These are also lessons that must be rigorously evaluated in order to ensure that the safety and well-being of children are truly enhanced by the reorientation in service delivery.

The remaining 10 quarters of the evaluation will continue to track the progress of the children who were assigned to the IB3 intervention group. In addition to monitoring the implementation integrity of the demonstration, a major investment will be made in surveying the well-being of children in both comparison and intervention groups. In this way, we hope to capitalize on the rigor of the evaluation design and strengthen the evidence-base for what works to promote the safety, family permanence, and well-being of children during their most sensitive years of development.