October 2010

better beginnings

EVERY CHILD DESERVES OUR BEST

UAMS

Evaluating Arkansas' Path to Better Child Outcomes

Prepared for DCCECE | by Partners for Inclusive Communities, UAMS

EVALUATING ARKANSAS' PATH TO BETTER CHILD OUTCOMES

UAMS
PARTNERS FOR INCLUSIVE COMMUNITIES
UNIVERSITY OF ARKANSAS FOR MEDICAL SCIENCES

Lorraine McKelvey
Melanie Chapin-Critz
Betsy Johnson
Patti Bokony
Nikki Conners-Burrow
Leanne Whiteside-Mansell

Crosswalks used for this report available in the online Technical Appendices at www.arbetterbeginnings.com

Appendix 1: PAS & CARF
Appendix 2: PAS & COA
Appendix 3: PAS & Head Start
Appendix 4: SF, ECERS-R, & PAS
Appendix 5: SACERS, Younger Youth PQA, Youth PQA Form B, & Youth PQA
Appendix 6: ECERS-R & CLASS

Prepared for
Arkansas Department of Human Services
Division of Child Care and Early Childhood Education
700 Main Street
Little Rock, AR 72201

Prepared with funding from The American Recovery and Investment Act of 2009

CONTENTS

TABLES

1 EXECUTIVE SUMMARY

Child care is more than a basic custodial duty. It is an important educational function with the potential to improve long-term individual and social health. Research in brain function and early learning shows that more rapid development takes place from birth to age 5 than at any other stage of life. Young children are more vulnerable to risks in their caregiving environments. Fortunately, they are also more resilient and responsive to intervention.[1]

★★★★★★★★★★

Similar to other state rating systems, Better Beginnings draws on elements common to successful preschool models.

★★★★★★★★★★

Evaluators of early model preschool programs have persistently followed participants into adulthood. They offer alluring evidence that high quality child care confers long-term cognitive and social-emotional benefits. Model programs share common elements: well-educated, well-paid staff; good teacher-student ratios (usually 1:3 for infants/toddlers and up to 1:6 for preschoolers); extensive education, training, and involvement of parents; and curricula with a strong theoretical basis.[2]

Some states have amplified or enhanced minimum licensing standards, but most do not come close to guaranteeing that all children experience these elements. The reauthorization of federal funding for Head Start and the establishment of state-funded Arkansas Better Chance for School Success (ABC) programs represent a societal shift towards acknowledgement of the importance of quality early care and education. However, legislative bodies, citizens, and parents, have not yet been willing to allot the level of funding necessary for the widespread implementation of comprehensive, intensive programs. Thus, a number of states are exploring less costly alternatives to enhance the quality of care and education for all children. Quality Rating Systems (QRS) or Quality Rating and Improvement Systems (QRIS) are such alternatives.

"A QRIS is a systemic approach to assess, improve, and communicate the level of quality in early and school-age care and education programs. Similar to rating systems for restaurants and hotels, QRIS award quality ratings to early and school-age care and education programs that meet a set of defined program standards. By participating in their State's QRIS, early and school-age care providers embark on a path of continuous quality improvement. Even providers that have met the standards of the lowest QRIS levels have achieved a level of quality that is beyond the minimum requirements to operate." (National Child Care Information and Technical Assistance Center)[3]

Similar to other states, Arkansas' Better Beginnings QRIS draws on elements common to the successful model preschool programs when deemed feasible for local providers. It includes standards for the education and training of staff, parent involvement, and use of a curriculum.

---

[1] Shonkoff, J. P., D. Phillips, et al. (2000). From neurons to neighborhoods: the science of early childhood development. Washington, D.C., National Academy Press.
[2] Yoshikawa, H. (1995). "Long-term effects of early childhood programs on social outcomes and delinquency." The Future of Children 5(3): 51-75.
[3] http://nccic.acf.hhs.gov/qrisresourceguide/index.cfm?do=qrisabout#1. Accessed 9/7/2010.

To aid in understanding possible effects on children's development, we classify each Better Beginnings standard according to the type of measure: **structural** (regulatable components of child care that form the foundation of good quality care), **process** (interactions that directly involve the child, such as presentation of activities and materials to children, behavior management, and the teacher's responsiveness or sensitivity to a child's needs), or **global** (both structural and process features of care such as the Environmental Rating Scales). In many studies, process measures are most capable of predicting child outcomes. Like many state-wide systems, Better Beginnings primarily contains structural measures.

★★★★★★★★★★

Look for the type of measure.
Process measures
predict outcomes best.

★★★★★★★★★★

1.1 COMPONENTS OF BETTER BEGINNINGS

There are five components of the Better Beginnings system: 1) Administration, 2) Qualifications and Professional Development, 3) Learning Environment, 4) Environmental Assessment, and 5) Child Health and Development. We review general findings and recommendations for each of the sections separately.

1.1.1 ADMINISTRATION

To capture a multi-dimensional picture of a program's quality, Better Beginnings incorporates two standardized instruments to measure leadership and management functions in Early Childhood Education (ECE) programs: *The Program Administration Scale* (PAS) for center-based programs and *The Business Administration Scale* (BAS) for family day care.[4] These are the first published instruments that focus solely on the administrative processes of early childcare programs. While these processes are intended to enhance the experiences of children, they are established and staged outside of the classroom. Thus, both scales are categorized as *structural* measures.

The PAS and the BAS are reliable, valid instruments that approach the quality of early childhood programs from a different angle than other scales used in the system. The PAS and the BAS will highlight strengths and weaknesses of business functions and steer administrators toward positive changes. The rationale behind these administrative scales is strong, but the instruments are relatively new. To date, PAS and BAS scores have not been analyzed alongside child outcome variables in general research or in evaluations of state quality rating systems. Also, they have not been validated for use in school-age programs. In centers, there is limited evidence that administrative support moderates teacher-child interactions. In family child care, regulatory status, which is linked to provider adherence to good business practices, showed a relationship to time spent with peers and in level of play but not to other cognitive and social outcomes.

[4] Talan, T. and P. Bloom (2004). Program Administration Scale. New York, Teachers College Press.
Talan, T. and P. Bloom (2009). Business Administration Scale for Family Child Care. New York, Teachers College Press.

Strong leadership and well-informed administrative practices contribute to the global quality of a program, which supports child development.

In the interest of encouraging wider Arkansas center participation, Better Beginnings does not require programs to be assessed on PAS items 22-25 that address administrator and teacher qualifications. Also, PAS items 5 and 6 rating staff benefits, staffing patterns, and scheduling will be assessed but not counted in the program's overall score. However, omitted PAS items may impact the measure's usefulness for Better Beginnings.

Using data collected as part of the Evaluation of the Arkansas Early Childhood Professional Development System (AECPDS), the evaluation team compared the original scoring with the Better Beginnings scoring of the PAS. We found the original scoring of the PAS and the Better Beginnings scoring of the PAS (BB PAS) to be similarly related to the ERS and to Arnett Caregiver Interaction (CIS) subscale scores. In all cases, correlations between the BB PAS, ERS, and CIS scales were weaker than with the original scoring of the PAS. The original PAS scoring was significantly related to teacher behaviors that support children's cognitive development and school readiness. The BB PAS scoring was not. These behaviors include engaging the children with open-ended questions and encouraging children in the use of symbolic/literacy materials, numbers and spatial concepts, and problem solving. Better Beginning's exclusion of PAS items may have a negative effect on the measure's validity.

Using the Better Beginnings scoring of the PAS, the evaluation team determined whether cut scores for the system were meaningful. Better Beginnings' Levels 1 and 2 do not require PAS observations while Level 3 requires a minimum score of 4. Analysis of the AECPDS data demonstrates that programs scoring lower than 4 on the scale have teachers who are less sensitive, more detached, and less supportive of socio-emotional development and classrooms with lower overall global environmental quality ratings.

The second component of the Better Beginnings Administration section requires program leaders to learn about and take actions to reduce child abuse and neglect. Administrators will use the Strengthening Families self- assessment tool and strategy developed by the Center for Study of Social Policy.[5] Strengthening Families is a national initiative to equip early childhood programs with knowledge and practices to prevent child abuse and neglect. The initiative's logic model was based on research highlighting five protective factors in families that correlate with greater child protection and on observations of model child care programs. The goal is to implement strategies used by model child care programs to enhance the factors that provide protection for children (parental resilience, social connections, knowledge of parenting and child development, concrete supports in times of need, and children's social and emotional development). The vast majority of items addressed in the self-assessment are focused on organizational policies, parent training, and communication with parents. These things generally occur outside of classroom interactions with children. Therefore, the Strengthening Families self-assessment is best classified as a structural measure.[6]

---

[5] Center for the Study of Social Policy (2008). "Strengthening Families online resources: Guidebook and self-assessment." Retrieved 05/11/2010, from http://www.strengtheningfamilies.net/index.php/online_resources/guide_assess/category/self_assessment/

[6] The section, Children's Social and Emotional Development, does contain a few items that instruct the type and quality of interactions between teachers and children. We ruled out a classification of SF as a global measure because these child interactions items are outnumbered by structural items not only in the grand scheme of the self-assessment, but also within that particular section.

Research related to Strengthening Families suggests that sustained increases in protection for children are likely if 1) parent involvement and supports are comprehensive, intensive, and sustained, and 2) non-maternal care is comprehensive, sustained, and of quality exceeding the highest level of Better Beginnings. We would expect use of the Strengthening Families model to heighten awareness and to improve relational help-giving skills, such as listening and demonstrating respect and empathy for the family, if all staff members, not just administrators as currently designated, receive more intensive training. It is unlikely that a webinar and self-assessment for administrators will increase the amount and quality of participatory helpgiving, which is individualized, includes the family as active participants in achieving goals, and is more tightly linked to change in family functions and behaviors. Empirical evidence suggests that webinar training, self-assessment and adoption of one or even a few of the Strengthening Families strategies is unlikely to produce detectable significant changes in child abuse and neglect. If administrators extend their training to teachers and adopt the Strengthening Families practices comprehensively, improvements in parental understanding of child development and parenting behaviors are likely.

1.1.2 QUALIFICATIONS AND PROFESSIONAL DEVELOPMENT

The ECE field is shifting from clock hours to college hours.

Within the past decade, a movement calling for a minimum of a bachelor's degree in early childhood education (ECE) classrooms has gained considerable traction. Advocates of this shift point to evidence that college programs focused on ECE or child development improve classroom quality, that teacher education is a better predictor of quality than years of experience, and that higher teacher education is related to better child outcomes.[7] This widespread appeal for degreed teachers represents a general policy shift from one that emphasized in-service training and annual clock hours to one that favors pre-service training. Head Start and the National Association for the Education of Young Children (NAEYC) have a timeline to increase the number of teachers with a bachelor's degree: 50% required by Head Start by 2013, and 100% required by NAEYC by 2020.[8] The early childhood debate about the importance of well-educated teachers is less applicable to school-age programs. As we reviewed the research and recommendations from leaders in the field, we found highly qualified staff to be considered a basic and necessary requirement for high-quality programs.[9]

---

[7] Howes, C., Whitebook, M., & Phillips, D. (1992). Teacher characteristics and effective teaching in child care: Findings from the national child care staffing study. Child and Youth Care Forum, 21(6), 399-414. Snider, M. H., & Fu, V. R. (1990). The effects of specialized education and job experience on early childhood teachers' knowledge of developmentally appropriate practice. Early Childhood Research Quarterly, 5(1), 69-78. Zill, N., Resnick, G., Kim, K., McKey, R. H., Clark, C., Pai-Samant, S., et al. (2001). Head Start FACES: Longitudinal findings on program performance. Third progress report. Washington, DC: Research, Demonstration and Evaluation Branch & Head Start Bureau, Administration on Children, Youth and Families, U.S. Department of Health and Human Services.

[8] Time line for implementation accessed on 10/4/10 at http://www.naeyc.org/files/academy/file/Time%20Line%20for%20Meeting%206_A_05.pdf

[9] Bodilly, S. and M. K. Beckett (2005). "Making Out-of-School Time Matter: Evidence for an Action Agenda." Accessed 06/24/2010, from http://www.rand.org/pubs/monographs/2005/RAND_MG242.pdf. Committee on After-School Research and Practice (2005). Moving Towards Success: Framework for After-School Programs. Washington, DC, C. S. Mott Foundation. Miller, B. M. (2005).

Arkansas' current minimum licensing standard for staff (a high school diploma or GED and 10 hours of in-service training) is far-removed from the field's best practices. However, only rewarding providers who can afford to make the substantial leap from simply hiring teachers with high school education to hiring teachers with four-year degrees in all classes would alienate many from the QRIS. Moreover, private providers who must pay more for better educated teachers would pass costs on to consumers, which might force lower-income families to choose informal or lower quality forms of child care.[10] To maximize inclusivity and to encourage provider participation in the quality improvement process, Better Beginnings calibrated its standards to existing levels in Arkansas. The Administrator/Staff Qualifications/Professional Development component of Better Beginnings encourages increased levels of training for teachers and administrators but with more emphasis on clock hours than on formal college hours. Requirements for school-age providers are identical to the center-based standards. Requirements are slightly lower for family child care providers. Primary caregivers in family child care homes must have 30 clock hours at the lowest level and an additional 15 clock hours plus 10 hours of ongoing professional development annually for the highest level.

Minding the gap between the quality of Arkansas programs and research-based quality indicators is a complicated task. Touting the merits of teacher education is one thing; paying for tuition subsidies, enhanced professional development programs, and increased wages for a better trained workforce is another. Yet, our evaluation team concludes that the Better Beginnings scale is tipped substantially more toward status quo than toward the field's best practices. In light of ECE research and state comparisons, the professional development and qualifications standards are likely insufficient. Incentive grants available to programs participating in Better Beginnings to help pay for college credits and credentialing are likely to prove more effective for promoting increased professionalism for teachers and optimal support for child development.

Research suggests that advantages for child outcomes begin at the Intermediate levels of the Arkansas Early Childhood Professional Development System, Traveling Arkansas Professional Pathways (TAPP), not at the Foundation levels emphasized in Better Beginnings. Findings are mixed about whether a particular level of specialized college education enhances teaching practices and child outcomes. However, we find general agreement that ECE college courses are better at preparing teachers to create developmentally appropriate environments and to interact with children in ways that promote their development than workshops or in-service trainings. Short, one-day workshops, such as those required for Better Beginnings, are not likely to be effective. Multi-day workshops lacking a fixed curriculum and offered at a large number of sites without customization for each group of participants are also unlikely to produce results. We recommend that professional development efforts continue to open avenues for college credits. As an intermediate step, proposed higher levels of Better Beginnings could require longer trainings, such as Pre-K Early Literacy Learning in Arkansas and Pre-K Social Emotional Learning for Young Children. These trainings already exist, meet more of the research-based criteria for effective trainings, and address teacher-child interactions.

Pathways to success for youth: What counts in after-school. Massachusetts after-school research study (MARS) report. Wellesley, MA, National Institute for Out-of-School Time.

[10] Kelley, P. and G. Camilli (2007). "The impact of teacher education on outcomes in center-based early childhood education programs: A meta-analysis." NIEER Working Papers.

1.1.3 LEARNING ENVIRONMENT

Better Beginning's Learning Environment section evaluates the level of planning and intentionality programs devote to developmental domains through four components of the Learning Environment section: planning and curriculum, portfolios, interest centers, and developmentally appropriate physical activities. The amount and quality of research conducted for each component varies across the three types of care. Findings related to the after-school environment are markedly different. We classify most of the Learning Environment standards as structural measures because they specify *what* tools are to be present, but they do not specify or observe *how* the tools are used to facilitate child development or peer and teacher-child interactions.

Better Beginnings designers intended each level of the Learning Environment category to indicate increasingly complex levels of intentionality, starting with routines and working up to daily planning. Although a quality rating system should address levels of quality above those expressed in a state's minimum licensing regulations, Better Beginnings language regarding routines is redundant of minimum licensing. We suggest clarification.

With the exception of curriculum use, we found minimal evidence that child outcomes are tied to items for the Learning Environment section. Absent of a curriculum, general planning is not associated with enhanced child outcomes in center-based care. One study suggested that family care teachers who make daily plans have better interactions that help children feel more secure and play better. For school-age care, one study found that daily plans may be associated with better academic outcomes.

The number of interest centers in a program has been linked to global quality but not to child outcomes. The kind of materials placed in the interest centers and how teachers guide children to use them are related to child cognition and social competence. For school-age children, if materials are adequate and accessible to youth, a sufficient variety of activities may be more related to outcomes.

Portfolios may be used for assessment to determine atypical development and/or to individualize care and instruction. Validation studies have identified problems with using portfolios for assessment purposes. We did not find studies examining whether the use of portfolios to individualize instruction helps children. We recommend that Better Beginnings define the intent of portfolio use. As currently written, this item may not discriminate between programs using portfolios with a developmentally appropriate intent from those arbitrarily collecting products or recording behaviors without further reflection. For school-age children, the Better Beginnings intent should also be clarified. If portfolios are to be considered a tool for tracking individual development, this standard contradicts the exemption of school-age programs from developmental assessment. If the intent is for portfolios to track individual progress and plan further programming, training will be necessary to ensure staff intentionality and consistency with a method unfamiliar to many school-age providers.

If programs are to be observed or inspected for compliance in planning and implementing developmentally appropriate physical activities, we classify Standard 2.C.3 as a process measure with greater influence on child outcomes. The school-age literature presents strong evidence that physical activity in care is associated with better outcomes. Literature related to physical activities and fitness for ECE is sparse and suggests that physical activity will have to increase above the amounts typically introduced in interventions and be combined with dietary education and parental outreach to curb current trajectories of children's weight gain.

The Learning Environment standard with the strongest ties to child outcomes requires a written curriculum plan. Curriculum use has been linked to global quality and to child outcomes more than any other item. The type of curriculum chosen will affect children in different ways, with the traditional nursery school approach being the least likely to produce academic, cognitive, and social outcomes. Curriculum is a distinguishing characteristic between poor and adequate care and is misplaced at the highest level of Better Beginnings. For school-age children, we would add that coordination with participants' schools is equally important to outcomes.

> There is strong evidence that **curriculum use** relates to developmental gains.

1.1.4 ENVIRONMENTAL ASSESSMENT

Of all Better Beginnings components, Environmental Assessment is the strongest because it utilizes the Environment Rating Scales (ERS) including:

- *Early Childhood Environment Scale, Revised Edition (ECERS-R)*,
- *Infant/Toddler Rating Scale, Revised Edition (ITERS-R)*,
- *Family Child Care Environment Rating Scale, Revised Edition (FCCERS-R)*, and
- *School-Age Care Environment Rating Scale (SACERS)*.[11]

These instruments measure the quality of care in a variety of settings. Over a period of 30 years they have become the most widely used quality measures in ECE practice and research. Empirical evidence has validated the relationship of ERS quality to child outcomes in child care research around the world. Findings are not always consistent and are modest in strength.

> Environmental Rating Scales lend strength to the QRIS in that they rely more on information collected by an independent observer than on provider reports.

Each ERS is a global measure. A global measure combines items rating structural aspects of the program—for instance, the physical layout of the space or staff qualifications—with observations of processes that directly involve children. ITERS-R and ECERS-R scores are modestly associated with child outcomes. Quality in the low range has been linked to children's elevated stress, anger and defiance, and setbacks in

[11] Harms, T., R. Clifford, et al. (1998). Early Childhood Environment Rating Scale, Revised Edition. New York, Teachers College Press. Harms, T., D. Cryer, et al. (2003). Infant/toddler environment rating scale, Revised edition. New York, Teachers College, Columbia University. Harms, T., D. Cryer, et al. (2007). Family Child Care Environment Rating Scale, Revised Edition. New York, Teachers College, Columbia University. Harms, T., E. Vineberg Jacobs, et al. (1996). School-Age Care Environment Rating Scale. New York, NY, Teachers College Press.

vocabulary and applied math development. Recent large, multi-state studies with sophisticated controls indicate that the higher the ERS quality, the greater the effect on child outcomes.

Better Beginnings is similar to other QRIS systems in permitting self-assessment for entry-level quality ratings. It diverges from other states by permitting scores in the adequate range for upper tiers of quality. The Missouri rating system evaluation found that all children lost social skills, and children in poverty lost vocabulary when enrolled in centers with lower ERS ratings. The lower quality ratings in Missouri are comparable to Better Beginnings Levels 1 and 2.

The evaluation team's validation analysis of Better Beginnings criteria for ERS scores identified some concerns. Findings from ITERS data collected as part of the national evaluation of Early Head Start (in which one of 17 sites included children in Arkansas) show a loss of emotion regulation and engagement skills for very young children at the lowest levels. Over time, children attending programs scoring lower than 3 on the ECERS had less optimal language, math, and social-emotional development. These findings were echoed in data collected from family child care programs in the national evaluation of Early Head Start. Children in family child care centers with scores lower than 3 had significantly lower cognitive, math, and language skills. Most worrisome, children in the lowest quality centers, regardless of program type, scored more than one standard deviation below the national average in cognitive, math, and language scores.

We found less evidence related to environmental quality in school-age children's development. A 2010 study showed that elements of SACERS scores are associated with growth in language, math, and social skills.[12] In lieu of the SACERS, Better Beginnings allows school-age programs to use the Youth Program Quality Assessment (Youth PQA) or the Younger Youth Program Quality Assessment (Younger Youth PQA).[13] The SACERS is a global measure that observes a program's structural and process features. Youth PQA and Younger Youth PQA focus more on process components. They emphasize staff-youth interactions, the socio-emotional climate of the program, and youth engagement. Data of SACERS, Youth PQA, and Younger Youth PQA were not available, so we could not validate Better Beginnings Levels for school-age children.

1.1.5 CHILD HEALTH AND DEVELOPMENT

Better Beginnings standards to improve child health and development involve sharing information with parents and documenting the implementation of medical and educational plans. Sharing information with parents is a structural measure. Level 1 programs share information on ARKids First and on child development and health. Level 2 programs share information regarding medical homes and on the stages of child development. Level 3 programs share information on nutrition and physical activity for children. Even when children are in pediatric care, children's developmental issues are often not adequately

[12] Pierce, K., D. Bolt, et al. (2010). "Specific Features of After-School Program Quality: Associations with Children's Functioning in Middle Childhood." American Journal of Community Psychology 45(3): 381-393.
[13] Adams, K., Brickman, N., & McMahon, T. (Eds.). (2005). Youth program quality assessment, Form A. Ypsilanti, MI: High/Scope Press. Adams, K., Brickman, N., & McMahon, T. (Eds.). (2005). Younger youth program quality assessment, Form A. Ypsilanti, MI: High/Scope Press.

addressed by medical providers.[14] Because child care providers have daily contact with parents and children, they are in a unique position to fill this gap.

Better Beginnings topics are appropriate to the needs of children, and there is some empirical precedence from ECE and pediatric literature to suggest that anticipatory guidance in the form of print information about child development or medical conditions increases use of medical and preventive care as well as parental willingness to communicate with providers. Some studies find that written information alone produces results. Others find that written guidance coupled with conversational guidance is more effective.

For young children in center-based care, Arkansas does not require screening as an element of care independent of the assessment of administrative practices (using the PAS/BAS). There is a requirement at the lowest level of Better Beginnings that "medical and educational care plans involving a child are written and on file, and implementation is documented" (1.E.3). In other words, programs need to adhere to an individualized plan (IFSP/IEP) for children with identified delays/disabilities.

1.2 RECOMMENDATIONS

Within any system, there is the opportunity for improvement. The evaluation team recommends refinements and revisions to reduce redundancy among components and to strengthen the influence of Better Beginnings on child outcomes. In making recommendations, we are mindful that changes to a system already implemented should be minimized to prevent possible resentment.

1. Reduce Redundancy

There are many elements of family involvement documented in Better Beginnings. There are items of the Strengthening Families training materials that are being assessed as part of the administrative and environmental assessments. We recommend that the Strengthening Families component be modified to exclude content areas already gauged with the PAS and ERS assessments.

2. Use Measures as Written and Tested

Better Beginnings should assess and score PAS items that are currently excluded. Teacher education measured by PAS is related to more optimal classroom practices. We recognize that providers may have difficulty achieving high scores on the items, but the original scaling of the instrument outperforms the scale with the excluded items. The state has already invested in the PAS and should take advantage of its validity and reliability testing.

3. Designate Teacher-Child Ratios

A key component present in other state rating systems but absent from Better Beginnings is required teacher-child ratios. States and organizations seeking to improve child outcomes via research-based practices have adopted guidelines for limiting the number of children in a teacher's care. Arkansas minimum licensing allows less optimal teacher-child ratios, especially for birth to two years, than most

[14] Schuster, M. A., N. Duan, et al. (2000). "Anticipatory Guidance: What Information Do Parents Receive? What Information Do They Want?" Archives of Pediatrics & Adolescent Medicine 154(12): 1191-1198.

key comparison states, the National Association for the Education of Young Children (NAEYC), and Head Start. Improving ratios is expensive. Within Arkansas, stakeholders have rejected attempts to adjust minimum licensing regulations. Nevertheless, in light of consistent evidence that ratios affect child outcomes, we recommend that Better Beginnings include requirements for teacher-child ratios that exceed those found in minimum licensing.

4. Incorporate Process Measures

The evaluation team recommends incorporating process measures because they are stronger predictors of child outcomes. There are structural measures in Better Beginnings that could be strengthened by altering how information is collected. One example that is strongly supported in the literature is use of a curriculum. There is evidence that relying on program-reported use of curricula will be less accurate than using independent observations of use. Developing methods to observe adherence to a curriculum during an assessment visit versus permitting self-report of curriculum is advised.

There is also strong evidence to support inclusion of teacher-child interactions assessments. The incorporation of a new instrument would be costly. As an alternative, Better Beginnings could more closely track teacher-child interactions already being observed with the ERS instruments. The evaluation team found evidence that children in programs with ERS Interaction subscale scores that do not meet the minimum criterion score for the overall ERS have less optimal cognitive and social development. These findings could be used to support revised Better Beginnings standards requiring programs to meet an overall minimum ERS score and also the same minimum for the Interactions subscale. Increased technical support for programs in the area of teacher-child interactions is warranted when scores on the Interaction subscale of the ECERS-R are substantially lower than the overall score.

5. Address Lower Levels of Quality

Analyses examining current cut scores on the ERS would lead us to caution accrediting programs with a quality rating when at least a minimum score of 3 has not been met. An implicit goal of Better Beginnings is to communicate to parents the importance of quality child care for their children's development. We recommend that Better Beginnings Level 1 be considered a "getting ready" level that invites participation but also communicates to parents that programs have not yet been assessed and may not reflect a minimal level of environmental quality.

6. Address Higher Levels of Quality

We recommend the development of levels beyond the current highest level of Better Beginnings to encourage programs to make improvements that promote optimal child development. The range of low scores typically recognized in other states' QRIS is either 3.0-3.75 (N=13) or 4.0-4.5 (N=6). The range of high scores is typically in the 5.0 – 5.5 range (N=13) or higher (N=4). [15] Better Beginnings Level 3, the highest rating in the system, requires average ERS scores of 4, a substantial divergence from other quality systems. Analyses showed that children in higher quality programs (meeting cut scores of 5 and 5.5 that the UAMS evaluation team proposes for future Better Beginnings Levels 4 and 5) had higher cognitive and academic skill scores than children in lower levels. In awareness of print materials and phonemic

[15] Tout, K., R. Starr, et al. (2010). Compendium of Quality Rating Systems and Evaluations. Child Trends, Mathematica Policy Research.

knowledge, children in our proposed Level 5 fared even better than children in our proposed Level 4 programs.

7. Include Child Screening

At the lowest level of Better Beginnings programs must adhere to an individualized plan (IFSP/IEP) for children with identified delays/disabilities. However, Arkansas does not require screening as an element of care independent of the assessment of administrative practices. (In PAS and BAS, programs with good to excellent scores facilitate screenings for children in their care). We recommend that quality ECE programs implement efforts to identify children with special needs and make referrals for early intervention. Without screening, delays and disabilities can stay unaddressed for years. Furthermore, young children are more responsive to intervention than at any other time. Early identification increases the possibility that applied intervention will be effective, reduces education costs, and alleviates hardship for children and their families.

Better Beginnings includes some characteristics of an overall level of quality that have been found to contribute the most heavily to child outcomes in recent years. Better Beginnings draws on the elements common to successful model preschool programs when deemed feasible for local providers. It includes standards for the education and training of staff, parent involvement, and use of a curriculum, as well as global measures of program quality as rated by independent observers. Moreover, the component choices that were made in Better Beginnings were similar to those of the other states we reviewed. Empirical evidence tells us that the quality of child care affects children in various ways, and multiple determinants of quality are included in the Better Beginnings system.

2 INTRODUCTION

Public and political beliefs are beginning to shift from viewing child care as a mere custodial duty to an important educational function with the potential to improve long-term individual and societal health. With greater understanding of brain function and early learning, we see that more rapid development takes place from birth to age 5 than at any other stage of life. Young children are vulnerable to risks present in their caregiving environments, but fortunately, are also more resilient and responsive to intervention than at any other time (Shonkoff & Phillips, 2000).

Evaluators of early model preschool programs have persistently followed participants into adulthood and offer alluring evidence that high quality child care can confer long-term benefits. For instance, the randomly assigned treatment group participating in the High/Scope Perry Preschool Project in the sixties has experienced fewer arrests and higher incomes with less use of public assistance than the control group (Nores, Belfield, Barnett, & Schweinhart, 2005). Another model program with a randomized evaluation, the Carolina Abecedarian Project, provided full-day, year-round care to children believed to be at risk for developmental delays from birth to age 5. Following these children into adulthood, researchers found that participants in the preschool treatment group were more likely to have sustained better math and reading abilities and to have completed college. They were also less likely to have repeated a grade, to have required special education, and to have become teenage parents (Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002; Ramey et al., 2000). While Perry Preschool and Abecedarian projects were small demonstration projects, the Chicago Child-Parent Center (CPC) project also has provided evidence that large-scale federally funded projects may also produce long-term positive effects. Low-income, mostly black children from urban areas who completed one or two preschool years in the program required less special education and grade retention, had higher high school completion rates, and had fewer arrests at age 20 (Arthur J. Reynolds, Temple, Robertson, & Mann, 2001).

These model programs shared some common elements:

- well educated, well paid staff;
- good teacher-student ratios (usually 1:3 for infant/toddlers up to 1:6 for preschoolers);
- extensive education, training, and involvement for parents;
- and curricula with a strong theoretical basis (Yoshikawa, 1995).

Some states have amplified or enhanced minimum licensing standards, but most do not come close to guaranteeing that all children experience these elements of model programs. The continued reauthorization of federal funding for Head Start and Arkansas' state-funded establishment of Arkansas Better Chance programs do represent a societal shift towards acknowledgement of the importance of early care and education. However, legislative bodies, citizens, and even parents, have not yet been willing to allot the level of funding necessary for the widespread implementation of comprehensive, intensive programs. Thus, a number of states are in the process of exploring less costly alternatives to enhance quality of care and education for all children. One such alternative that many states are designing or have already implemented is the Quality Rating System (QRS) or Quality Rating and Improvement System (QRIS).

"A QRIS is a systemic approach to assess, improve, and communicate the level of quality in early and school-age care and education programs. Similar to rating systems for restaurants and hotels, QRIS award quality ratings to early and school-age care and education programs that meet a set of defined program standards. By participating in their State's QRIS, early and school-age care providers embark on a path of continuous quality improvement. Even providers that have met the standards of the lowest QRIS levels have achieved a level of quality that is beyond the minimum requirements to operate" (National Child Care Information and Technical Assistance Center).[16]

Similar to other states, Arkansas' Better Beginnings QRIS (BB) draws on the elements common to the successful model preschool programs when deemed feasible for local providers. It includes standards for the education and training of staff, parent involvement, and use of a curriculum.

The purpose of this investigation is to summarize the findings from early childhood education (ECE) and school-age research relevant to the BB standards and to use pre-existing data to predict the impact of this new system on children in Arkansas. A multitude of standards exist in the ECE world. Accrediting bodies, state governments, national advisory councils, and federally funded programs have made varying recommendations. The overlap and interplay of these regulations complicates the design and implementation of a state QRIS, which must consider how to improve the lives of children in the most efficient and economically responsible way possible without crushing individual child care providers with the weight of high costs or bureaucratic procedure. The good news is this: empirical evidence tells us that the quality of child care affects children in various ways, and some determinants of quality are already included in the Better Beginnings system.

[16] http://nccic.acf.hhs.gov/qrisresourceguide/index.cfm?do=qrisabout#1. Accessed 9/7/2010.

3 METHOD

The BB QRIS "is designed to help programs improve their day to day environment for children, and to establish proven administrative practices" [17] with the desired output being long-term improvements for child outcomes. We approached the evaluation of BB with three main questions.

★ Key Evaluation Questions ★

1. Are the Better Beginnings standards in alignment with current research on child outcomes related to all types of child care?

2. What can we learn from states that have already implemented and evaluated a quality rating and improvement system?

3. Are child outcomes predicted by Better Beginnings levels?

3.1 LITERATURE REVIEW

Our first step in answering these questions was to conduct a literature review on child care related to child outcomes. Measuring child outcomes is expensive, methodologies and measurements are debated, and the measured impact of elements of child care is small in comparison to the effect of elements of the home environment and parental characteristics. Most projects, especially state QRIS evaluations facing tight funding, simply cannot afford the process of tracking child outcomes. Given these limitations, we also reviewed research examining classroom interactions and global quality, in light of evidence that both elements affect child outcomes in different ways.

★ Approach to the Literature Review ★

- What elements are empirically linked to child outcomes?
- What elements affect teacher or parent interactions with children?
- What elements raise or reduce global quality?

Initially we conducted searches for literature related to outcomes, interactions and quality through online databases, giving preference to peer-reviewed, published articles. Given the necessary involvement of state and federal governments and advisory councils, we also reviewed numerous reports commissioned by federal, state, and local agencies and position statements or policy recommendations disseminated by advisory councils.

[17] http://www.state.ar.us/childcare/bb/full%20book.pdf

Finally, we conducted additional micro-searches to review literature pertinent to each of BB sections and items. For instance, Section A. Administration queries included "Program Administration Scale", "child care administration" and "child care and business practices" to address items 1.A.1, 2.A.1, and 3.A.1; and "Strengthening Families," "child abuse and neglect prevention," "parent involvement" and "family involvement" to address Items 2.A.2, 3.A.2, and 3.A.3.

3.2 COMPARISONS

The next step in our evaluation was to compare Arkansas standards with current best practices manifested in standards of national accreditation organizations, Head Start, or other state rating systems. Throughout this report, crosswalk summaries identify areas where BB aligns with these standards. Crosswalks may also be used to identify ways that BB will prepare programs for accreditation and to identify areas that may prove redundant or cumbersome for programs already participating in other accreditation. Also, DCCECE is considering automatically conferring Level 3 status to accredited programs with verification of only a few items. The UAMS team has conducted crosswalks to verify whether accreditation standards meet or exceed Better Beginnings standards.[18]

Recognizing that all children deserve high quality care, states are experimenting with systems to improve all types of non-maternal care. As of 2009, 26 states were operating a QRS (Kathryn Tout et al., 2010). In addition to reviewing the formal academic research, we gathered information about other states' quality rating systems, collecting standards and evaluations when available to determine what we could learn from predecessors. Other important resources were the National Child Care Information and Technical Assistance Center (NCCIC) and personal communication with QRIS officials or independent researchers who may have contributed to the design or evaluation of quality measures or state rating systems.

To narrow the scope of our states' comparisons, we selected six key states to highlight in this report. Colorado, North Carolina, Ohio, Oklahoma, and Pennsylvania were identified and compared as QRS "early adopters" by Zellman and Perlman (2008). Each of these states has implemented and evaluated a state-wide quality rating system with which Arkansas may compare standards and use to predict outcomes. We additionally included Missouri because the state recently released a QRS evaluation that included child outcomes. To date, Missouri and Colorado are the only states to have reported on child outcomes. Ohio's study of child outcomes is ongoing through 2011.

Three of the key comparison states highlighted in our report—North Carolina, Missouri, and Oklahoma—have built a QRS into their licensure. Whereas states with combined systems reap about 60% participation among state child care providers, states with voluntary systems, like BB, usually achieve 30% or less. Table 3-A, on the following pages, describes characteristics of the key comparison states' rating systems and summarizes their evaluation findings.

[18] Full crosswalks are contained in this report's Technical Appendices available at the Better Beginnings website (http://www.arbetterbeginnings.com/).

Table 3-A Key States QRIS Description

State, QRIS Name	Categories of QRIS Standards	QRIS Levels/Point Systems	Rating Assignment Method	Evaluation Results
Colorado **Qualistar Rating System**	• Learning environment • Family partnerships • Training & education • Adult-to-child ratios & group size • Accreditation	• Provisional • Star 1 • Star 2 • Star 3 • Star 4	**Point system:** Providers earn points in 5 areas of standards, up to a total of 42 points. In 4 areas of the standards, 0–10 points can be earned. Zero or 2 points are earned for program accreditation. Stars are earned when a provider has a total of at least 10 points. A provider who earns 0–9 points is designated as provisional. **Voluntary participation** Approximately 20% participate; 73% are rated in the two highest levels.	No strong links between QRIS components and child outcomes or between star ratings and outcomes were found. Problems with the measurement of family involvement and teacher education and training were identified. Ratings and QRIS components did not relate well to measures of teacher-child interactions. Quality improved over time. Findings are not generalizable because of non-randomized design, mid-study change of a measure, provider attrition problems, and very high attrition of children (G Zellman, Perlman, Le, & Setodji, 2008).
Missouri **Missouri Quality Rating System**	Separate standards for centers, home-based programs, and school-age centers • Director education & training • Staff education • Education specialization • Annual training • Learning environment • Intentional teaching • Family involvement • Business & administrative practices	• Tier 1 • Tier 2 • Tier 3 • Tier 4 • Tier 5	**Point system:** Providers earn points in 8 areas of the standards to obtain each tier. Components are progressive; standards in a component at a lower level must be met to achieve the next higher level. **Tier 1 required for licensure** % participation not available	Compared to children in low tiers, children in high tiers made significant gains in social and behavioral skills, motivation, self control, and positive adult relationships. Children in low tiers lost skills in all these areas. Children in poverty made vocabulary gains in high tiers and lost vocabulary in low tiers. Those attending medium and high levels gained vocabulary skills, but those in low quality lost vocabulary. Children in poverty also gained in letters/sounds and gross motor regardless of level, but magnitude of gains was greater in high quality (Thornburg, Mayfield, Hawks, & Fuger, 2009).

North Carolina **North Carolina Star Rated License**	There is one set of standards for all facilities in two areas: program and education. The following is a summary of what is included in each area
	Program standards
	• Program environment
	  ○ Sufficient space for activities
	  ○ Variety of play materials
	  ○ Clean and comfortable play area
	• Number of staff per child
		• 1-star license
		• 2-star license
		• 3-star license
		• 4-star license
			Points system:
			Points are earned in two standard components–education and program–to obtain each level of the star rated license.
			
			1-star mandatory for licensure
			100% of licensed programs participate; 47% of participating centers are rated in the two highest levels.
				Overall quality in preschools and school-age classes was good but not so for infant-toddlers. The study revealed a strong connection between teacher education and rating scale scores. A two-year degree was found to be necessary to achieve good quality. One- and 2-star centers were not well represented (D. Cassidy, Hestenes, Mims, & Hestenes, 2003).

State, QRIS Name	Categories of QRIS Standards	QRIS Levels/Point Systems	Rating Assignment Method	Evaluation Results

	• Interactions between the following:
	  ○ Adults and children
	  ○ Children with other children
	  ○ Children with activities and materials
	Education standards
	• Administrator education & experience
	• Number of lead teachers with child care credentials
	• Number of lead teachers with more EC education and experience
	• Number of teachers with formal education and/or experience
		• 5-star license		

Ohio **Step Up to Quality**	There are separate standards for child care centers and FCC homes. • Ratio and group size • Staff education and qualifications • Specialized training • Administrative practices • Early learning	• Licensing • Step one • Step two • Step three	**Building blocks:** To be eligible for the QRIS, providers must not have had serious noncompliance during their last regular licensing inspections. Accreditation is an alternative pathway for meeting child-staff ratio and group size requirements for steps two and three. **Voluntary** 24% of licensed programs participate; 12% of participating program are rated in the two highest levels.	Wave 1 research found steps were not significantly different in overall ECERS-R quality, possibly because of measures chosen may not be a good match for the state's goal of increasing school readiness. All benchmarks were important to the system. Director's attitudes toward SUTQ were generally positive. Wave 2 research found children in Steps 2 and 3 had higher problem solving skills than Step 1. Children in Step 3 had higher independence than lower steps. Children in Step 1 had lower adjusted mean scores on a self-regulation measure (Ohio Collaborative, 2009).
Oklahoma **Reaching for the Stars**	There are separate standards for child care centers and FCC homes. • Licensing status and compliance • Administrative • Qualifications and training (for directors, staff, and providers) • Learning environment • Parent involvement • Program evaluation	• One star • One star plus • Two star • Three star	**Building blocks:** All licensed providers are automatically designated as one star. For each remaining level, providers must meet the requirements of that level and those of previous levels. Three star providers are accredited. **One star mandatory for licensure** 100% of licensed programs participate; 46% are rated in the two highest levels.	Increased financial support was a significant motivator to participate. High quality centers had teachers with higher levels of EC education. Subsidy density, Master teacher-child ratio, parent involvement and interest centers were variables with strongest relationship to quality. The study validated different levels of quality in star quality in family child care. Two star FCC homes have more positive interactions than one star, and specialized education and ongoing professional development were found crucial to ratings (Norris & Dunn, 2004).

State, QRIS Name	Categories of QRIS Standards	QRIS Levels/Point Systems	Rating Assignment Method	Evaluation Results
	• Master teachers (for centers only)			
Pennsylvania **Keystone STARS**	There are separate sets of standards for child care centers, group FCC homes, and FCC homes. • Staff qualifications and professional development • Learning program • Partnerships with family and community • Leadership and management • Continuous quality improvement (for FCC only)	• Start with STARS • STAR 1 • STAR 2 • STAR 3 • STAR 4	**Building blocks:** Providers in the Start with STARS level have current licenses. For each remaining level, providers must meet the requirements of that level and those of previous levels. To reach the STAR 4 level, providers must meet all STARS performance standards, or be accredited and meet STAR 4 accreditation standards. **Voluntary** 60% participate; 18% are rated in two highest levels.	The study validated the system as a reliable indicator of quality. Ratings for centers and homes related to ERS. The system helped reverse a declining trend in quality. FCC and centers with a defined curriculum and teachers with at least an AA have higher quality. Center-based teachers with 5 years experience and FCC teachers with 20 years experience had higher ERS (Barnard, Smith, Fiene, & Swanson, 2006).

Sources: The majority of this table is borrowed from "QRIS Standards, Levels, and Rating Systems" by the National Child Care Information and Technical Assistance Center (NCCIC) at http://nccic.acf.hhs.gov/poptopics/qris_systems.html. For evaluation summaries and links see NCCIC's: "QRIS and the Impact on Quality in Early and School-Age Care Settings" (http://nccic.acf.hhs.gov/poptopics/qrs-impactqualitycc.html). A description of Missouri's QRIS is available at www.openinitiative.org. Ohio's unpublished report of findings in 2009 was supplied by Jamie Gottesman, Assistant Chief of Ohio's Bureau of Child Care and Development.

As specified above, there are three structural systems currently used for state QRIS:

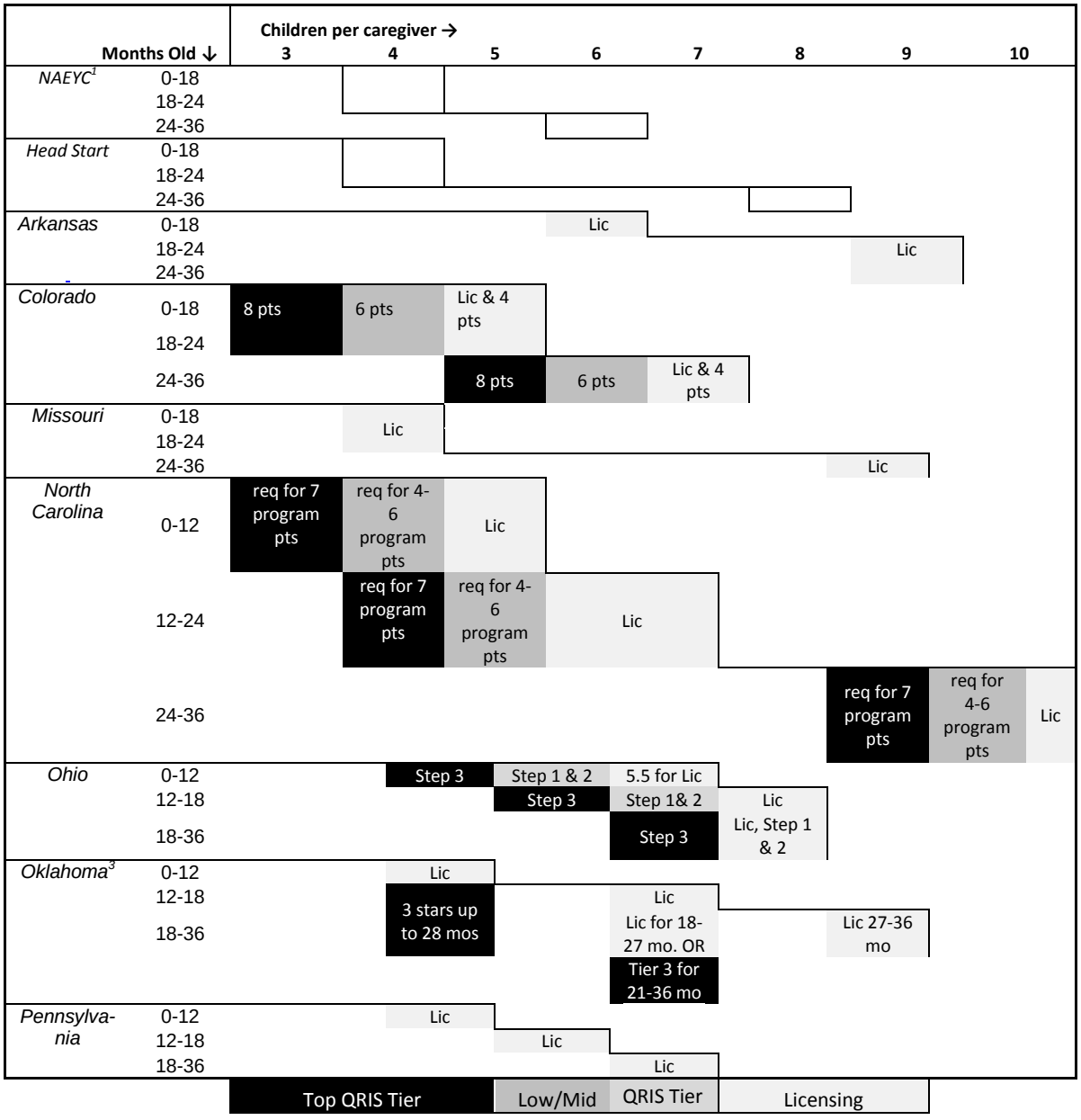
- A **building block system** requires a program to meet all standards for one level before moving to the next level. BB is a building block system.
- A **point system** assigns various numbers of points to each of the standards and adds them together for composite scores. Programs that have earned enough points across standards may move from one level to the next.
- A **combination system** may use points and building blocks.

We suspect that differences may emerge as QRISs mature and usage expands, but thus far, research has not yet addressed varying effects that the different system types may produce. In addition to consistently reporting on these key states throughout our review, we also mention significant features or findings from other states when useful.

BB does have many categories in common with other states. However, a key feature it does not share with many other states is required teacher-child ratios. Improving ratios is expensive, and within Arkansas, stakeholders have rejected attempts to adjust current minimum licensing regulations. There is enough consistent evidence that ratios affect child outcomes that states and organizations seeking to improve child outcomes via research-based practices, for instance, The National Association for the Education of Young Children (NAEYC) and Head Start, have adopted guidelines for limiting the number of children in a teacher's care. [19] Presumably to increase providers' willingness to participate, BB authors did not include a ratio component. In comparing states and their evaluations, though, it is important to remember this difference. Table 3-B Child-Caregiver Ratios for States Licensing and QRIS**,** on the following page, shows that Arkansas holds lower standards for ratios, especially for birth to two years, than most key comparison states, NAEYC, and Head Start.

[19] (for example, M. R. Burchinal & Roberts, 2000; Frede, 1995; Helburn, 1995; C Howes, Phillips, & Whitebook, 1992; John M. Love et al., 2004; John M. Love et al., 2003; Mosteller, 1995; Phillips, Mekos, Scarr, McCartney, & Abbott-Shim, 2000)

Table 3-B Child-Caregiver Ratios for States Licensing and QRIS



State licensing and QRIS sources:
http://nccic.acf.hhs.gov/pubs/cclicensingreq/ratios.html
http://www.naralicensing.org/displaycommon.cfm?an=1&subarticlenbr=160
http://nccic.acf.hhs.gov/poptopics/qrs-criteria-websites.html

[1]**NAEYC** has age ranges that overlap. If a group includes children whose ages range beyond the overlapping portion of two age categories, then the group is a mixed-age group. For mixed-age groups, universal criteria and criteria relevant to the age categories for that group apply. In mixed-age preschool groups of 21 2-year-olds to 5-year-olds, no more than 4 children between the ages of 30 months and 36 months may be enrolled. The child-teacher ratios are lowered when children with special needs are present or when square footage is low.

[2]**Head Start** "Performance Standard Grantee and delegate agencies must ensure that each teacher working exclusively with infants and toddlers has responsibility for no more than four infants and toddlers and that no more than eight infants and toddlers are placed in any one group," (45 CFR 1304.52(g)(4)). When a class has predominantly 3-year-olds in a standard program, an average of 15-17 children may be enrolled per class in a standard class [with two staff members required in each class]. For 3-year-olds in a double session, 13-15 children may be enrolled per class in these classes. No more than 15 children enrolled in any class (45 CFR 1306.32).

[3]**OK** centers must be accredited to earn 3 stars. NAEYC ratios used here.

3.3 CLASSIFICATION OF MEASURES

The Administration for Children and Families Office of Planning, Research and Evaluation (OPRE) recommends that all accountability systems, including state QRISs, define their purposes, articulate desired outcomes, and carefully choose measures that align with those outcomes (Child Trends, 2009). Measures that are aligned with the desired outcome will have greater predictive power. If the purpose of a system is to give parents information about quality across settings, a global measure that provides a summary will be best. On the other hand, if school readiness is the primary goal, then an instrument must be chosen that measures instructional quality or classroom practices. Measures in ECE research are typically broken into three types: structural, process, and global.

- ***Structural features*** are regulatable components of child care thought to be the foundation of good quality care. Examples of structural features are teacher-child ratios and group size, staff education or credential, and regulations for physical space, such as square foot per child. Many rating systems rely heavily on structural features as measures because they are less time consuming and less costly to implement and enforce.
- ***Process measures*** are those that observe interactions that directly involve the child. Activities and materials presented to children, behavior management, and the responsiveness or sensitivity of teachers to a child's needs are examples of process features. Within research literature, measures of these processes are typically found to be most directly related to child outcomes.
- ***Global measures*** rate both structural and process features of care. Environmental rating scales are considered global measures because they observe the physical layout and routines as well as interactions.

Research typically identifies an indirect relationship between structural features and child outcomes. In other words, structural features improve child outcomes by setting the stage for good processes to occur (Child Trends, 2009). There is evidence that structural features influence the quality of teacher-child interactions in family- and center-based care (Fischer & Krause Eheart, 1991; E. Galinsky, Howes, & Kontos, 1995; NICHD Early Child Care Research Network, 1999, 2000; Phillips, et al., 2000). For example, when Florida implemented stricter standards for child care, such as lower teacher-child ratios and higher credentialing requirements, the quality of caregiving and child outcomes improved considerably (C. Howes & Smith, 1995). It is thought that some structural measures enhance child development via their improvement of interactions. Those testing the paths of influence find that structure affects process, which then affects child outcomes (for example, cognitive competence as in C Howes, et al., 1992; NICHD Early Child Care Research Network, 2002). In the case of ratios, teachers with fewer students may be able to respond faster and more positively to children's needs than teachers juggling more students and subsequently receive more cooperation in learning processes.

Although easier to mandate, there are downsides to structural measures. Sandra Scarr's research teams identified vast violation of compliance with state regulations and found that direct observation of process measures was much more accurate in assessing program quality (S. Scarr, Eisenberg, & Deater-Deckard, 1994; Sandra Scarr, Phillips, McCartney, & Abbott-Shim, 1993). For systems evaluation purposes, there is probable "noise" between the structural measure and the outcome. This noise will limit the statistical size of the find and may even create confusion as to whether the regulation or something else is producing an outcome. Structural measures are appealing for their lower up-front price tags, but they may ultimately

cost accountability and technical assistance systems more money by not directly addressing and improving what some investigations identify as the heart of the ECE matter: classroom interactions.

In comparative studies, measures of teacher-child interactions show more robust relationships with child outcomes than structural and even global measures that incorporate items related to interactions (Beller, Stahnke, Butz, Stahl, & Wessels, 1996; Mashburn, 2008; Mashburn et al., 2008). Interactions are typically classified into those that support the social-emotional well-being of the child and those that stimulate cognitive or academic processes. We summarize research related to both types of teacher-child interactions in the Learning Environment and Environmental Assessment sections.

With this body of evidence in mind, we classify each BB standard according to the type of measure—structural, process, or global. Similar to many state-wide systems, BB largely deals in structural measures. However, the process-oriented standards within BB should have greater predictive power for outcomes. In some cases, there is evidence that a BB structural measure will enhance the quality of classroom interactions, which lends greater validity to the standard. Within all the BB measures, subscales and items dealing with process components of quality, such as teacher-child interactions, should be paid close attention in evaluative, technical assistance, and professional development efforts.

4 ADMINISTRATION

The Administration section of BB invites programs to improve their programs through self study and action steps in two areas:

1. **business and management practices** as measured by the Program Administration Scale (Talan & Bloom, 2004) and the Business Administration Scale (Talan & Bloom, 2009), and

2. **child abuse and neglect prevention** through parent support and involvement according to guidelines provided by the Strengthening Families initiative.

Within this section, we will address related outcomes and conclusions for each of these two items separately.

4.1 ADMINISTRATION SCALES

To capture a multi-dimensional picture of a program's quality, Better Beginnings incorporates two standardized instruments to measure leadership and management functions in ECE programs, *The Program Administration Scale* (PAS; Talan & Bloom, 2004) for centers and *The Business Administration Scale* (BAS;Talan & Bloom, 2009) for family day care. These are the first published instruments that solely focus on the administrative processes of early childcare programs. These processes, while intended to enhance the experiences of children, are established and staged outside of the classroom, so both scales are categorized as *structural* measures. Because these scales were not developed for school-age programs, BB may consider folding in a third measure, Youth Program Quality Assessment (Youth PQA), Form B (Adams, Brickman, & McMahon, 2005c).

4.1.1 PROGRAM ADMINISTRATION SCALE (PAS)

The PAS includes 25 items clustered in 10 subscales using a 7-point Likert scale similar to that of the ERS, 1 being inadequate and 7 being excellent. Subscales include Human Resources Development, Personnel Cost and Allocation, Center Operations, Child Assessment, Fiscal Management, Program Planning and Evaluation, Family Partnerships, Marketing and Public Relations, Technology, and Staff Qualifications. Within Staff Qualifications, Teacher and Teacher/Aide items are considered optional depending on the center's staffing patterns.

The instrument was tested for reliability and validity in 67 centers. The PAS had acceptable internal consistency (coefficient alpha .85), distinctive but related subscales and items, and 90% inter-rater reliability among 8 assessors. An analysis of variance (ANOVA) using NAEYC-accredited programs to differentiate programs of different quality showed that accredited programs had significantly higher

scores than unaccredited programs in the sample (M=92.12, S.D.= 19.43 and M=72.06, S.D. 20.83 respectively). [20]

Separate from the author's initial tests, Lower and Cassidy correlated the PAS with global quality (2007). Their concurrent validity tests comparing the PAS to the ECERS-R Parents and Staff Subscale and to the *Early Childhood Work Environment Survey* (ECWES;Bloom, 1996) indicated that the PAS was related to but not redundant of these other measures of organizational quality. **This study also confirmed the importance of formal assessment to verify stated practices.** "A paired sample t-test revealed that the mean score of directors' stated practices (M=3.25, SD=1.04) was significantly different than the mean PAS score following document verification (M=2.87, SD=.88), t(29)=-6.73, p=.00" (p.7). [21]

In the interest of encouraging wider Arkansas center participation in quality improvement, BB excludes items 22-25 that address administrator and teacher qualifications from the assessment process. Also, items 5 and 6 that rate staff benefits and staffing patterns and scheduling will be scored to advise program improvements, but these items will not be counted in the program's overall score. Additionally, when the PAS is applied to programs serving school-age children, the subscale measuring child assessments will be excluded. Once children enter the school system, assessments should be a typical part of their educational experiences, and most school-age programs will not complete additional assessment.

Using data collected as part of the Evaluation of the Arkansas Early Childhood Professional Development System (AECPDS), we have examined the scoring of the PAS and the subscales that are being excluded from BB in relationship to other elements of program quality. Our goal is to determine if the exclusions to the BB scoring of the PAS might change the external validity of the measure. Across most indicators of program quality, the original scoring of the PAS and the BB scoring of the PAS (BB PAS), which excludes the Staff Qualifications subscale and items 5 and 6 of the Personnel Cost and Allocation subscale (addressing benefits and staffing patterns/scheduling) was similarly related to the Environmental Rating Scales (ERS) and Caregiver Interaction Scale (CIS) subscale scores (Table 4-A). In all cases, the bivariate correlations between the BB PAS and ERS and CIS scales were weaker than with the original scoring of the PAS. One difference between the scoring was evident for Arnett Cognitive Support of the Child, with BB PAS scores not significantly related to these teacher behaviors. This is worrisome as the Cognitive Support scale assesses behaviors that are key to the child's cognitive development and school readiness (items such as dialoguing with and engaging the children with open-ended questions and encouraging children in the use of symbolic/literacy materials, numbers and spatial concepts, and planning and problem solving) and suggest that changes to the PAS as instituted in BB may have deleterious outcomes on the measure's external validity.

Examining the excluded scales and items separately, we find that both Staff Qualifications and Personnel Cost and Allocation are related to elements of program quality. Correlations between the Personnel Cost and Allocation scale and the ERS and Arnett Scores are higher than for the full PAS and BB PAS. Furthermore, the Personnel Cost and Allocation scale is the only measure that is significantly related to Harshness of Interaction scale on the Arnett. This scale measures open irritation and hostility with the

[20] Reported scores are based on 23 items (possible range 23-161) because sections 24 and 25 related to assistant teachers and aides were not relevant to all centers.
[21] Reported scores are based on the overall PAS score of 1-7.

children, including threatening the children in attempts to control behavior and punishing children without explanation.

Table 4-A Bivariate Correlations between Program Administration Scale and Classroom Quality

	PAS: Original Scoring	PAS: BB Scoring (no 5-6 and 22-25)	PAS Subscale: Staff Qualifications[a]	PAS Subscale: Personnel Cost and Allocation[b]
Environmental Ratings (Infant-Toddler/Early Childhood)	.44**	.42**	.30**	.50**
Arnett Caregiver Interaction Subscales				
Sensitivity	.39**	.37**	.21**	.45**
Harshness	-.13	-.12	-.05	-.23**
Detachment	-.48**	-.46**	-.27**	-.50**
Permissiveness	-.17*	-.17*	-.08	-.18*
Cognitive Support	.18*	.15	.21*	.34**
Social-Emotional Learning	.31**	.28**	.22*	.40**

Note: *p<.05, **p<.01; a=Staff Qualifications subscale is completely omitted from the scoring of the PAS in Better Beginnings, b=two of three items in the Personnel Costs subscale are omitted from scoring of the PAS in Better Beginnings.

Given the evidence that teacher education is related to more optimal classroom practices and the validity of the instrument may be detrimentally impacted, the UAMS evaluation team recommends re-introducing the items that have been excluded from BB. While we recognize that providers may have difficulty achieving high scores on the items, the original scaling of the instrument outperforms the scale with the excluded items. Formal PAS assessment does not occur until programs apply for a BB Level 3 rating; therefore, inclusion of the items for programs at the lowest levels would not adversely impact ratings' cut scores. The state has already born the burden of the cost of the assessment (both the costs associated with its use and its application state-wide) and should reap the maximum benefit it offers through its validity and reliability testing, and comparability with the QRIS of other states.

4.1.2 BETTER BEGINNINGS' LEVELS: PROGRAM ADMINISTRATION SCALE

Using the Evaluation of the Arkansas Early Childhood Professional Development System (AECPDS) data, we developed cut scores on the PAS (using the BB scoring; BB PAS) to determine whether the BB Levels are meaningful for predicting teacher-child interaction and global quality ratings on the ERS. Levels 1 and 2 do not require scoring on the PAS, while Level 3 requires a PAS score of at least 4 on the items discussed above (1-21, 5 and 6 not included in computation of the average score). These analyses are based in an assumption that programs in Levels 1 and 2 would score lower than 4 on the BB PAS as this is the criterion score required to attain Level 3.

We examined Caregiver Interaction Scores (CIS; Arnett, 1989) and ERS that were collected for the relevant types of care from the AECPDS. Each of the CIS and ERS scale scores was significantly different across BB Levels (with the noted exception of Harshness and Permissiveness, which was trend level, on the CIS). Table 4-B presents means, standard deviations, and sample sizes for each of the mean comparisons.

Table 4-C shows the BB Levels' relationship to CIS scores. CIS scores range from 1 to 4 with higher scores indicative of higher levels of behavior. Across most teacher behaviors, centers achieving Level 3 of BB

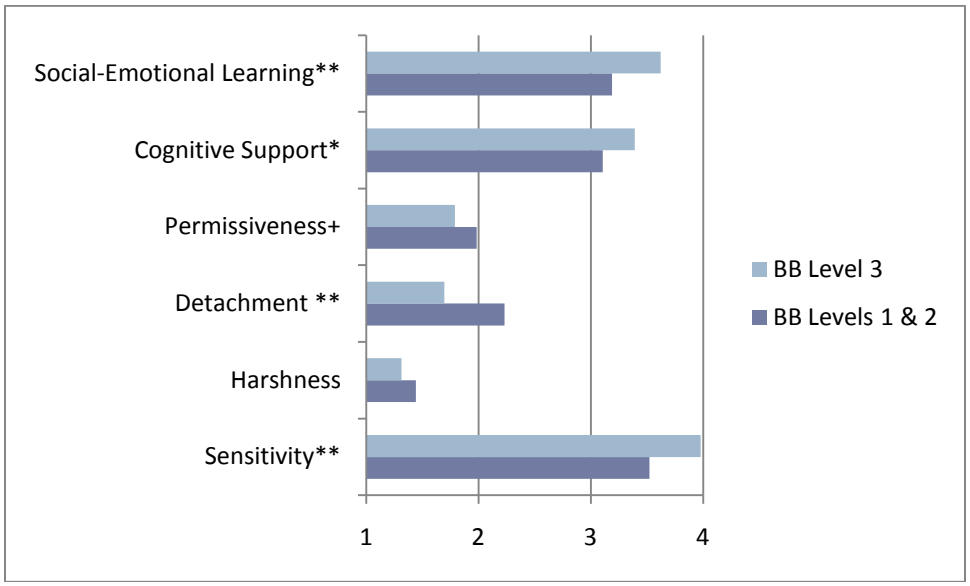
score significantly different than those in Level 1 and 2 centers, and in the optimal direction. Teachers in Level 3 centers provide care that is more sensitive, more supportive of cognitive and social-emotional development, and less detached. Further, there was a trend towards Level 3 teachers being less permissive than those in Levels 1 and 2. CIS scores for centers with PAS scores lower than 4 (BB Levels 1 and 2) were still relatively high in this sample. It is important to keep in mind that the majority of centers in the original study design were drawn from Arkansas Better Chance (ABC) and Head Start programs that were already Quality Approved by the state of Arkansas. It is unknown how these scores would look in for-profit programs and, unfortunately, existing data are not available.

Table 4-B: Validation of Better Beginnings Cut Scores on the Program Administration Sale

	BB: Levels 1 & 2			BB: Level 3			Total		
CONSTRUCTS	Mean	SD	N	Mean	SD	N	Mean	SD	N
Caregiver Interaction Scores									
Sensitivity**	3.52	.76	57	3.98	.70	91	3.80	.76	148
Harshness	1.44	.49	57	1.31	.48	91	1.36	.48	148
Detachment **	2.23	.79	57	1.69	.53	91	1.90	.69	148
Permissiveness+	1.98	.77	57	1.79	.56	91	1.86	.65	148
Cognitive Support*	3.11	.85	57	3.39	.84	91	3.28	.85	148
Social-Emotional Learning**	3.19	.79	57	3.62	.82	91	3.45	.83	148
Environmental Ratings Scales									
Infant/Toddler ERS-R*	4.17	1.05	18	4.86	.89	29	4.59	1.00	47
Early Childhood ERS-R**	4.39	1.19	32	4.97	.77	84	4.80	.93	116
School Age Care ERS-R**	3.83	.92	26	5.95	.56	6	4.23	1.20	32
Average of All ERS Scores**	4.22	1.09	57	5.03	.77	91	4.72	.99	148

Note: *p<.05, **p<.01.

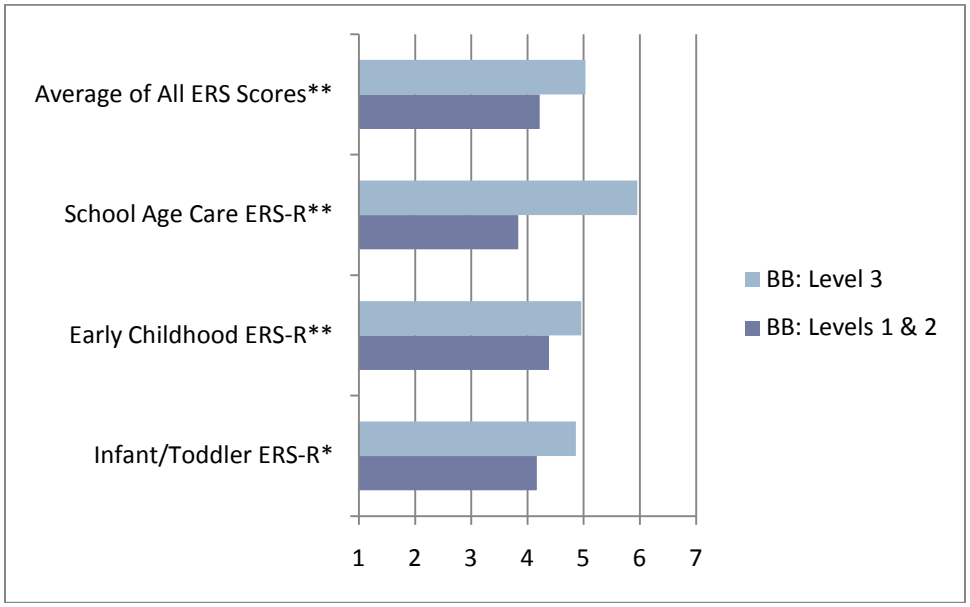
Table 4-C: Better Beginnings Levels on the Program Administration Scale and Caregiver Interaction Scores



Note: *p<.05, **p<.01

Table 4-D shows the cut scores relationship to ERS scores. The scores on the ERS range from 1 to 7, with 7 indicating higher quality. For ease of interpretation, the developers guide scoring of quality as 1=inadequate, 3=minimal, 5=good, and 7=excellent. Across each of the total ERS scores there were significant differences in favor of programs meeting Level 3 in BB. Scores for programs in the lower levels were, on average, between minimal and good. As discussed above, the programs recruited for the original study were mainly already in the state of Arkansas' Quality Approval system and were more representative of programs such as the federally funded Head Start and state-funded ABC programs. Therefore, the findings are hard to generalize to for-profit centers.

Table 4-D: Better Beginnings Levels on the Program Administration Scale and Environmental Rating Scale Scores



Note: *p<.05, **p<.01

Taken as a whole, BB Levels 1 and 2 do not require programs to attain a stated criterion score on the PAS, and analysis of the AECPDS demonstrates that programs scoring lower than 4 on the scale have teachers who are less sensitive, more detached, and less supportive of socio-emotional development. Further, ERS scores for those programs are all significantly lower than for programs achieving a minimum score of 4.

4.1.3 BUSINESS ADMINISTRATION SCALE (BAS)

Family child care providers who approach their program with intentionality and professionalism are more likely to provide high quality care supportive of child development. The BAS is designed to monitor and help improve business practices in family child care and is designed for tandem use with the *Family Child Care Environment Rating Scale, Revised (FCCERS-R; Harms, Cryer, & Clifford, 2007)*. The BAS contains 37 indicator strands clustered in 10 items scored on a 7-point Likert scale (1 indicating inadequate business practices up to 7 indicating excellent business practices), like PAS and FCCERS-R. Items include

Qualifications and Professional Development, Income and Benefits, Work Environment, Fiscal Management, Recordkeeping, Risk Management, Provider-Parent Communication, Community Resources, Marketing and Public Relations, and Provider as Employer (scored only if the provider employs other assistants).

The authors conducted reliability and validity tests for the BAS with 64 family child care providers in Illinois. They made refinements and tested again with 83 providers in four states that varied in regulatory stringency. The instrument demonstrated acceptable internal consistency (coefficient alpha .77 for the 10-item scale and .73 for the 9-item scale), distinctive but related subscales and items, and 94% inter-rater reliability among 21 assessors. Against the Parents and Provider subscale of FCCERS-R, BAS items were related but not redundant (.49, $p < .01$). In a subsample (n=33), full FCCERS-R observations were conducted and compared to the BAS. Programs that had lower FCCERS-R global quality scores also had lower BAS scores.

4.1.4 PAS/BAS CONFIRMATORY EVIDENCE

ECE research has yet to link the PAS and the BAS measures directly to child outcomes. These scales were based on research exploring how leadership and management functions affect global quality, staff continuity, and even teachers' interactions with children, all of which have been related to child outcomes in ECE literature.

Most notably, The National Child Care Staffing Study (NCCSS; Phillips, Howes, & Whitebook, 1991) investigated working conditions that predicted staff job satisfaction and 6-month turnover in EC programs via staff interviews and the Adult Needs subscale of ECERS, which contains items much like the Parents and Staff subscale in the current revised version. The team identified two factors of quality within ECERS/ITERS, appropriate caregiving and developmentally appropriate activities, but only reported on the developmentally appropriate activities factor, which examines characteristics such the availability of materials, scheduling, and the types of activities planned for children.

NCCSS analysis revealed that teachers who had higher wages provided higher quality care. "Paid preparation time was a consistent predictor of job satisfaction, and for teachers, good co-worker relations and advancement opportunities were negatively associated with turnover. Centers that provided for adult needs, such as offering opportunities for professional development and separate adult space, in addition to paying higher salaries, also offered higher quality care" (p. 67). For all staff only actual wages predicted turnover; staff perception of the fairness of their wages did not. A complex mix of characteristics influenced lead teacher turnover; consequently, wages accounted for 3% of the variation of turnover among all types of teachers. For assistant teachers, however, wage was the only predictor of 6-month turnover, calculated as 17% of the variation for that group.

A separate project examining staff intent to remain in the field (Holochwost, DeMott, Buell, Yannetta, & Amsden, 2009), rather than observed turnover, revealed salary to be less important. Among 846 educators, incentives most likely to promote staff retention were health insurance, disability insurance, and a pension plan. Unlike the pension plan, the availability of an investment plan was not a significant predictor of intent to stay, and the authors note that employees prefer defined, guaranteed benefits versus defined contributions.

Within BB scoring of the PAS, co-worker relations, advancement opportunities connected with salary increases and professional development opportunities will be scored. However, paid preparation time, associated with job satisfaction in NCCSS centers, and insurance benefits, associated with intent to remain in the field, will be omitted from the final score.

This year, the institution sponsoring PAS training, the McCormick Center for Early Leadership, reported on a new study assessing the quality of administrative practices in 138 Chicago Head Start centers (McCormick Center for Early Childhood Leadership, 2010). The study linked administrator professional qualifications and leadership characteristics measured by the PAS to ECERS-R scores. The mean PAS score was 3.42 (reflecting minimal quality) with administrative quality accounting for 26% of the variance in global quality ($t=3.62$, $p=.0001$). "Correlations between ECERS-R scores and dimensions of director qualifications suggest that higher classroom quality is associated with directors who had a B.A. degree or higher ($r = .218$, $p = .01$), had completed 24 or more hours of ECE coursework ($r = .192$, $p = .02$), and had made at least four professional contributions during the past three years ($r = .20$, $p = .018$). Lower classroom quality was associated with directors who had earned less than an A.A. degree ($r = .-189$, $p = .026$)" (p. 2). Directors with more formal education were more likely to hire teachers with higher levels of education. Head Start programs are already accountable to a comprehensive set of regulated professional standards and business practices, which in many areas meet or exceed PAS standards and certainly exceed common minimum licensing regulations. We would expect that general compliance with Head Start standards would dilute connections between PAS and global quality, so the variance in this study is a striking find. Again, this study identifies correlations between a center's global quality and PAS items that will be omitted from BB.

In addition to having some relationship with global quality, administrative quality may affect teacher-child interactions. One study confirmed what anecdotal experience often suggests: tension or lack of cooperation between administrators and teachers has spillover effects into the classroom. Mill & Romano-White (1999) found that teachers who felt they had authority to make decisions, were challenged by their jobs, and were supported by their administrator were less likely to express anger in the classroom. In this regard, making administrators accountable for excellent support systems may indirectly contribute to better child outcomes by reducing negative teacher-child interactions.

Within the family child care literature, the strongest evidence to corroborate use of the BAS in BB comes from the Family Child Care and Relative Care Study (FCCRC; S. Kontos, Howes, Shinn, & Galinsky, 1995). The team collected data from more than 800 family child care providers in Texas, North Carolina, and California. Regulated providers using a greater number of good business practices were more likely to provide higher quality care and had higher quality interactions with children than unregulated providers. The examined business practices included declaring income with the IRS, using parent contracts, keeping doctor's phone numbers and emergency authorizations, holding liability and property insurance, and maintaining records of immunizations. All of these items except the last are addressed in the BAS. The FCCRC team also observed children's social and cognitive development among regulated and unregulated providers. They found statistically significant differences in the amount of time children spent with peers and in the level of play in regulated versus non-regulated care. However, they found no significant differences in children's secure attachments, behavior problems, or language development across types of providers.

4.1.5 SCHOOL AGE: PAS OR YOUTH PQA?

The PAS was developed for early childhood programs, and we did not find any studies validating the PAS for use in school-age programs. As noted above, BB has already made the decision to exclude the child assessment portion of the PAS for school-age programs, as it would be redundant to the child assessments already completed during the school day, and it would be an additional financial burden. There is discussion about whether the High/Scope Youth Program Quality Assessment (Youth PQA), Form B (Adams, et al., 2005c) might be a suitable alternative.

Form B is consistent for both the Youth and Younger Youth versions, which are acceptable measures of the environment in BB. It consists of three subscales: Youth-Centered Policies and Practices, High Expectations for Youth and Staff, and Access. (See the Environmental Assessment section 7.4.2 for detailed information about the Youth PQA.) This is a relatively new instrument, and findings from the only significant validation study to date (Smith & Hohmann, 2005) identified problems with inter-rater reliability and the level of internal consistency (Cronbach's alpha .54).

There are some notable difference between the PAS and Form B. PAS requires both interview and document review, while Form B requires only the administrator interview. PAS indicators are typically more specific and structural. For instance, Form B requires pre-service orientation, while PAS provides explicit criteria for the content and structure of new employee orientation. Appealing aspects of the Youth PQA for BB designers is involvement of youth in all aspects of program planning and its flexible application to a wide variety of school-age programs.[22]

The Youth PQA has been piloted in 30 central Arkansas after-school programs by the Arkansas Out-of-School Network. In addition, a workgroup led by ASU and comprised of school-age specialists from around the state is working to examine Part B in light of structural components that might be added to the instrument, including components of PAS. A revised version of Part B may serve as a practical alternative to PAS, particularly for those programs using Youth PQA in lieu of the ERS. However, unvalidated instruments are not recommended for high-stakes systems, such as the QRIS. We would urge formal testing and validation of any revisions prior to BB inclusion.

4.1.6 STATES COMPARISON

Comparing state quality rating systems provides no additional confirmatory evidence for either the PAS or the BAS. Of the five key states highlighted in this review, Ohio is the only one using the PAS in its quality rating system. All levels of the Ohio Step Up to Quality (SUTQ) system require completion of an action plan based on PAS self-assessment results, but no formal assessments or minimum scores are required for entry or advancement across levels. The SUTQ evaluation (Ohio Collaborative, 2009) did not include the PAS as a variable so we do not know what influence it may have had on global quality in Ohio.

[22] Noted by Diana Courson, Arkansas State University Childhood Services, in an unpublished comparison of PAS, Better Beginnings (S-A), and YPQA Form B.

Looking beyond the key states, Illinois is the only state with a quality rating system incorporating the PAS. Illinois calls for PAS training for centers applying for a 3- or 4-star rating. To receive a 3-star rating, programs may either receive a PAS score of 4.25 or be nationally accredited. A PAS score of 5 and national accreditation is required to earn a 4-star rating. The state has not conducted an evaluation of its rating system.

The BAS was published even more recently than the PAS, and Illinois is the only state using it in a state quality rating system, requiring either a score of 4.25 on the BAS or NAFCC accreditation for a 3-star rating. For 4 stars providers must score 5.0 on the BAS score and be NAFCC-accredited. Although they are not using the BAS, most states do recognize the importance of efficient, professional practices in family care. Missouri, North Carolina, Ohio, Pennsylvania all include business standards within their rating systems.

In a recent review of the use of program quality measures in states with school-age QRIS, 16 states were found to identify and require program quality measures. This group included 12 fully implemented state programs, 3 pilot state programs, and 1 proposed program (Arkansas). The SACERS is the most commonly used instrument; only one state (Ohio) requires the PAS, and they allow self-assessment.

4.2 CROSSWALKS

4.2.1 HEAD START

There are areas of considerable overlap between the Head Start Act and Head Start Performance Standards and items in the Administration section of BB. A crosswalk between Head Start and PAS shows that Head Start requirements meet or exceed the scope of the following PAS items:

- Staff Development
- Screening Identification of Special Needs
- Assessment in Support of Learning
- Budget Planning
- Accounting Practices
- Program Evaluation
- Strategic Planning
- Family Support and Involvement

In the event that BB administrators need to reduce assessment research and evaluation costs, one possible route would be to excuse a Head Start program on these PAS items when results from the program's three-year evaluation can be obtained and alignment with Head Start performance standards verified.

4.2.2 NAEYC

NAEYC Leadership and Management standards cover in greater detail the same areas as the PAS. Centers with high PAS scores will have laid a decent groundwork to begin advancing toward NAEYC's standards.

Standard 10, Leadership and Management, is the section most closely aligned with PAS items. The contents of this standard are summarized:

- **10.A Leadership** outlines education qualifications for the administrator.
- **10.B Management Policies and Procedures** details the use of written policies, insurance requirements, strategic planning processes, orientation processes, provisions for working with consultants, and policies addressing ratio and group size staffing.
- **10.C Fiscal Accountability Policies and Procedures** requires accounting processes and operating budgets.
- **10.D Health, Nutrition, and Safety Policies and Procedures** designates a host of written policies that must be provided. Examples are management plans to prevent infectious disease and child abuse and neglect, arrival and departure procedures, emergency procedures, and handling medications.
- **10.E. Personnel Policies** requires written policies defining job responsibilities, hiring and evaluation practices, and benefits. For example, employees must participate in health assessments. Programs must provide benefits such as health insurance, vacation, and retirement.
- **10.F. Program Evaluation, Accountability, and Continuous Improvement** requires an annual comprehensive evaluation of the program that receives input from staff and families.

4.2.3 NAFCC

Although there is some alignment in business practices and family support, NAFCC accreditation standards are less specific and comprehensive than the BAS.

Like BAS, NAFCC has standards to address record-keeping (5.21-5.25), some risk management measures, for example, required contents of a First Aid Kit required (4.10), and the maintenance of telephone and emergency phone numbers (4.11). NAFCC is also similar to BAS in the areas of the enrollment process (5.14), contracts with families (5.16), receipts for parents (5.17), accident reports (5.18), written policies that must be give to parents (5.19), and required insurance coverage (5.20). Unlike BAS, NAFCC does not suggest the number of modes of communication that must be used.

In terms of provider qualifications, BAS exceeds NAFCC standards. Whereas BAS gives quality points for college credits and attaining specific numbers of clock hours in annual professional development, NAFCC providers need only to show evidence that they seek continuing education and training (5.6).

4.2.4 CARF

The CARF standards include substantial areas of overlap with PAS. CARF standards meet, partially meet, or exceed all PAS standards considered mandatory for BB Level 3.

4.2.5 COA

COA after-school programming standards overlap PAS substantially; however, COA is less detailed. For example, orientation is required, but content is not specified, and annual performance evaluation guidelines do not specify that classroom observation is required. There is no discussion of space for staff, written minutes, specific assessment tools for evaluation, or the provision of daily conferences with

parents. With COA, guidelines are broad, and specific programs have some leeway as to how they comply with the standards but must provide documentation of compliance.

COA standards partially meet or fully meet many of the PAS criteria in the following sections:

- Fiscal Management
- Program Planning and Evaluation
- Marketing and Public Relations

In the event that BB administrators need to reduce assessment costs, one possible route would be to excuse a COA-accredited program on these PAS sections when results from the program's accreditation can be obtained and alignment with COA standards verified.

4.3 STRENGTHENING FAMILIES

The second component of the Administration section of BB requires program leaders to learn about and take actions intended to reduce child abuse and neglect. This is done through the Strengthening Families self- assessment tool and strategy developed by the Center for Study of Social Policy (2008b) and through an online tool to train child care staff in Arkansas. Strengthening Families (SF) is a national initiative to equip early childhood programs with knowledge and practices to prevent child abuse and neglect. The initiative's logic model was based on research highlighting five protective factors in families that correlate with greater child protection and observations of model child care programs. The goal is to implement strategies used by model child care programs to enhance these factors that provide protection for children:

- Parental resilience
- Social connections
- Knowledge of parenting and child development
- Concrete supports in times of need
- Children's social and emotional development

The vast majority of items addressed in the self-assessment are focused on organizational policies, parent training, and communication with parents. These things generally occur outside of the scope of classroom interactions with children. The SF self-assessment is best classified as a structural measure because these types of processes occur apart from interactions directly involving children in the center or in the family day care environment. [23]

Within BB, there are no SF requirements for Level 1. Administrators can achieve credit for Level 2 by completing a webinar on SF. Self-assessment, action planning, and one action step are required at Level 3. There are no provisions regarding the scope or quality of the action step, so a program that completes an

[23] The section, Children's Social and Emotional Development, does contain a few items that instruct the type and quality of interactions between teachers and children. We ruled out a classification of SF as a global measure because these child interactions items are outnumbered by structural items not only in the grand scheme of the self-assessment, but also within that particular section.

action step unlikely to affect child outcomes could potentially receive the same "quality credit" as another program choosing a more salient action. For instance, in terms of probable outcomes, having a comfortable space for families to meet informally (item 1 in Facilitate Friendships and Mutual Support) would be much less likely to influence positive parent or child outcomes than "teaching about children's social and emotional development in parenting classes and informal discussions" (item 3f in Facilitate Children's Social and Emotional Development).

The SF logic and content are based on empirical research pointing to protective factors that lead to better child outcomes. A review of this literature is available on the SF website (Horton, 2003). [24] Although non-maternal care can provide modest support for a child's well-being, the combination of parent and familial characteristics such as parents' income, mental health, education, cultural beliefs and quality of caregiving have far greater influence on the trajectory of a child's development (Klebanov & Brooks-Gunn, 1998; Linver, Brooks-Gunn, & Kohen, 2002; NICHD Early Child Care Research Network, 2001, 2002; Yeung, Linver, & Brooks–Gunn, 2002). Traditional quality measures address the cognitive, social-emotional, and physical wellness of children, but have not yet gone far enough in how the parents' wellness in these same domains will shape the child's development. SF is intended to bridge this gap.

4.3.1 EVIDENCE FOR ECE

The SF assessment tools have not been tested for reliability or validity, and none of the programs that have implemented the strategy have examined change in abuse and neglect rates or for other child outcomes. Because it is prohibitively expensive for most states and programs to conduct experimental or quasi-experimental evaluations of child outcomes, those that have adopted the formal SF model and tools report only on parent and provider self-reported attitudes. For instance, family support programs in New Hampshire adopted the SF strategy through a curriculum that trained staff to identify and provide support for families at risk. After implementation, parent surveys from 10 programs indicated improvements in parents' self-reported attitudes and confidence (Brandt, 2005). Centers also reported improvements in supportive relationships, accessing community resources, parental confidence, sharing parental concerns, standing up for family needs, and reduced family stress. Meeting Family Needs (centers providing concrete support in time of need) showed the lowest percentage of improvement. The Center for the Study of Social Policy also describes improvements in self-assessments in Alaska and Wisconsin (2008a).

There is only one published peer-reviewed evaluation of a state's SF approach in early childhood programs (Roach, Kim, O'Connor, & Laurion, 2009). Wisconsin introduced the model to more than 2,000 child care providers from 812 programs via two-hour community trainings. Centers completed the SF self-assessment and then were able to apply for mini-grants out of a pool of $35,000 from the state, and $43,700 from local agencies was distributed to centers to improve areas of weakness identified in self-assessments. Of 56 programs that completed pre- and post-self-assessments, NAEYC-accredited programs and mainstream programs combined (n=38) reported improvement in facilitating friendships and mutual support, strengthening parenting, responding to family crises, and linking families to services and opportunities. Head Start programs (n=18) only showed significant improvement in valuing and supporting parents, but no significant change in six other areas: facilitating friendships and mutual

[24] http://strengtheningfamilies.net/images/uploads/pdf_uploads/LiteratureReview.pdf. Accessed 08/03/10.

support, strengthening parenting, responding to family crises, linking families to services and opportunities, facilitating children's social and emotional development, and observing and responding to early warning signs of child abuse/neglect. Head Start regulation requires careful documentation and observation of these strategies, so the majority of Head Start programs were already able to report "Excellent" performance on 6 out of 7 categories at pre-assessment. This was not true of NAEYC centers where self-reported scores were much lower than Head Start at pre- and post-assessments. Parent and caregiver surveys are an inexpensive method to track changes in attitude, but more research incorporating observations of teacher, parent and child behaviors and verification of child outcomes are needed to empirically validate the SF model.

Because the literature related to the SF model was slim, we widened our scope and reviewed other programs and lines of study that might shed light on whether the approach would improve child outcomes. We selected large-scale and model ECE programs, nurse home visitation programs, and recently piloted programs in parent support and child mental health. Additionally, we found research on helpgiving practices to be helpful. We will summarize findings related to each area.

For young children, there is substantial correlative evidence that programs which involve and support parents in a culturally sensitive way produce more potent positive long-term effects than those isolating care to the child (Bronfenbrenner, 1974; Kumpfer, Alvarado, Smith, & Bellamy, 2002; A. J. Reynolds & Robertson, 2003; Shonkoff & Phillips, 2000). Frede observed that options for parent involvement varied widely by center auspice (1995). Head Start programs were highly likely to offer opportunities for parent volunteering, parent workshops, and home visits. On the other hand, the percentage of for-profit centers offering these opportunities was 12%, 23%, and 7% respectively.

EC experts widely agree on the risk factors that contribute to child abuse and neglect, but little is known about the most effective and practical means to incorporate measures of prevention of child abuse and neglect (PCAN) into large-scale ECE programs. While parent involvement is an integral part of Head Start programs, it is a neglected area of Head Start research (M. Cochran, 2007; Zigler & Muenchow, 1992). Generally, examinations of ECE programs with a parent involvement component monitor other outcomes besides abuse and neglect, most commonly academic success (John M. Love, et al., 2004) but also juvenile delinquency and substance abuse (Zigler, Taussig, & Black, 1992). Abuse and neglect cases are under-reported and difficult and expensive to track, so it is rare to find an experimental or quasi-experimental study of family involvement in ECE programs that includes analysis of documented cases of abuse and neglect.

An exception is a study of the Title 1 Child-Parent Centers (CPC) in Chicago (A. J. Reynolds & Robertson, 2003). Children who had participated in the preschool treatment group (n=913) had 52% fewer court petitions for child abuse than control group children (n=495) through age 17. The study identified two statistically significant mediators for PCAN: 1) fewer moves from one school to another and 2) parent involvement, which included parent volunteering, parenting classes, home visits, and other methods to increase parent participation. Many children in the control group attended Head Start or other intervention programs, so effects would probably be even greater if compared to children who do not participate in any kind of intervention. Duration and timing were influential variables. Children who were in the program more than one year experienced greater protection. Children who participated in the school-age CPC did not experience reduced maltreatment. Early Head Start programs for children birth to 3 years likewise bear out improvements in parent and child outcomes with earlier enrollment (J.M. Love et al., 2002; U.S. Department of Health and Human Services, 2001a).

Interventions documenting meaningful, lasting changes have been costly. For instance, CPCs employed a parent resource teacher, and all teachers in the program had BAs and certificates in early childhood education. Comparing combination ECE/family support programs that produced long-term positive effects on delinquency and anti-social behaviors, Yoshikawa (1995) identified common characteristics of effective interventions. They are intensive, for instance conducting weekly or monthly home visits, and they are high quality, using curriculum with a strong theoretical base, maintaining teacher-child ratios of 1:3 for infant/toddlers and 1:6 for preschoolers, and employing staff with extensive training.

Outside of ECE programs, home visitation programs have produced more results with PCAN that other methods, although their success is still inconsistent. Researchers find that intensity and duration is as important to ECE programs as it is to home visitation programs seeking to improve parent-child outcomes (Gomby, Culross, & Behrman, 1999). Old's Nurse Home Visitation Program (NHVP) has been successful at reducing abuse and neglect, even though PCAN was not the original goal for the program. Participants in these programs access more services, have more positive parent/child interactions, abuse less, are less likely to remain in welfare, and are more likely to get more education and take on stable jobs. However, the programs experienced more success when operated by highly trained professional staff and offered intensively and consistently. In the Elmira, New York trial, low-income mothers were visited 9 times during pregnancy and 23 times after their children were born. Nurses addressed health behaviors, care for children, family planning, general education, and work force issues. At a 15-year follow-up, the frequency of verified reports of abuse and neglect perpetrated by low-income, single mothers was .29 for treatment group compared to .54 ($p<.001$) for the control group (David L. Olds et al., 1997). Effects failed to reach significance for other groups of women. In a controlled, randomized study, similar results were not replicated when visits were conducted by paraprofessionals rather than nurses. Paraprofessional visits did result in small positive effects on mother-child interactions and on the caregiving environment, but these changes did not translate to less domestic violence or to improved child outcomes (D. L. Olds et al., 2004).

A Centers for Disease Control and Prevention meta-analytical study of 22 visitation programs (2003), including the Old's programs, found that home visitation programs "delivered by nurses demonstrated a median reduction in child abuse of 48.7% (interquartile range: 24.6% - 89.0%); programs delivered by mental health workers demonstrated a median reduction in child abuse of 44.5% (interquartile range not calculable). For paraprofessional visitors, effects were mixed: the median reduction in child abuse was 17.7%, but the variability of the findings is reflected in the interquartile range of -41.2% - 65.7%. In programs using paraprofessionals, beneficial effects were consistently evident only when programs were carried out for $\geq$2 years…Evidence from the single study of the effects of home visitation on partner violence indicated that home visitation might not prevent child maltreatment in the presence of ongoing partner violence."[25] Many of the studies included for review were not randomized and did not measure abuse outcomes, rather they measured other outcomes that may not be adequate proxies for abuse and neglect, such as hospitalization (Chaffin, 2004).

Unfortunately, the majority of home visitation programs have failed to reduce abuse and neglect or to conclusively show positive effects on child outcomes (Duggan et al., 2004; Fraser, Armstrong, Morris, & Dadds, 2000; St. Pierre & Layzer, 1999). We do not have sufficient evidence that home visitation programs

[25] Cited from http://www.cdc.gov/mmwr/preview/mmwrhtml/rr5214a1.htm. Accessed 9/10/2010.

are capable of addressing more complex serious issues like substance abuse and domestic violence, which are proximal risk factors for child abuse and neglect (Centers for Disease Control and Prevention, 2003; Daro & Donnelly, 2002). For example, the goal of The Hawaii Healthy Start Program (Duggan, et al., 2004), was to reduce abuse by training parents to be more accepting of child behavior and to use nonviolent discipline, but the program had no impact on participants in these areas. The home visitors, paraprofessionals with at least a high school diploma and five weeks of training, were not able to detect or report proximal risk factors for neglect and abuse in homes they visited.

An alternative model of family involvement is being tested locally in ECE programs. Teaching Important Parenting Skills: TIPS for Great Kids! (TIPS) trains ECE teachers to support families and to improve child outcomes via program handouts containing parenting and family tips and guidance cards for teachers. Teachers in the North Little Rock School District Early Childhood Program (NLRSD ECP) participated in one day of training per month and in two pre-service days on helpgiving, on the TIPS model, and on tips delivery. Trainers provided instruction in many of the TIPS topics, with special attention devoted to those that were sensitive in nature, such as divorce and trauma to the child. Preliminary data from this program (P. Bokony, McKelvey, & Patrick, 2009) indicate that teachers were trained on helpgiving and supported with tools to elicit families' needs and information to provide support, teachers were the most comfortable with and provided the greatest number of tips that were more closely aligned with their role as teachers – those related to child development and school readiness outcomes. Parents were unlikely to express the need for help in the areas of family stress, mental health, and substance abuse. Less than half of the families that reported needing help in these three areas on parent surveys received help from their teachers in these areas. TIPS has not documented child abuse and neglect reports, but parents reported improvements such as increased communication and fewer battles over routines (McKelvey, Bokony, & Patrick, 2010).

Another model that incorporates family support is early childhood mental health (ECMH) consultation. The ECMH consultant is a professional with expertise in mental health and child development who builds a collaborative relationship with caregivers, child care center directors, other providers, and parents. The consultant typically works to assist teachers with children who have challenging behaviors, focusing on the classroom environment as well as on the individual child and his or her family. The consultant often serves as a link between centers and families and makes referrals to outside resources when needed. A study of six effective early childhood consultation programs with demonstrated positive outcomes identified key skills of consultants, including the ability to "link children/families/providers to other services and systems as needed" (Duran et al., 2009). Seventy-two percent of consultants in these model sites reported making referrals to other community supports and services for families at least monthly, and they reported following up to assist families in accessing the needed services.

In 2005, the Arkansas Early Childhood Mental Health Consultation Pilot Project began a partnership with three community mental health centers to provide consultation services in early childhood centers around the state. Teachers participating in focus groups reported that referrals made by the consultants were very helpful. They also said the consultant helped them see that what was going on at home had an impact on a child's behavior at the center (Conners-Burrow, McKelvey, Amini-Vermani, & Sockwell, In review).

Recent studies suggest that ECMH consultation can help improve children's classroom behaviors, reduce rates of preschool expulsion, and help teachers create a more positive classroom climate (Alkon, Ramley, & MacLennan, 2003; Brennan, Bradley, Allen, & Perry, 2008; Green, Everhart, Gordon, & Garcia Gettman,

2006; D. Perry, Dunne, McFadden, & Campbell, 2008; Raver et al., 2008). Similar results from the Arkansas ECMH study found improvements in positive behaviors and reductions in problem behaviors for children and improvements in sensitive behaviors and reductions in permissive and detached behaviors for teachers (Conners-Burrow, et al., In review).

Other investigations into general family-centered helpgiving practices back up these pilot program findings and are relevant to BB. In a meta-analysis of 19 studies of child programs, outcomes that had the strongest associations with family-centered practices were parent self-efficacy beliefs, satisfaction with staff and services, parenting capabilities, and child behavior and functioning. Child development outcomes were not directly affected by family helpgiving but were indirectly mediated by self-efficacy beliefs (C. J. Dunst, Trivette, & Hamby, 2006).

A larger, follow-up meta-analysis by the same team (C. J. Dunst, Trivette, & Hamby, 2007) reviewed family-centered practices and outcomes in multiple fields, including early childhood education. The team classified two types of helpgiving: relational and participatory. The former includes listening and demonstrating respect and empathy for the family and believing in family strengths. The latter is "individualized, flexible and responsive to family concerns and priorities" and involves "informed choices and family involvement in achieving desired goals and outcomes" (p. 370). More providers are better at relational than participatory helpgiving, which occurs less often (C. Dunst & Trivette, 2005) but is a stronger predictor of family functions and behaviors of parents and children and outcomes mediated by self-efficacy beliefs (C. J. Dunst, et al., 2006). In both meta-analyses, parents' judgments regarding children's behavior were affected positively by strengths-based family helpgiving practices.

Of course, a typical stand-alone childcare program in Arkansas will not have the financial means to implement family support models like the NHVP or CPC. Given the magnitude of child abuse and neglect, mandates for widespread institutional shifts to protect the well-being of children and families are vitally needed, but there will likely be pitfalls along the way, especially with programs of limited scope, or without trained professionals. Even when specialized professionals facilitate programs, there may be unintended negative outcomes. For instance, one studied parent support group for mothers of infants with disabilities increased participants' stress (Krauss, Upshur, Shonkoff, & Hauser-Cram, 1993).

4.3.2 EVIDENCE FOR AFTER-SCHOOL PROGRAMS

After-school programs that facilitate increased parental involvement may see improved outcomes for their school-aged participants. However, apart from the CPC program, we did not find research examining PCAN as an outcome for quality after-school services.

Parent involvement with school makes a difference in the academic outcomes of children. When parents are involved, children across all ethnic and socioeconomic boundaries benefit from higher grades and test scores, better school attendance, more time spent on homework, reduced drop-out rates, increased rates of college attendance, and improved behavior and social skills (Alexander & Entwisle, 1988; Comer, 1984; J. L. Epstein, 1996; Henderson & Berla, 1994; Henderson & Mapp, 2002; Keith et al., 1993; Miller, 2003; Putnam, 2000; Steinberg, Brown, & Dornbusch, 1996).

Does parental involvement make a difference to the *quality* of after-school programming? The after-school literature supports the belief that it does. For instance, the Massachusetts After-School Research Study (MARS) developed a list reflecting "growing consensus on the ingredients of successful programs"

and included parental involvement as one of those core ingredients (Miller, 2003). *The researchers noted that larger programs were less likely to maintain communication with parents, and also included "low adult to youth ratios" as a requirement for successful programs.* A RAND study was designed to determine whether California's Stone Soup Child Care Programs were following best practice in their after-school programming. In an extensive literature review they identified 15 quality indicators, including parent involvement (Beckett, Hawken, & Jacknowitz, 2001).

Some of the benefits of parental involvement with after-school staff may be based on the improved connection with school. Parents often have more contact with after-school staff than with their children's teachers. After-school programs can encourage positive communication between schools and parents by helping parents understand a school's expectations, curriculum, and culture (Fiester, White, Reisner, & Castle, 2001; Miller, 2003; Noam, Biancarosa, & Dechausay, 2001).

Of the 96 schools in The After-School Corporation (TASC) partnership in New York, 79% of school principals said that parents expressed more positive feelings about the school because of their child's involvement in the after-school program. Principals reported a 31% increase in parents' attendance at school events, and a 15% increase in parents' attendance at parent-teacher conferences (Reisner, White, Russell, & Birmingham, 2004). TASC participants in the elementary and middle grades showed improved performance in mathematics and increased their school attendance. Although parental support and involvement was required for TASC programs, in an analysis of what program characteristics were linked with student gains, these elements were not mentioned (Reisner, et al., 2004).

Clearly children benefit when their parents are involved in their school and after-school activities, and programs that encourage parental involvement see results. However, outcomes for providing support to families of school children have not been well explored. The TASC evaluation did find, when they looked at the ten highest-performing programs in the TASC network, that "intentional relationship-building," which included "regular communication with and the provision of support services to families," was one of five characteristics shared by all of those high quality programs (Birmingham, Pechman, Russell, & Mielke, 2005). The SF approach to providing family support would include building close relationships with parents and connecting parents to appropriate resources when needed.

4.3.2.1 STATES COMPARISON

No other quality rating system in the U.S. incorporates the SF self-assessments or training tools. Colorado, Missouri, Oklahoma, and Pennsylvania include a family involvement or partnership component that address some of the same standards included in the SF strategy to improve family-provider partnerships and to foster good parenting and increased protection for children. We provide each of the comparison state requirements for the family components below. Similar to Arkansas, within some of these items there is a choice for the means and content of communication and/or training for parents. In contrast with Arkansas, centers within these three states must demonstrate adoption of parent involvement procedures in lower levels and do not offer credit for self-assessment.

4.3.2.2 COLORADO

Initially Colorado used parent satisfaction surveys in its family component but then found the surveys biased toward positivity and moved to a different partnership measure (G Zellman & Perlman, 2006; G. Zellman & Perlman, 2008). The current Family Partnerships category receives 1-10 points based on parent

questionnaires and center documentation. Criteria include provision of information regarding the program, parenting and child development, community resources, and individual child's activities and progress. Centers must also demonstrate that parents participate in planning and decision-making from the child level up to the program level.

4.3.2.3 MISSOURI

Specific methods of communication and cooperation with families are required to move beyond minimum licensing, Tier 1. Tier 5 centers must be accredited.

- **Tier 2** centers must offer one communication method for parents, one family educational workshop and one social event per year, and family volunteer opportunities.
- **Tier 3** centers must offer monthly child-specific written communication, have a communication center, have three communication methods, offer two family educational workshops and two social events per year, have one family-teacher conference, a family resource center, and offer one of the following: a home visit, a family needs assessment, a family advisory board, or family support groups.
- **Tier 4** centers must offer weekly child-specific written communication, three communication methods, three family educational workshops per year, two family-teacher conferences, a family advisory board, and a family resource center with three required items. They must also offer at least one of the following: home visits, family needs assessment, or family support groups.

Although Missouri did not isolate the effects of its Family Involvement component, confirmation of relationships between their system tiers and child outcomes suggests that it is a meaningful component to child development.

4.3.2.4 OHIO

Ohio's system was originally designed with a family involvement component, but it was later removed because it failed to detect meaningful information. Currently, the only related items are sharing child screening and assessment results with families as part of the Early Learning category for **Step 2** and **Step 3** providers.

4.3.2.5 OKLAHOMA

- **Centers at all levels** must have a system for sharing and communicating with parents issues related to the child's emotional and physical state, welcome parents in the child's class at all times, document annual conferences, hold at least two meetings or special events for parents each year, use two methods of communication, have parent participation in the board or in planning meetings or yearly questionnaires, and make licensing requirements available to parents.
- **For levels 2 and 3,** centers must also give parents a written report about the child annually, maintain a list of available community resources, and assist parents in locating and connecting to services.

4.3.2.6 NORTH CAROLINA

North Carolina programs can earn one quality point by having evidence of an infrastructure of parent involvement that includes at least two of the following elements: parent newsletters offered at least quarterly, a parent advisory board, periodic conferences for all children, or parent information meetings offered at least quarterly.

4.3.2.7 PENNSYLVANIA

- **Star 1** centers are required to give parents information about public, social, and community services. Upon enrollment parents are offered a "Getting to Know You" meeting within 60 days of enrollment.
- **Star 2** centers give and explain written information on health and human services, wellness, nutrition and fitness, and/or development to staff and parents. Copies of children's IEPs or IFSPs, written plans, and/or special needs assessments are completed by professionals to inform classroom practice. Individual information is shared in written form daily; specific group information is shared daily using a visual communication format. One parent conference is offered annually.
- **Star 3** centers have a written and implemented plan to refer parents to appropriate social, mental health, educational, wellness, and medical services. One annual group activity must involve parents in meeting program learning goals. At least two parent conferences are offered with a written report required for one.
- **Star 4** centers, if applicable, in conjunction with parent and service providers, implement activities to meet IEP or IFSP goals or special needs plans and objectives. Programs must have policies that demonstrate engagement and partnership with parents in program planning and decision making.

4.4 CROSSWALKS

4.4.1 HEAD START

The current paradigm for best practices in ECE parent involvement began with Head Start (M Cochran, 1992). We see evidence of this in our crosswalks in that Head Start standards align with the SF scope of family supports more closely than NAEYC. Head Start parental involvement is typically carried out through one of three ways: volunteering in the classroom, serving on the Policy Council, or attending parent education programs and meetings (M. Cochran, 2007). Head Start standards

- require staff to identify family strengths and set goals, provide emergency resources, make referrals for mental health or employment services, follow up to ensure services were provided, and establish procedures to make parents aware of a comprehensive list of community resources (45 CFR 1304.40),

- call for collaborative partnerships with community organizations (45 CFR 1304.41),

- and demand that all staff receive training in "methods for identifying and reporting child abuse and neglect that comply with applicable State and local laws using, so far as possible, a helpful rather than a punitive attitude toward abusing or neglecting parents and other caretakers." (45 CFR 304.531). [26]

4.4.2 NAEYC

Standard 7, Families, and Standard 8, Community Relationships, have the most overlap with the SF approach. These NAEYC topic areas highlight the need for communication with families and their involvement in program planning and in meeting children's individual needs. NAEYC standards do very little to address family crisis management in comparison to SF model.

4.4.3 NAFCC

NAFCC standards are less comprehensive than SF. NAFCC providers must share information about common child-rearing issues with parents (5.10), detect and report possible child abuse and neglect (5.11), help families access community and medical services (5.11), incorporate a family's cultural traditions into the program (1.6), be alert and responsive to children's needs (1.5), and encourage families to visit or to communicate with the provider (1.9). Standards 1:14-1.19 address positive communication with families. Also, providers are expected to demonstrate ties to social supports (1.28).

4.4.4 COA

Within COA after-school programming standards, section ASP-PS 10 covers family connections with general guidelines for treating family members with respect, helping them to feel welcome, offering orientation sessions and information about the program, encouraging family members to visit and give input, sharing information about resources and services that can help meet families' needs, and working together with families so that arrivals and departures are smooth. There is no discussion of family crisis management. ASP-HR 3.05 requires that staff receive training on mandatory reporting and indicators of suspected abuse and neglect. COA-accredited programs must ensure and provide documentation that all of the general guidelines are met, but how the guidelines are met may vary with the program, so there are fewer specifics than in SF.

4.4.5 CARF

The overall CARF emphasis on family involvement and partnership fits well with the SF approach. A CARF-accredited program would have policies and procedures in place to ensure that *child and family-centered care* was provided, and would be held accountable for implementation of those policies and procedures by on-site review. However, CARF does not include detailed descriptions of program design and implementation specific to child care settings, as does SF. An individual program chooses how it will carry out the broad CARF objectives.

[26] Head Start standards available at http://eclkc.ohs.acf.hhs.gov/hslc.

4.4.6 PAS

There is some redundancy between the PAS and SF, the two components of the Administration section of BB, and between ECERS-R. Most of the overlap occurs with PAS section 17, Family Support and Involvement. Within the PAS, the more family supports implemented, the more points earned for that section. Almost all of the forms of support suggested by PAS are listed at some point in SF, but whereas the PAS suggests the mode of communication or participation, SF goes much further by also specifying the agenda and tone. SF items could be more meaningful to child outcomes because they suggest agendas and sensitive behaviors, but we have no formal evidence to prove this. SF is not a formal measure and there little accountability in the system for applying or measuring standards in a comprehensive programmatic way.

4.5 CONCLUSIONS

The PAS and the BAS are reliable, valid instruments that approach early childhood program quality from a different angle than environmental rating scales. Multiple measures will generate a more comprehensive and refined picture of the overall health the program. The PAS and the BAS will be useful in highlighting strengths and weaknesses of business functions and steer administrators toward changes that may benefit staff, parents, and children.

Rationale behind these administrative scales is strong, but the instruments are new. To date, PAS and BAS scores have not been analyzed alongside child outcome variables either in general research or in evaluations of state quality rating systems. In centers, there is limited evidence that administrative support moderates teacher-child interactions. In family child care, regulatory status, linked to provider adherence to good business practices, showed a relationship to time spent with peers and in level of play but not to other cognitive and social outcomes. There is evidence that strong leadership and well-informed administrative practices contribute to the global quality of a program, which in turn supports child development. This evidence also suggests that PAS items omitted from Better Beginnings may be important to the measure's validity in Arkansas centers.

In the interest of encouraging wider Arkansas center participation in quality improvement, Better Beginnings excludes items 22-25 that address administrator and teacher qualifications from the assessment process. Also, items 5 and 6 that rate staff benefits and staffing patterns and scheduling will be scored to advise program improvements, but these items will not be counted in the program's overall score. Findings from our validation efforts suggest that elements of the PAS that are not being scored may relate to high quality supports for children in the classroom. Given the evidence that teacher education is related to more optimal classroom practices and the validity of the instrument may be detrimentally impacted, we recommend re-introducing the items that have been excluded from Better Beginnings.

Better Beginnings Levels 1 and 2 do not require programs to attain a stated score and analysis of the Arkansas Early Childhood Professional Development System demonstrates that programs scoring lower than 4 on the scale have teachers who are less sensitive, more detached, and less supportive of socio-

emotional development. Further, ERS scores for those programs scoring lower than 4 on the scale are all significantly lower than for those that achieve at least a minimum score of 4.

Research related to Strengthening Families suggests that sustained increases in protection for children are likely if parent involvement and supports are comprehensive, intensive, sustained, and also combined with non-maternal care that is comprehensive, sustained, and of exceptional quality, i.e., quality that exceeds Better Beginnings Level 3. We would expect use of the Strengthening Families model to heighten awareness and to improve *relational* helpgiving skills if all staff members, not just administrators, receive more intensive training than currently designated. It is unlikely that the webinar and self assessment for administrators will increase the amount and quality of *participatory* helpgiving, which is more tightly linked to change in family functions and behaviors. Empirical evidence suggests that webinar training, self-assessment and arbitrary adoption of one or even a few of the Strengthening Families strategies is unlikely to produce detectable significant changes in child abuse and neglect. If administrators extend their training to teachers and adopt the Strengthening Families practices comprehensively, there are likely to be improvements in parental understanding of child development and parenting behaviors.

5 QUALIFICATIONS AND PROFESSIONAL DEVELOPMENT

Within the past decade, a movement calling for a minimum of a bachelor's in ECE classrooms has gained considerable traction. Reviews by prominent committees and researchers support policies increasing teacher education requirements (Barnett, 2003; National Research Council, 2001; Whitebook, 2003). By the year 2013, 50% of all Head Start teachers must have a BA. State pre-kindergartens are increasingly folded into the public school system, where minimum requirements for BAs are the norm and preschool teachers are paid on the same scale as all other teachers. Community child care centers participating in the NAEYC accreditation process will also be affected by this shift. NAEYC's current official position statement on professional development acknowledges great diversity among members of the ECE field and supports a varied approach to professional development (National Association for the Education of Young Children, 1993), but accreditation standards emphasize the importance of specialized college preparation. Similar to Head Start, NAEYC is on a timeline to increase teacher qualifications so that by the year 2020 all accredited classrooms will have a teacher with a BA.[27]

This widespread appeal for degrees represents a general policy shift in the ECE field from one that emphasized in-service training and annual clock hours to one that favors pre-service training. Advocates of this shift point to evidence that college education focused on ECE or child development improves classroom quality, to evidence linking teacher education to child outcomes, and to evidence that teacher education is a better predictor of quality than years of experience (Carolee Howes, Whitebook, & Phillips, 1992; Snider & Fu, 1990; Nicholas Zill et al., 2001). Many analyses show that BA-level teachers have the most optimal interactions with children. Advocates also contend that increasing educational requirements will make ECE a professional field, increase wages for currently underpaid employees, and reduce the very high turnover rates that negatively affect child attachments.

Arkansas' current minimum licensing standard for staff education, a high school diploma or GED and 10 hours of in-service training, is far-removed from the field's best practices. However, only rewarding providers who can afford to make the substantial leap from simply hiring teachers with high school education to hiring teachers with four-year degrees in all classes would alienate many from the QRIS. Moreover, private providers who must pay more for better educated teachers would pass costs on to consumers, which might force lower-income families to choose informal or lower quality forms of child care (Kelley and Camilli, 2007). Sudden adoption of more stringent policies could also have unintended consequences. For instance, in 1998 the New Jersey Supreme Court ordered that all children 3 years and up receive a high quality education in the high poverty Abbott districts, New Jersey. Under this mandate, teachers in all programs—whether public, private or Head Start—were required to obtain a BA and teacher certification. As teachers increased their qualifications, Head Start programs in the area lost many teachers, who were lured by the school district's offer of higher pay for teachers with BAs (Ryan & Ackerman, 2004).

The early childhood debate about the importance of well-educated teachers takes a slightly different turn in the school-age literature, as after-school programs typically have developed alongside the school day, with many after-school programs utilizing school teachers, other professionals from the community, and

[27] Time line for implementation accessed at http://www.naeyc.org/files/academy/file/Time%20Line%20for%20Meeting%206_A_05.pdf.

college students on a part-time basis. Yet, as we reviewed the research and recommendations from leaders in the field, we still found highly qualified staff to be considered a basic and necessary requirement for high-quality programs (Bodilly & Beckett, 2005; Committee on After-School Research and Practice, 2005; Miller, 2005).

To maximize inclusivity and encourage provider participation in the quality improvement process, BB calibrated their standards to existing levels in Arkansas. The Administrator/Staff Qualifications/Professional Development component of BB encourages increased levels of training for teachers and administrators but with more emphasis on clock hours than on formal college hours:

At the lowest level of quality

- All staff register for TAPP or the Arkansas Department of Education Registry (1.B.1).
- Administrators should have a CDA, 135 clock hours or 9 semester hours. Topics for some of the clock hours are designated (1.B.2).
- All other staff should have 15 clock hours that includes orientation (1.B.3).

At the highest level of quality

- On top of requirements for Level I, administrators must have 24 additional clock hours (3.B.1).
- 50% of staff must have 45 clock hours or 3 semester hours (3.B.2).[28]

Requirements for school-age providers are identical to the center-based standards. Requirements are slightly lower for family child care providers. Primary caregivers in family child care homes must have 30 clock hours at the lowest level and an additional 15 clock hours plus 10 hours of ongoing professional development annually for the highest level.

In this section, we present a spectrum of findings related to professional development. Results are not uniform, and there are considerable limitations within the research. The combined evidence offers no definitive threshold for the level of education necessary to achieve developmentally-enhancing quality. Examining the low end of the spectrum, we consistently see that a high school diploma or partaking in professional development that lacks a clear curricular focus and operates without individualization does not adequately equip ECE teachers. In general, findings demonstrate that more training is better, especially when the training is formal college education specialized in ECE or child development.

A number of studies identify a linear trend—the higher the degree, the better the outcomes—with teachers who possess a BA facilitating the best child development. However, others find no significant link between teacher qualifications and outcomes at all. There are mixed findings correlating teacher qualifications with child outcomes, possibly because education is serving as a proxy for teacher-child interactions, which the literature shows is a salient predictor of child outcomes. Our literature review explains that teachers who have more specialized training, in particular at the college level, do establish higher quality learning environments and interact with children in ways more supportive of child development.

---

[28] Better Beginnings standards for teacher qualifications are drawn from the Traveling Arkansas' Professional Pathways (TAPP) levels were accessed 08/10/2010 at http://www.arkansas.gov/childcare/services/aecpds/pdf/tapp%20map.pdf

The Qualifications and Professional Development component is a structural measure of quality. It relies on the assumption that teachers who have increased training will establish more positive environments for development and interact with children in more positive ways, but the component does not directly observe and confirm whether these results are produced.

5.1 TAPP REGISTRY

For entry into Level 1 for all types of care, all administrators and teaching staff must join the Traveling Arkansas' Professional Pathways (TAPP) registry (1.B.1). The registry provides professional development opportunities for early childhood educators and has expanded to include school-age educational development. The registry enables decision-makers to track trends in the workforce and to tailor trainings to meet the needs of staff and provides a centralized means of communicating opportunities for ongoing professional development. TAPP classes conform to the Arkansas Department of Education (ADE) rules governing professional development. The TAPP registry provides a way for members to track and document training hours.

5.2 TEACHER EDUCATION AND OUTCOMES

We found limited evidence that college-level training in ECE or a child-related major may confer benefits to children in cognitive, language, and social development. In some studies examining the relationship of teacher's degree to child development, the higher the degree, the greater the benefit to the child. Advantages begin at the CDA or AA level. Results from the Cost, Quality, and Outcomes Study (CQO) and the National Institute of Child Health and Development (NICHD) studies both identified relationships between teacher qualifications and outcomes.

Two analyses of NICHD data showed that teachers' college education was better for child development (NICHD Early Child Care Research Network, 1999, 2002). The 1999 results found significant relationships between college education in ECE and child outcomes at age 3, but not at age 2. Some college education was associated with language, school readiness, and behavior, but there was no distinct advantage afforded by a BA. The 2002 study demonstrated that caregiver training influenced children's cognitive competence and teacher rating of social competence with quality of care serving as a mediator between training and child outcomes.

CQO data revealed significant but modest associations between teacher education and their interactions. Howes' examined relationships between teacher education and outcomes of children included in the CQO study and the Florida Quality Study (Howes 1997). In analyses controlling for maternal education and classroom teacher-child ratios, CQO children had better receptive vocabulary when teachers possessed at least an AA in ECE. Among Florida Quality Study children, those who had teachers with either CDA or BA engaged in more language play. In both cases, teachers with an AA or CDA training were more effective than teachers who only had high school education or some college courses without a degree.

Another team later conducted applied hierarchical linear model analyses using the CQO data (M. R. Burchinal, Cryer, Clifford, & Howes, 2002). Variables included teachers' formal and informal training,

family selection factors, and state and classroom characteristics, including teacher-child interaction. Results indicated that children's receptive vocabulary performance was significantly higher if teachers had either a BA in ECE or training in community workshops. While this language outcome is notable, teachers who attended workshops were still less sensitive and had lower quality classrooms than teachers with college degrees who reported no workshop training. Similar to many studies within the qualifications and professional development literature, limitations of this study are that teacher report of education and training are likely less accurate than documented evidence of training, that the number and content of courses were not collected, and that more opportunities for training may be offered in centers that strive to maintain quality in ways not accounted for within the study.

There is debate about where the bar should be set for ECE teacher qualifications, with most of the debate centered on preschool care. One of the most vocal advocates for the BA as a minimum is W. Steven Barnett, Director of the Rutgers National Institute for Early Education Research (NIEER). NIEER measures state policies according to 10 structural benchmarks, including teacher's bachelor's degree. Barnett argues that empirical evidence supports the use of teachers with BAs, whether from a child development perspective or from an economic perspective (2003). Using evidence from 1989-2000, he writes, "The key finding is that only teachers with at least a four-year college degree consistently provide the good-to-excellent quality linked to future school success" (p.10).

However, in demonstrating the impracticality of a BA mandate for California childcare workers, Fuller, Livas and Bridges (2006) point out that Barnett has overstated the strength of the evidence. Some studies he has cited include small samples, are isolated to a particular geographic area, and/or do not adequately control for teacher and/or child background characteristics. Making generalizations for policy is also confounded by other limitations within the professional development research:

- Lack of common definitions or measurements of teacher education and training, whether in years or degrees, course hours or clock hours, etc.
- Scarcity of large-scale, long-term, and/or experimental studies on change in child outcomes even across the period of one year.
- Lack of knowledge regarding the content and quality of teacher education and training. For instance, insubstantial instruction related to interactions and the teacher-child relationship may affect child outcomes negatively.
- Some researchers suggest that structural features of the center, such as ratios and wages (Phillipsen, Burchinal, Howes, & Cryer, 1997) may diminish positive associations previously observed. For instance, more qualified teachers or talented teachers may choose to work at higher quality centers.

A few recent studies discussed here tackle some of the limitations by using more extensive and sophisticated controls and larger sample sizes, sometimes combining multiple datasets to gain more representative samples. The CQO and NICHD studies were both large and examined a range of child ages, but with more states implementing or investigating the feasibility of offering large-scale or universal state-sponsored pre-kindergarten programs, there are opportunities to better understand ECE teacher qualifications at a broader level. However, these studies tend to focus on preschool rather than on infant-toddler care and education. Results indicate either very small, significant associations or null findings between teacher education and child outcomes.

For instance, a meta-analysis of 32 studies of pre-kindergarten teacher education found that a teacher BA had an effect of only .16 standard deviations ($p < .05$) on child outcomes compared to lower teacher education levels (Kelley & Camilli, 2007). The National Center for Early Development and Learning Multi-state Study (NCEDL; Early et al., 2006) used a stratified random sample of more than 800 children, testing children's pre-academic skills, receptive vocabulary, oral and written language, academic achievement, rhyming, and identification of color and letters. Teachers' BAs in preschool classrooms related to significant gain in math skills regardless of major, and CDA related to gains in basic skills, but this study found no other consistent relationships with quality or with other measures of development. In another study using NCEDL Multi-state Study data and the NCEDL-NIEER Statewide Early Education Programs (SWEEP) Study, only one NIEER indicator of quality, teacher's BA, was associated with social competence in state pre-K programs. The authors suggest that NIEERS' use of dichotomous variables to designate teacher education, rather than scoring across a range, and the fact that the studied programs were built on NIEER indicators as possible explanations for the lack of associations with other outcomes (Mashburn, et al., 2008).

In a separate analysis, the lead investigator of the NCEDL study, also participated in a study combining seven major datasets (Early et al., 2007). [29] The analysis produced mostly null associations between teacher education, major, classroom quality, and child development. The team hypothesized that teacher preparation programs may be inadequate, teachers likely do not receive the support they need to implement what they have learned, and higher quality teachers with BAs may be choosing to work in higher-paid positions working with older children.

Although recent findings should temper passionate directives for BAs in all classrooms, they should not be used to downplay the importance of college education. As we will discuss in subsequent sections, specialized college education appears to outfit teachers with important interaction skills that influence child development and provide a springboard for making later professional development efforts more successful.

5.2.1 EDUCATION'S INFLUENCE ON INTERACTIONS AND QUALITY

In addition to direct benefits that teacher education may have on child outcomes, there is also evidence that education may indirectly affect child outcomes by molding the way teachers interact with children and by enabling them to raise the overall quality of care in the program. Teacher interactions influence child outcomes. Children's language, cognitive, and social-emotional skills thrive when teachers are more sensitive and responsive to children's needs and when they use fewer harsh behaviors. Higher levels of teacher education appear to improve the quality of teacher-child interactions and the caregiving

[29] Early Head Start, Family Child and Experiences Survey, Georgia Early Childhood Study, More at Four, National Center for Early Development and Learning, National Institute of Child Health and Development, and Preschool Curriculum Evaluation Research.

environment across multiple quality measures.[30] In the ECE literature, there is considerable overlap between studies investigating teacher education's effect on teacher behaviors and on global quality. A number of studies report on both types of outcomes simultaneously. To avoid unnecessary repetition, we will combine them for this section.

Certain populations are more likely to experience less educated caregivers who do not provide high quality care. NICHD data shows that children from lower income families are less likely to experience more sensitive behaviors and cognitive stimulation and more likely to experience negative behaviors (Dowsett, Huston, Imes, & Gennetian, 2008). Likewise, NCEDL Multi-State Study data indicated that when 60% of a classroom was populated with children from homes below poverty line, teachers had less specialized training or college education and had fewer appropriate beliefs about child development (Robert Pianta, et al., 2005). The Multi-State Study team found that higher education was a significant predictor of overall classroom quality. Thirty-one percent of the teachers in the sample had MAs or above, yet the mean ECERS score was 3.86, only average quality with the lowest factor scores in teaching and interacting and provisions for learning (R. Clifford, Bryant, & Early, 2005). These findings were focused only on pre-K children. The teacher characteristics required to produce better care for infant/toddler may be different.

Findings related to interactions are fairly consistent between small and large studies. The Bermuda Study was an early but seminal study into professional qualifications (Arnett, 1989). Among 59 teachers, some community college training improved interactions, but teachers with a four-year degree in ECE were less authoritarian, engaged in more positive interactions, were less detached, and were less likely to react punitively than teachers with a few college-level specialized courses with the comparison groups. On a much larger scale, the Head Start Family and Child Experiences Survey (FACES) found that rooms with higher ECERS scores were staffed with teachers who had higher qualifications and more sensitive interactions (Nicholas Zill, et al., 2001). The correlation coefficient approached .20 for total ECERS score ($p<.05$) and ECERS language subscale ($p<.01$). Nearly one third of the teachers had a BA or an MA. Howes' 1997 analysis of the CQO and Florida Quality studies revealed that teachers with CDA training initiated more interactions with children but were not always responsive to the individual needs of the child. Teachers with BAs appeared more capable in providing appropriate, individualized interactions.

In addition to enabling teachers to establish higher quality classroom environments, teacher education may also be an important antidote for other risk factors that can otherwise disrupt classroom processes. One such risk is teacher depression, which increases the likelihood of harsh behaviors or greater detachment (Hamre & Pianta, 2004). Fortunately, there is evidence to suggest that depression is less likely to negatively affect teacher's attunement and positive expression when s/he has more specialized ECE training (Gerber, Whitebook, & Weinstein, 2007). Similar findings were reported by Mill and Romano-

[30] Examples of studies using each measure: ERS (D. I. Cassidy, Buell, Pugh-Hoese, & Russell, 1995), ORCE (Clarke-Stewart, Vandell, Burchinal, O'Brien, & McCartney, 2002; NICHD Early Child Care Research Network, 1999), the Caregiver Interaction Scale (de Kruif, McWilliam, Ridley, & Wakely, 2000), CLASS (Robert Pianta et al., 2005), and Assessment Profile (Huffman & Speer, 2000).

White (1999), “though educator training did not predict anger or affection, training did in fact become very important to the quality of the interactions the educators had with the children when other risk factors were present” (p. 155). These risk factors included instabilities of the workplace, such as low wages or poor relationships with a supervisor. It is possible that a similar moderating effect may be possible for a different type of structural risk, high student-teacher ratios.

The relationship of teacher education to interactions may vary by children’s ages. There is some evidence that the level of degree is not as important in infant-toddler rooms as in preschool rooms as long as the teachers receive college-level specialized training, in which case they provide more appropriate instruction and sensitive care (Whitebook & et al., 1989) and higher quality environments (Phillips, et al., 2000). Again, in the CQO study, education had a significant correlation with higher process quality scores in preschool classrooms but not in infant/toddler classrooms. Process quality of infant/toddler rooms was associated with more experienced, better paid teachers and more experienced directors (Phillipsen, et al., 1997).

Of course, investments must be made to increase teacher qualifications, and this is a big hurdle to cross. In this regard, the incentive grants for professional development related to CDA or college-level training are a vital element of BB. In addition to potential benefits for children, staffing better educated teachers could save centers money. Although based on 1989 data, a cost-benefit analysis estimated that increasing teacher education requirements by one year raised total center annual cost by 3.4%, which included a 5.8% increase in teacher wages. However, the authors maintain this investment could substantially offset the even higher costs affiliated with teacher turnover. In this study, when there was 10% turnover in the center (a very conservative percentage) annual costs increased 6.8%. We know that many centers have much higher turnover, so investing in teachers who see their job as a career and are more content with their salary may actually save money in the long run. [31]

When examining school-age care, Beckett and colleagues (2001) identified that hiring and retaining educated staff and providing adequate compensation had strong support in the literature. One study looking at 30 school-age care programs in the Madison, Wisconsin area found more frequent negative interactions between staff and participants in programs with less educated staff (R. Rosenthal & Vandell, 1996). The authors noted that much of the literature in their review was not based on empirical evidence, but rather on the recommendations of panels and experts, and they drew from both areas for their discussion.

The MARS study, conducted in 78 after-school program sites in ten geographically and economically diverse school districts in Massachusetts, found that programs with more highly educated staff both at the program director level and direct service levels (defined as number of staff having a college degree) were rated significantly higher on several elements of program quality, including staff engagement, youth engagement, activities, and homework time. Programs that utilized certified teachers and other school staff also tended to rate higher on these measures of quality (Miller, 2005). When looking at participant outcomes, the evaluators found that increases in children’s homework persistence and completion were related to programs with more highly educated staff and directors. The quantity of after-school research studies has grown substantially in the years following this meta-analysis, but the input of expert panels

[31] The study sample included accredited centers providing care for 4- and 5-year-old children with 70% of the centers being from the South and Midwest.

remains a strong influence, and the relationship between individual program components, overall quality, and youth outcomes is still a subject of debate (Bodilly & Beckett, 2005; Miller, 2003).

5.3 ADMINISTRATOR QUALIFICATIONS

Very little attention has been devoted to administrator education and qualifications. We found one study that identified an association between administrator experience and child social competence (Phillips, McCartney, & Scarr, 1987), but we did not find administrator education related to outcomes. Slightly more evidence is available relating administrator qualifications to quality. In 122 California programs, directors' credential levels predicted classroom quality (Vu, Howes, & Jeon, 2008). The technical report of the CQO Study described a significant relationship between administrators and child care quality. In particular, administrators with more child care experience and more active involvement in curriculum planning tended to run higher quality centers (Helburn, 1995). Among Chicago Head Start centers, higher classroom quality was observed in centers where directors had a BA or higher ($r = .218$, $p = .01$), had completed 24 or more hours of ECE coursework ($r = .192$, $p = .02$), and had made at least four professional contributions during the past three years ($r = .20$, $p = .018$). Classrooms with low quality were more likely to have directors with less than an AA ($r=.-189$, $p = .026$) (McCormick Center for Early Childhood Leadership, 2010).

The after-school literature also emphasizes strong leadership (Birmingham, et al., 2005; Bodilly & Beckett, 2005; Jordan, Parker, Donnelly, & Rudo, 2009). In the MARS Study, a highly qualified coordinator or program director was found to be key for promoting youth engagement, staff engagement, and high quality activities (Miller, 2005). The 2003 After-School Summit found that strong program management, including adequate compensation of qualified staff, was essential in sustaining high quality in after-school programs (Committee on After-School Research and Practice, 2005).

Education and training are important, but not sufficient to produce strong leadership. Other qualifications considered essential for best practice in school-age care include experience in youth development and a strong connection to the community and families served. A dynamic leader maintains strong connections with staff through activities like regular staff meetings, supervision, and feedback about what is and is not working. Their vision for what the program should be provides motivation and support for program staff (Birmingham, et al., 2005).

5.4 CLOCK HOURS

We know the tales of teachers who have left one workshop jumping with joy, inspired and equipped to implement positive practice, but then left the next one resentful of the time they wasted. In-service professional development for the ECE field is under-researched. The effects of workshop training are not typically studied in depth, and when they are, they are often found ineffective in comparison to college-level specialized education. For instance, the evaluation of Pennsylvania's rating system showed that completion of requisite professional development was linked to ERS quality in family child care, but center-based teachers who *did not* complete the professional development requirements (n=29) scored

significantly better on two ECERS-R subscales, Interaction (5.99 versus 5.03) and Parent and Staff (5.89 versus 4.98)(Barnard, et al., 2006). No doubt, this is an exasperating find.

More optimistic conclusions came out of the NCCSS and CQO studies. Within the NCCSS, teachers with 15 hours of specialized in-service training within a one year period were found to provide more appropriate care and were more sensitive and less detached than teachers without the training (Whitebook & et al., 1989). For CQO teachers in infant/toddler and preschool care, attendance of workshops was modestly related to higher global quality and more sensitive interactions regardless of the educational background of the caregiver (effect sizes .21 to .43), and the magnitude of differences between teachers who did and did not attend workshops was similar regardless of how much formal education the teacher possessed (M. R. Burchinal, et al., 2002). Children's receptive language scores were positively related to teacher BA or to attendance of community workshops. Language outcomes did not correlate to in-service workshops that occurred in the teachers' centers or to training that occurred within professional development meetings. Limitations of this study are that the duration and contents of workshops were not observed and that centers willing to send their teachers to community workshops may also be hiring better teachers.

There is no evidence of a simple linear relationship between the number of clock hours input and the output of levels of quality or enhancement to child outcomes, but there is general agreement that more positive gains are produced when training for center-based or FCC teachers has the following characteristics:[32]

- an extended and continuous format with each session building on earlier sessions rather than one-day, "one-shot" type courses.
- a curriculum that is fixed yet is individualized to its participants.
- participants have opportunities to apply their knowledge
- includes trainer observation and feedback related to classroom implementation.
- participants have opportunities to reflect on what they have learned and to share their accomplishments and challenges.

Additionally, Garet and colleagues (2001) identify evidence that training which has a reform-type format (classified as committee task force, study group, mentor relationship) versus a traditional workshop or course is more effective than other types of workshops.

The curriculum and individualization of training matter. Results of a meta-analysis of experimental and quasi-experimental studies published between 1980 and 2005 showed that specialized training in interaction skills for caregivers did produce a significant positive effect (*d*=.45, S.E. =.10) on caregiver pedagogical competencies across caregiver educational levels (Fukkink & Lont, 2007). Training periods in the studies lasted as short as 4 days or as long as 80 weeks with the average period lasting 6 months. Most had "a broad scope and often an introductory nature" (p.306). Positive results were often insignificant, or even null, for courses that lacked a fixed curriculum and were delivered at multiple sites

[32] (A. S. Epstein, 1993; National Research Council, 2001; Spodek, 1996; Zaslow & Martinez-Beck, 2005)

to a wide variety of learners. No clear relationship between caregiver-level and child-level effects was identified, perhaps because many of the tests used to measure development were not properly aligned with the subject of the training. This and other studies underscore the importance of an individualized dialogue between the trainer and trainees. For example, Howes, James, and Ritchie (2003) studied a sample of African-American and Latino teachers working with low-income children in centers. Although their sample was not representative of the greater ECE field, the researchers were particularly looking for alternative pathways to formal degrees for teachers who would have trouble accessing college education even with increased wages. Although results still reinforced the importance of a BA, teachers who had experienced a combination of being mentored and being supervised in a reflective fashion were as responsive and as engaging of children in emergent literacy activities as teachers who possessed a BA.

Like many other large-scale programs, BB designers are turning toward Web-based training as a cost-effective, convenient medium for professional development. SF and Framework Basics trainings are forecasted to be offered online. A few clear advantages are that teachers do not have to travel, space is not required, and the curriculum is standardized. Nevertheless, inherent weaknesses are that training may be applied out of context and that anonymity will weaken results. We investigated online ECE trainings and found that mentorship is a key ingredient for optimizing results. MyTeachingPartner (MTP) is one of the few evaluations of Web-based trainings available. The goal of MTP training is to enhance instructional, language, and social interactions by providing teachers opportunities to view and analyze video examples of classroom interactions using the Classroom Assessment Scoring System (CLASS; RC Pianta, La Paro, & Hamre, 2008). Within one state pre-K program, teachers who only received access to the site and video clips showed significantly less improvement in their own interactions than those who also received individualized online consultations (RC Pianta, et al., 2008). The combination of video access and individual feedback was particularly effective in classrooms attended by higher percentages of children from low income families. Another evaluation of an ECE online course targeted at teachers of high-risk preschool children tested four conditions of a language-literacy training program (Landry, Anthony, Swank, & Monseque-Bailey, 2009). All teachers experienced small-group online training with a facilitator and a comparison group additionally experienced in-class mentoring and/or feedback and recommendations based on child monitoring data. Teachers' instructional practices were observed for portfolio use and modeling and support of language/literacy activity. The most enhanced professional development produced more positive results in teaching practices and children's academic readiness[33] across sites in four states with varying student populations and teacher education levels.

5.4.1 SCHOOL-AGE CARE AND CLOCK HOURS

After-school programs have been around since at least the 1880s, when settlement houses provided services to immigrant families (Bodilly & Beckett, 2005). Yet only during the last decade, corresponding with an increase in federal funding, has research begun to examine what makes for quality programming. Typically the development of quality standards parallels existing education and training structure in the wider education field.

Some states are developing training programs specific to after-school. For instance, the Pennsylvania STAR program has developed a 15-hour self-learning module specifically for after-school care that is

[33] Language comprehension, phonological awareness, vocabulary, print and letter awareness.

designed to complement program level orientation and training. In Ohio, the Career Pathways professional development system is being adapted to include a school-age care track. Vermont is developing after-school education and care coursework in conjunction with local community colleges and universities (Afterschool Investments Project, 2007). Additionally, a network of non-profit organizations committed to helping establish, support, and evaluate quality service provision in after-school programming has developed due to increases in demand and federal funding. It is committing considerable resources to the field, including "research, information, technical assistance, and advocacy focused on improving and expanding the offerings in the market place" (Bodilly & Beckett, 2005).[34] Among the varied stakeholders, there is some general consensus emerging as to what helps a program succeed, and well-trained staff is one necessary component.

"Ongoing Training and Staff Development" is listed as one of the six critical components indicating program quality in the Arkansas Standards for Quality Afterschool Programs (Arkansas Standards) developed by the Arkansas Out of School Network. National experts in the field have likewise emphasized training of after-school staff as an essential component of quality programming. The Rand Corporation's literature review of key program elements associated with quality included "stable, trained personnel" (Bodilly & Beckett, 2005). The Harvard Family Research Project similarly emphasized the importance of ongoing staff training, as opposed to orientation and pre-service training: "Intentional, focused programming entails a clear vision and goals for the program from the start, as well as strong, directed leadership and sustained training and support to staff" (Little, Wimer, & Weiss, 2007).

The C.S. Mott Foundation's Committee on After-School Research and Practice found "qualified afterschool program staff and volunteers with regular opportunities for professional development and career advancement" to be a necessary condition for high-quality in after-school care (Committee on After-School Research and Practice, 2005). And the National After-School Association standards for quality school-age care require that a quality program have "effective staff and administration, with committed and well-trained staff and volunteers, frequent and efficient staff meetings, and ongoing training opportunities" (National Institute on Out-of-School Time, 2000). In a review that assessed the literature in three fields: school-age care, youth development, and education, the authors found that school-age care and education literatures both emphasized the importance of staff or teacher training, and in fact found this component to have the strongest empirical support in existing research, along with one other quality characteristic: limiting the size of the program or classroom (Bodilly & Beckett, 2005).

The quantity of training hours may not be sufficient to affect program quality. A meta-analysis of best practice in after-school programming (Beckett, et al., 2001) found that training staff was one of the essential components of quality, but they found that for training to be effective, it must be specific to school-age children and relevant to the employee's job position. In the MARS Study, the number of hours of staff training was found to be related to higher quality staff engagement, but not related to other elements of program quality (Miller, 2005; Sheldon, Arbreton, Hopkins, & Grossman, 2010). Similarly, the timing of training may be crucial. When after-school programs adopt the school model of emphasizing pre-service training and do not provide ongoing staff training and consultation, attempts to implement a well-structured and consistent program, particularly a formal curriculum, will be difficult (Sheldon, et al., 2010). In The After School Corporation evaluation, the highest-performing programs were found to be

[34] See p. 25, Table 2.2 for a description of this network.

those led by strong, experienced leaders committed to ongoing consultation, supervision, and training of staff, in addition to regular staff meetings throughout the program year (Birmingham, et al., 2005). In programs with diverse offerings, more training focused on the "point of service" is important, so staff become well grounded in the specific subject matter they intend to convey to youth and in the unique developmental needs of the children (Yohalem, Wilson-Ahlstrom, & Yu, 2005).

As a whole, evidence from school-age programs parallel those from early childhood programs. Namely, training that is ongoing, reflective, and individualized is more likely to impact classroom quality.

5.4.2 BETTER BEGINNINGS REQUIRED TRAININGS

In addition to meeting clock hour requirements, requisite professional development courses for BB include administrator training in developmentally appropriate physical activities for children (1.B.5), ERS Training for the administrators and teachers (1.B.4, 2.B.6), Frameworks Basics for 50% of the teaching staff (2.B.5) and two clock hours of nutrition training for administrators and kitchen staff (2.B.7). In lieu of Frameworks Basics training, school-age staff participate in Developmental Assets Basics training for Level 2 (2.B.5).

- **ERS Training**
  Administrators entering Level 1 in center or family-based care (1.B.4) and 50% of center-based staff in Level 2 (2.B.6) take part in ERS training. We would expect this training to raise awareness of factors that contribute to global quality, enable staff to make improvements prior to assessment, and to inform staff of what to expect during the formal assessment process. We did not find evidence that training prior to assessment increases global quality or contributes to better child outcomes.

- **Youth PQA Training**
  School-age program administrators may choose to substitute Youth PQA training. We did not find evidence that training prior to assessment contributes to better participant outcomes. However, becoming knowledgeable about the tool will help administrators and staff in terms of "creating a common language and getting a conversation started" about building quality in after-school programs (Yohalem, et al., 2005).

- **Developmentally Appropriate Physical Activities**
  The second training assigned administrators in Level 1 is in developmentally appropriate physical activities for children (1.B.5). Motor development and physical fitness are important aspects of a child's wellbeing and support development in other domains. As obesity rates among young children rise, it is important that teachers increase their knowledge of effective means to promote physical fitness. However, we found no data on the effects of administrator training in physical activities on child outcomes. Two studies associated teacher's college education (M. Dowda, Pate, Trost, Almeida, & Sirard, 2004) and recent training in physical activities (Marsha Dowda et al., 2009) as factors that increase preschoolers' moderate-to-vigorous physical activity (MVPA) during the school day, but it is unclear how these findings might relate to education and training at the administrator level.

Likewise, for school-age programs, we did not find evidence that this training would make a difference in outcomes for youth. However, it may be true that administrators need to be better educated about the benefits of teaching our youth healthy practices. One report found 77% of superintendents and 54% of principals felt that most schools were doing an adequate job implementing wellness policies, while at least 72% of community health professionals and physical education teachers said they were not (Action for Healthy Kids, 2008). Our youth need to be encouraged to be physically active, and what administrators think is important will influence program curricular and activity choices.

- **Frameworks Basics**
  For Level 2, 50% of center-based teaching staff or the primary teacher in family child care should participate in a two-hour Framework Basics training introducing staff to a developmental framework intended to enhance academic preparation and transition into kindergarten, design and development of curricula, and child assessment (2.B.5). BB designers hope that this short, introductory course will pique interest in and drive enrollment into the full 12-hour Frameworks training. In Arkansas ABC programs, Honeycutt (2008) reported that the full Framework training correlated with an increase of .17 on the ECERS-R, but no statistically significant increases in student proficiency as measured by the Work Sampling System. [35] ABC programs are state-funded with additional resources for professional development and requirements for quality. As such, it is difficult to generalize the findings to other centers. The abbreviated Framework training has not been tested for effectiveness in enhancing classroom practice.

- **Nutrition Training**
  Two clock hours of training in nutrition are required for the administrator and kitchen staff in centers or for the primary caregiver in child care homes (2.B.7). We know that children do not get the recommended minimum of five servings of fruit, juice, and vegetables each day. Lack of nutritious foods coupled with an abundance of unhealthy foods put children at risk for becoming overweight and other diseases such as cancer and diabetes. The availability and accessibility of nutritional foods, meal structure, provider food modeling, and socialization practices within day care affect children's eating habits and nutrition. For further evidence, see a review by Nicklas and colleagues (2001).

  Better nourished children are less likely to experience mental and physical setbacks and are more likely to excel in school (Glewwe, Jacoby, & King, 2001). However, as discussed previously, one-session, one-size-fits-all trainings are unlikely to change staff practices. To ensure change in practice, and ultimately in child outcomes, staff would need to participate in long-term, ongoing professional development with onsite, individualized coaching and feedback.

  More comprehensive yet low-cost programs, such as Nutrition and Physical Activity Self-Assessment for Child Care (NAP SACC) and Color Me Healthy, have produced mixed results. For example, NAP SACC implementation in one Arizona county involved seven workshops, assessments, and customized action plans for each site. Many sites then offered three workshops for parents. Centers increased the number of nutritional best practices from a median of 25 to

[35] Our reading of the dissertation produced questions about the methods used to analyze the data that remain unanswered, so we convey these findings with caution.

30, with many centers choosing to offer a greater variety of foods with more nutritional value served family style (Drummond et al., 2009). On the other hand, a treatment group in a NAP SACC implementation in North Carolina did not show significant effects (Ward et al., 2008). Color Me Healthy was designed to enhance physical activity and nutrition. When introduced in North Carolina, with less extensive training than NAP SACC, provider reports of children trying new foods and recognizing more fruits and vegetables increased (C. Dunn et al., 2006).

- **Developmental Assets Basics**

  For BB Level 2, 50% of teaching staff must complete Developmental Assets Basics training. The framework of 40 Developmental Assets developed by the Search Institute pulls together the diverse elements of human experience that have long-term, positive results for youth into a "comprehensive vision of what young people need to thrive."[36] The Search Institute collected data on youth attitudes and behaviors of more than 350,000 youth Grades 6-12. In developing the assets, the authors looked at three areas of applied research: (1) youth development (what are the core processes that lead to healthy maturation), (2) prevention (what protective factors work to prevent high-risk behavior), and (3) resiliency (what facts increase a youth's ability to bounce back in the face of adversity).

  Although the original research was with adolescents, the Search Institute asserts that the basic strength-based approach and assets framework is consistent with research on what kids need to succeed throughout childhood, and it has continued developing separate frameworks for early and middle childhood. The 40 Developmental Assets are available for four age groups: Early Childhood (ages 3-5), Grades K-3 (ages 5-9), Middle Childhood (ages 8-12), and Adolescents (ages 12-18). Each list of 40 Developmental Assets is organized within two broad areas, external assets and internal assets, that are further divided into subsections. External assets include support, empowerment, boundaries/expectations, and constructive use of time. Internal assets include commitment to learning, positive values, social competencies, and positive identity.

  The assets have been used to predict whether youth will thrive, based on the premise that the more positive developmental factors that youth are exposed to, the more likely they will be to also report "thriving indicators" like school success, leadership, and good health (Scales, Benson, Leffert, & Blyth, 2000). Developers of the Youth PQA considered staff training in an "explicit youth development approach" to be an important element contributing to best practice in youth programming (Smith & Hohmann, 2005). And the Harvard Family Research Project's seven recommendations for high quality after-school programs included the development of quality standards that are asset based (Westmoreland & Little, 2006).

  Although the Developmental Assets have sound underpinnings, we did not find research suggesting that the related BB training would result in improved youth outcomes.

---

[36] http://www.search-institute.org/developmental-assets/lists

The selection of topics chosen for BB staff training is appropriate and valid to the needs of children. However, these trainings do not appear to possess the structural characteristics of trainings likely to affect child outcomes described above in Clock Hours section. As an intermediate step to adopting college-level ECE credits for BB professional development, we recommend that curriculum technical assistants steer providers toward courses provided by DCCECE that meet more criteria for effective trainings, especially those related to teacher-child interactions, to fulfill TAPP professional development clock hours. Examples include Pre-K Early Learning in Arkansas (Pre-K ELLA, 30 hours) and Pre-K Social Emotional Learning for Young Children (45 hours).

5.5 VARIATIONS IN FAMILY CHILD CARE

There is some variation in the research and advocacy related to the education of family child care teachers. Family child care literature reflects the importance of college-level training, but we do not see evidence that a BA is necessary to achieve good interactions and positive child development. The research is mixed as to whether college education must be specialized in ECE. Limited evidence points to the value of ongoing professional development throughout the family child care provider's tenure. Family child care professional development research reflects similar findings as center-based research: training should be extensive and individualized. Most studies have associated family day care training with global quality and/or to caregiver-child interactions. Very little data on child development in family day care has been collected.

More formal education above the high school level translates to higher quality care (M. R. Burchinal, et al., 2002; Ellen Galinsky, Howes, Kontos, & Shinn, 1994). In positive caregiving and in global quality measured by the *Home Observation for Measurement of the Environment Inventory* (HOME, Caldwell & Bradley, 1984), NICHD family day care participants who had college educations were significantly different from those who had none. More teacher education related to children's higher cognitive and language development scores among children (Clarke-Stewart et al., 2002). Additional benefits of college-level specialized education appear to be higher social competence among children (Susan Kontos, Hsu, & Dunn, 1994).

The Pennsylvania quality rating system evaluation found that family child care teachers with an AA or higher had significantly higher FDCRS scores than those without (mean 4.82 vs. 4.21, $p<.001$). Difference on the Language and Reasoning subscale was 5.22 versus 4.41 (Barnard, et al., 2006). A prior study of all types of care in Pennsylvania revealed that family child care home providers with an AA or higher were also more likely to use positive practices in arrangement of indoor space, safety, television use, language stimulation, and professional development activities (Fiene et al., 2002).

The Study of Children in Family Child Care and Relative Care (FCC Study) (Ellen Galinsky, et al., 1994) reported that out of 226 providers, 44% of the teachers studied had a high school diploma or less, 38% had some college or an AA, and 17% had a BA or higher. Teachers with more formal education were rated as more sensitive and less detached and observed to be more responsive. Burchinal and colleagues combined the FCC Study sample with the California Licensing Study sample and found that teachers with more education exhibited higher global quality and less detachment as measured by the CIS (2002). For

these two studies, the models did not include variables for college major, so they give us no insight whether specialization is necessary, but other studies that have investigated teacher major found that college-level specialization makes a difference. Doherty and colleagues (2006) used linear regression to examine the following variables' effects on quality: level of general education, degree of intentionality, training and experience in family child care, use of support services, and work environment. Among the 231 regulated teachers studied, only ECE college training emerged as a predictor of quality. Level of general education did not. A prior study of the Vancouver Day Care Research Project also found that training specifically related to family day care, rather than general college education, predicted quality (Pence & Goelman, 1991).

Outside of college education, family child care training programs can foster a sense of professionalism and the interest in further education and development (E. Galinsky, et al., 1995; Modigliani, 1993). Training should be ongoing throughout the family child care teacher's career. Clarke-Stewart observed that learning environments were of higher quality and teachers were more sensitive to children when training was more recent (Clarke-Stewart, et al., 2002). This was also found to be the case in an Oklahoma evaluation (Norris, 2001), where caregivers who received workshop training on a regular, ongoing basis had higher FDCERS scores than those who participated intermittently. "On the learning activities subscale, for example, continual participants have a greater variety of materials and opportunities for eye-hand coordination, art, and music. In addition, these providers are more likely to specifically plan a balance of activities inside and outside. These providers are also more likely to interact with the children while playing rather than just supervising activities" (p. 119). Burchinal, Howes, and Kontos also observed significant positive effects on FDCR scores and teacher detachment scores when family teachers had participated in a workshop in the past year (2002).

Professional development trainings for family child care may not always produce results. For example, a validation study of Family-to-Family training involving 15-25 hours of class time and home visits found no significant effect of the training on process quality, and only 19% of the sample made statistically significant but very small improvement in the overall quality (Susan Kontos, Howes, & Galinsky, 1996). Similar to findings for center-based professional development, evaluations of family child care trainings show that professional development should be individualized to the participant to maximize results (Fukkink & Lont, 2007). The Midwest Childcare Consortium found that trainings with an "in-person" component were more effective than those that took place through video or study materials (H. H. Raikes et al., 2003). With a mentoring intervention in infant caregiving, family care teachers were more sensitive and responsive to infants than the control group, which only received state-provided workshop training (Fiene, 2002). In a language and literacy program for family care, there were no significant effects on teacher knowledge within two treatment groups (n=73) compared to control group (n=55), but a coursework plus coaching intervention had a significant effect on provider practice. No significant differences were observed between quality practices in language and literacy for coursework only group and control group (Koh & Neuman, 2009).

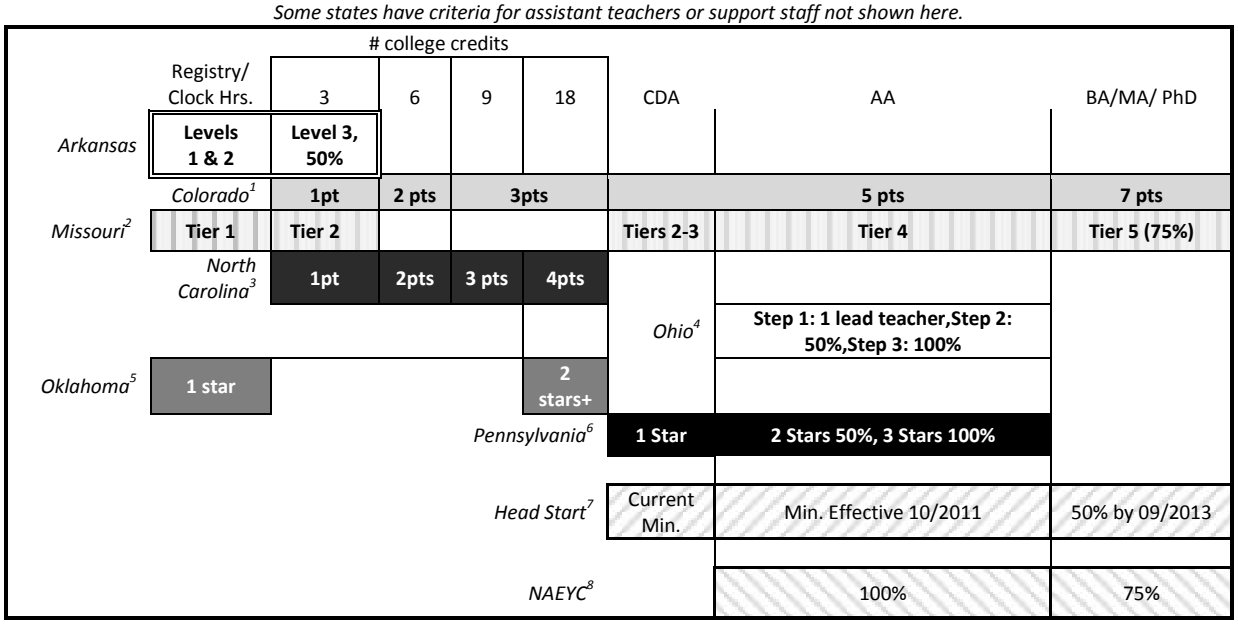
5.6 KEY STATES COMPARISON

The trend toward college credits rather than in-service professional development is evident within state quality rating systems. States such as Missouri have developed complex lattices for professional development with options combining credentials, higher education, and years or hours of classroom

experience. Table 5-A displays a summary of current requirements of the key comparison states, Head Start programs, and NAEYC. BB has the least stringent standards but is most closely aligned with Oklahoma's star system.

Table 5-A Lead Teacher Education Levels by State

When only a portion of staff must reach a level, percentages are indicated.
Some states have criteria for assistant teachers or support staff not shown here.



[1]**CO**: teachers scored individually on number of hours worked and on professional development. Individual teachers awarded 1 point (45 hours of training in past 3 years/Non-credit CDA/3 ECE credits) up to 7 points (BA/MA/PHD in ECE or BA/MA/PhD in non-ECE field with 24 ECE credits). Scores are combined with teachers' scores to calculate total number of points for the Training and Education component.

[2]**MO**: Tier 1 must meet minimum licensing standards of 12 clock hours. For Tier 2, if only 1 lead staff member, that employee should have CDA or 3 college credits. Lead teachers must also have 9 approved college credits or 14 clock hours of approved curriculum training. For tier 3 with one lead staff, employee must have CDA. For tier 4, AA. Requirements are higher for centers with more staff. Tier 5 centers must be accredited. NAEYC standards shown.

[3]**NC**: All lead teachers must have credential involving one 4-hour semester course or other work experience and education. Lower requirements listed for other teachers counted in ratios. For 2 education points, 75% lead teachers must have credential and 3 semester hours in ECE.

[4]**OH**: For step 1, one lead teacher with an AA in ECE or 150 points (e.g. MA degree = 50 pts; 20 hours in-service training = 1 point) and admin and teachers receive 10 clock hours specialized training annually. For step 2, 50% of lead teachers have AA in ECE or 150 pts. Career Pathways substitution available.

[5]**OK**: All staff have 20 hours of job-related training in the last year. For 2 stars, one Master Teacher required per 30 children in year 1 and per 20 children thereafter. For 3 stars, Master Teachers must be full-time and have additional credentials (e.g. a CDA or 4-year degree).

[6]**PA**: For 2 stars, 50% teachers/supervisors have AAs, including 18 credits in EC; 50% of assistant teachers at 45 hours professional development (workshop, seminar or college). For 3 stars, 100% teachers have AAs with 18 EC credits; 75% teachers have CDAs or specialized certificates; 25% of assistants at 45 hours of professional development. For 4 stars, 100% teachers have AAs; 25% teachers have AAs; 50% Aides at 45 hours professional development.

[7]**Head Start:** see Head Start Act Sec. 648A; http://eclkc.ohs.acf.hhs.gov.

[8]**NAEYC:** Standard 6.A.05. Number of teachers who must meet these requirements depends on number of groups in the center. Requirements are moving toward all classrooms having a teacher with a BA by 2020. Timeline found at http://www.naeyc.org/files/academy/file/Time%20Line%20for%20Meeting%206_A_05.pdf.

The picture is similar with family day care, but the spectrum of education identified in state QRISs is narrower for family day care than for centers. Although the states have similar starting points for quality, other states present higher levels of education.

Table 5-B Family Day Care Primary Caregiver Education Levels by State

Some states have criteria for assistant teachers or support staff not shown here.

	HS	Clock Hrs.	# college credits: 3	6	9-12	15	18	24	CDA	AA	BA/ MA/ PhD
Arkansas[1]		**Levels 1**	**Levels 2 & 3**								
Colorado		**1pt**		**2.5 pt**		**4pt**	**3pt**	**5.5**		**7pt**	**8.5-10pt**
Missouri[2]		**Tier 1**	**Tier 2**						**Tier 3**	**Tier 4**	
North Carolina[3]			**1-3pt**	**4pt**	**5pt**		**6pt**			**7 pt**	
Ohio[4]	**Step 1**									**Step 2: 1 lead teacher Step 3: All teachers**	
Oklahoma[5]		**1 star**							**1-2 Stars**		
Pennsylvania[6]		**1 Star**	**2 Stars**	**3 Stars**	**4 Stars**						

[1]**AR**: 30 clock hours for Level 1, 45 clock hours or 3 college credits Level 2, 60 clock hours or 3 college credits plus 15 clock hours for Level 3.
[2]**MO**: Tier 1, 12 clock hours. Tier 2, 3 college credits or 45 clock hours in past year. Tier 3, CDA or equivalent, 1 year certificate of proficiency, or 30 college credits. For tier 4, AA or 60 college credits. Tier 5 centers must have NAFCC accreditation, which does not have specific education requirements.
[3]**NC**: Emphasis is on the NC Family Child Care Credential. For first 2points, 4 EC college credits for 1 point or 5 years experience may be substituted. 12 hours college credits for 5 points.
[4]**OH**: High school diploma for Step 1. For Steps 2 and 3, a Career Pathways substitutions is available.
[5]**OK**: 20 clock hours within past year for any level, plus CDA or other substitutions such as OK valid teaching certificate or 30 credit hours with 12 in ECE or a related field. NAFCC accreditation, which does not have specific education requirements, is required for 3 stars.
[6]**PA**: 12 clock hours within past 2 years for 1 star. For 2 stars, complete Keystone Core Training Series and be working toward a CDA OR 3 college credits. For 3 stars, 6 EC college credits or be enrolled in a CDA program. For 4 stars 9 college credits or current CDA.

States with T.E.A.C.H Scholarship Programs: Colorado, North Carolina, Missouri, Pennsylvania, Ohio

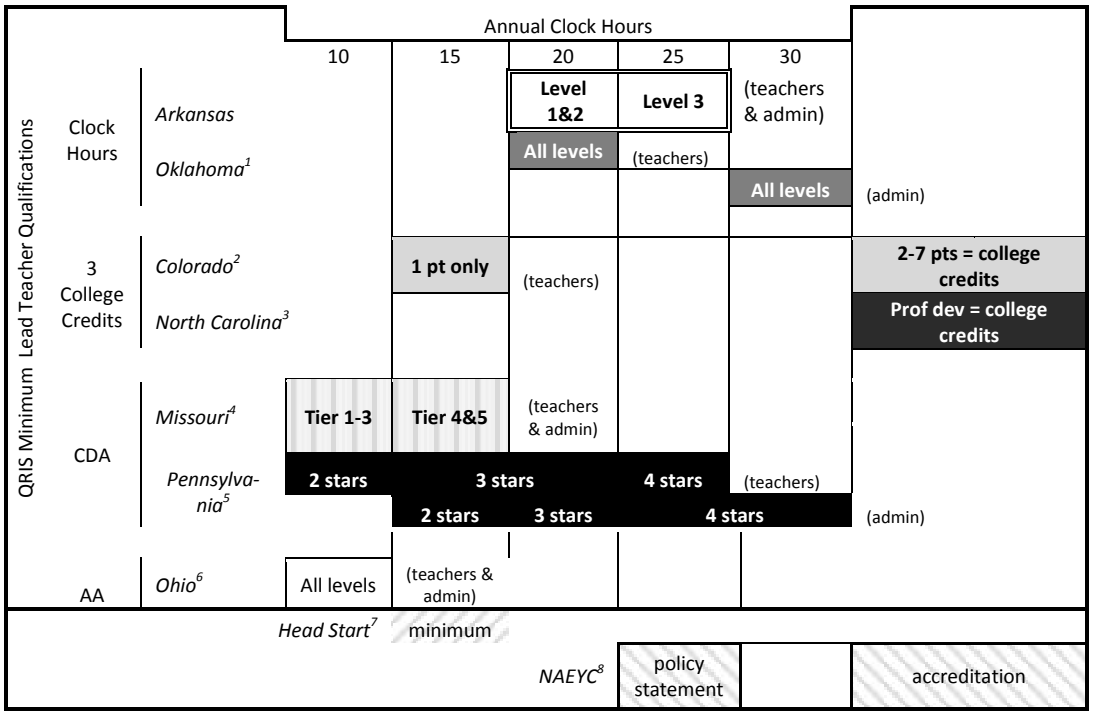
A trend in our comparisons is that states with higher formal education requirements mandate fewer annual clock hours of in-service professional development (Table 5-C). Professional development systems in Colorado and North Carolina are tied into the college system. As the cost of tuition is a great obstacle in raising the level of teacher education, most of the states shown here have developed infrastructures to subsidize tuition, most notably Teacher Education and Compensation Helps (T.E.A.C.H.) Programs. The states with the lowest education requirements but the most annual clock hours, Arkansas and Oklahoma, are the only states in our comparison *not* connected with T.E.A.C.H.[37]

None of the comparison states require administrator training in developmentally appropriate physical activities in the QRIS, and only Pennsylvania requires ERS training. There was more alignment with other states in terms of Framework training. We found that Missouri, North Carolina, and Pennsylvania provide

[37] For more information about T.E.A.C.H., see http://www.childcareservices.org/ps/teach_ta_qac.html.

training in state learning standards or core competencies. In North Carolina, this content is incorporated into college courses required for the rating system's administrator and teacher credentials. Professional development trends across states are similar in family day care (Table 5-D).

Table 5-C Center-Based Professional Development Clock Hours by State

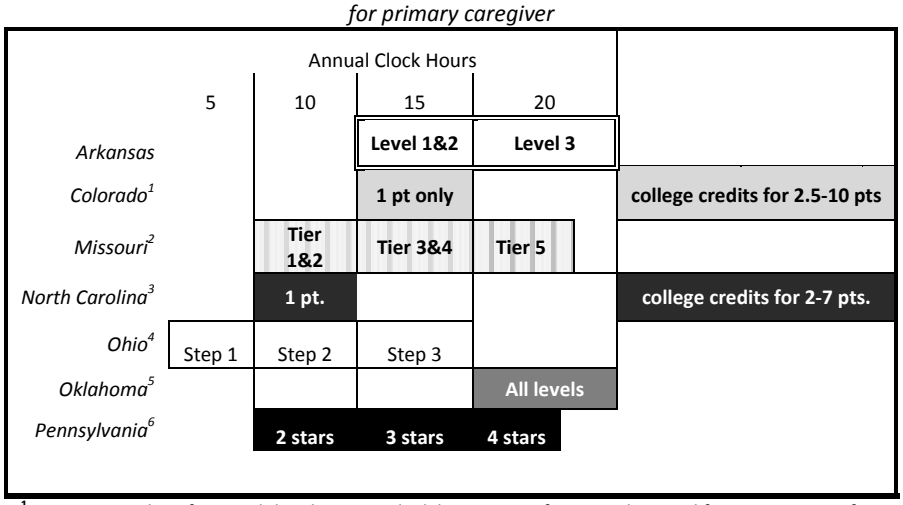


[1]**OK**: For all levels, staff receive 20 hours; directors receive 30 hours.
[2]**CO**: No annual professional development clock hours apart from teacher qualifications, except for the minimum teacher qualification, which is 1 point given for teacher training of 45 clock hours within past 3 years; 2-7 points given only for college credits or degrees. No points given for admin clock hours.
[3]**NC:**No ongoing yearly clock hours specified; emphasis is on college credits for professional development.
[4]**MO:** Tier 1 (minimum licensing) & 2, 12 hours. Tier 3, 14 hours. Tier 4&5, 16 hours. Number of minimum hours raised if less than 100% of the staff receive training.
[5]**PA:** Teachers: 12 hours for 2 stars, 18 for 3 stars, 24 for 4 stars. Directors: 15 hours for 2 stars, 21 hours for 3 stars, 27 hours for 4 stars.
[6]**OH:** Step 1- 3, admin and teachers receive 10 hours annual training.
[7]**Head Start**: Head Start Act 648A
[8]**NAEYC:** current standards emphasize credit-bearing coursework (10.E.11). A number of annual development hours is not included in accreditation standards, but an NAEYC policy statement recommends 24 hours annually (National Association for the Education of Young Children, 1993).

States with T.E.A.C.H. Scholarship Programs: Colorado, North Carolina, Missouri, Pennsylvania, Ohio

Table 5-D Family Day Care Professional Development Clock Hours by State

for primary caregiver



[1]**CO**: No annual professional development clock hours apart from teacher qualifications, except for the minimum teacher qualification, which receives 1 point for teacher training of 45 clock hours within past 3 years; 2-7 points given only for college credits or degrees. No points given for admin clock hours.

[2]**MO:** Tiers 1&2, 12 hours (plus 45 in last year if the provider doesn't possess 3 college credits; see Table 5-C). Tier 3, 14 hours. Tier 4, 16 hours. Tier 5, 18 hours.

[3]**NC:**12 clock hours per year for 1 pt.; emphasis is on college credits for professional development for more points.

[4]**OH:** Step 1, 5 hours. Step 2, 10 hours. Step 3, 15 hours.

[5]**OK:** 20 hours within the year for all levels.

[6]**PA:** 12 hours for 2 stars, 15 for 3 stars, 18 for 4 stars.

5.6.1 COLORADO

The RAND 2008 validity study found no relationships between center-based teacher qualifications and ECERS-R quality (G Zellman, et al., 2008).

5.6.2 MISSOURI

The Missouri QRS reports did not provide information specifically relating teacher education to child outcomes or to quality. However, all types of children in centers rated Tier 1 and Tier 2 lost social and behavioral skills from fall to spring, and children in poverty lost vocabulary. Teaching levels in Missouri's Tier 1 are comparable to Arkansas' Levels 1 and 2, but Missouri's Tier 2 exceeds Arkansas' Level 3 with regards to administrator and staff qualifications.

5.6.3 NORTH CAROLINA

The North Carolina Rated License Assessment Project identified a strong connection between teacher education and rating scale scores in participating programs. A two-year degree was found to be necessary to achieve good quality in centers (D. Cassidy, et al., 2003).

5.6.4 OHIO

The evaluation of the Ohio pilot did not provide information about teacher education related to quality or to child outcomes.

5.6.5 OKLAHOMA

High quality centers had teachers with higher levels of EC education. The ratio of Master teachers to children had a stronger relationship with quality (Norris, Dunn, & Eckert, 2003).

5.6.6 PENNSYLVANIA

The 2006 evaluation (Barnard, et al., 2006) showed that center and family care teachers had statistically higher ERS ratings when they had at least an AA. There were no significant differences between teachers with AAs and those with BAs in either type of care. The study did not investigate the major of the degree. Family child care teachers who had completed professional development requirements had higher ratings, but this did not prove to be the case for center-based teachers. Center-based teachers who had not completed the professional development requirements scored higher on the ECERS-R Interaction scale (5.99 compared to 5.03) and on the Parents and Staff scale (5.89 compared to 4.98).

5.7 CROSSWALKS

NAEYC and Head Start standards align with policy recommendations to raise the level of professionalization within the ECE field and with research finding that teachers who have higher levels of college education specialized in ECE create higher quality learning environments and contribute to better child outcomes. These standards exceed BB qualifications levels. NAFCC guidelines do not stress formal education, rather they encourage ongoing training and supportive connections with other family child care providers. COA standards emphasize a college education and terminal degrees.

5.7.1 HEAD START

The only Head Start standard regarding director qualifications is that they "must have demonstrated skills and abilities in a management capacity relevant to human services program management" (45 CFR 1304.52). Standards regarding teacher qualifications, on the other hand, are extensive and reflect time frames for increasing professionalization.

Head Start employees must have at least 15 hours of professional development each year (648A(f)). Agencies are required to create and implement individualized development plans that are regularly evaluated (1304.52(l)(2)). Agencies must also provide a structured training and development system that attaches academic credit to employee professional development as much as possible (45 CFR 1306.23).

Early Head Start and Head Start staff working as teachers with infants and toddlers must obtain a (CDA) credential for Infant and Toddler Caregivers or an equivalent credential within a year of being hired. They

must also possess training and experience in relationships, safety, and communication with infants/toddlers and their families (1304.52(f)).

Regulations from 648A(a)(2)(A) explain the timelines for raising the level of teacher qualifications. As of October 2011, 50% of all teachers in center-based programs must have a CDA, an AA in ECE or a related field, state certification, or have a BA and participate in the Teach For America program. In 2013, the qualifications will rise to 50% having a BA in ECE or a related major.

5.7.2 NAEYC

NAEYC recommends that administrators possess a BA with at least 9 credit-bearing hours in administration, management, or leadership and 24 credit-bearing hours in ECE or a specialized related area. Five-year completion plans or alternative pathways are accepted (10.A.02). All teachers are expected to have an AA or the equivalent, and 75% are expected to have a BA in ECE or a related specialty (6.A.05). Specialized college-level or professional training should address the following content areas: cultural/language diversity (6.A.07), the program's curriculum and collaboration skills (6.A.08), knowledge of the specific ages or special needs of the children they work with (6.A.10), and knowledge about working with children who have special needs (6.A.12).

Individual professional development plans must be updated annually (10.E.10) and emphasize credit-bearing coursework (10.E.11). Numbers of development hours are not included in accreditation standards, but an NAEYC policy statement recommends 24 hours of professional development annually (National Association for the Education of Young Children, 1993).

5.7.3 NAFCC

BB standards exceed NAFCC standards. NAFCC does not require minimum levels of education or experience for providers. Providers should seek "continuing training and education" and be "open to new ideas about family child care" (5.6), stay "up-to-date with topics related to program quality" (5.7), and be "actively involved with other providers or a related professional group" (5.8).

5.7.4 COA

COA standards include minimum qualifications for five job positions and require a BA, an AA, or two years of college for the top three positions. There are some provisions for experience in the field and training in youth development to count. For the top position, program administrator, a BA is required regardless of experience. For the next position, site director, an AA or two years of college with 18 months of experience plus training in child and youth development may be substituted for a BA. For the third position, senior group leader, an AA or two years of college with 6 months experience plus training in child and youth development may be substituted for a BA. For the fourth position, group leader, an AA with 9 months of experience plus training in youth development may be substituted, or if the individual has a high school diploma or GED, 18 months of experience plus child and youth development training is required. At the lowest level, employees work only under the direct supervision of a group leader and may be as young as 16.

5.8 CONCLUSIONS

Minding the gap between where Arkansas programs are versus where we would prefer them to be is a thorny task. It is one thing to tout the merits of teacher education and quite another to pay for tuition subsidies, enhanced professional development programs, and the increased wages that would result from a more professionalized workforce. Yet, our evaluation team still concludes that the Better Beginnings scale is tipped substantially more toward the status quo than toward the field's best practices. In light of ECE research and state comparisons, the system is likely insufficient to promote increased professionalism for teachers and optimal support for child development.

Research related to center- and family-based teacher qualifications and professional development suggests that advantages for child outcomes will begin at the Intermediate levels of Traveling Arkansas Professional Pathways (TAPP), not at the Foundational levels emphasized in Better Beginnings. Although findings are mixed as to whether a particular level of specialized college education will enhance teaching practices and child outcomes, research is in general agreement that college courses in ECE are more likely than workshops or in-service trainings to equip teachers with the knowledge and skills to create developmentally appropriate environments and to interact with children in ways that will promote their development. There is general concurrence that short, one-day workshops, such as those related to specific topics in Better Beginnings or even longer workshops that lack a fixed curriculum and are offered at a large number of sites without being customized to each group of participants are not considered likely to be effective. We recommend that professional development efforts continue opening avenues for college credits and that required state-sponsored trainings include mentorship components where individualized feedback or onsite assistance is provided. As an intermediate step to requiring more college credits, proposed higher levels of Better Beginnings could require longer trainings, such as Pre-K Early Literacy Learning in Arkansas and Pre-K Social Emotional Learning for Young Children. These trainings already exist, meet more of the research-based criteria for effective trainings, and deal with teacher-child interactions. Thus, they are more likely to produce positive change in child outcomes.

6 LEARNING ENVIRONMENT

Developments in neuroscience demonstrate that children begin learning at birth, and early caregiving environments and practices may enhance or deter learning in different domains of development (Shonkoff & Phillips, 2000). This is a shift away from the longer-held common societal belief that the primary function of non-maternal care is to attend to children's physical needs, such as food and rest, and to protect the them from bodily hazard. We know that even in circumstances where physical needs are well attended, a lack of cognitive and social-emotional stimulation can do long-term harm to children. Studies of children in Eastern European orphanages offer an extreme demonstration. After extended periods in institutions that offer consistent physical care but no emotional, physical, or cognitive stimulation, children who were otherwise born healthy are observed to have long-term impairments in social-emotional and cognitive development (Fries & Pollak, 2004; Hunt, 1998; Kadlec & Cermak, 2002; O'Connor, Rutter, Beckett, Keaveney, & Kreppner, 2000).

BB's Learning Environment section evaluates the level of planning and intentionality programs devote to the spectrum of developmental domains in childcare through four components embedded within the Learning Environment section: planning and curriculum, portfolios, interest centers, and developmentally appropriate physical activities. The amount and quality of research conducted for each of these components varies within the three types of care, and findings related to the after-school environment are markedly different.

We classify the learning environment standards as a structural measure because the standards specify *what* tools are to be present, but they do not specify or observe *how* the tools are used to facilitate child development or peer and teacher-child interactions. One item, the planning and implementation of developmentally appropriate physical activities (2.C.3.), may be an exception. If programs are to be observed or inspected for compliance, this item would be considered a process measure. However, we found no information as to how this item will be verified in the BB regulations or Toolkit materials.

6.1 PLANNING AND CURRICULUM

BB Level 1 calls for all types of programs to post a developmentally appropriate daily program schedule in each program area (1.C.1) and to develop and implement written daily plans for each group (1.C.2). The BB Guide further states: "A predictable routine is essential to children's optimal growth and development. An age appropriate daily schedule is the foundation for a predictable routine" (p.23). The language of 1.C.1 and the BB Guide appear to be redundant of minimum licensing requirements. Minimum licensing for centers states, "There shall be a written daily routine listing developmentally appropriate activities for children. The program shall offer alternating periods of active play and quiet times throughout the day" (400.1). This similarity is also be found in school-age and family care regulations. With careful reading, it seems that the difference is that minimum licensing is only required to print the schedule, but BB providers should implement the schedule. Because quality rating standards should clearly exceed those found in licensing, our evaluation team would recommend further clarification for the language of these Level 1 Learning Environment items.

Although we found evidence that routines at home predicted success in Head Start (Keltner, 1990), we did not find any studies assigning generic daily routines or planning as variables in analysis of center-based care. Results of secondary data analysis of NICHD, CQO, NCEDL 11-State Prekindergarten Evaluation, and Head Start FACES data revealed that the ECERS scale describing program structure, which includes schedule, free play, group time, and provisions for children with disabilities, had a modest correlation with child outcomes, albeit not as strong as the interactions scale (M. Burchinal, Kainz, & Cai, in press).

In school-age programs staff members are more engaged when there is a "good pace to their well-organized schedule," and youth are more engaged in "well-organized programs with clear routines" (Miller, 2005). A group of 96 after-school projects in New York City operated by The After-School Corporation (TASC) went through a 4-year external evaluation focused on program effects on participating students. The student sample totaled 52,000 after-school participants and 91,000 students who were enrolled in TASC host schools but were not participating in the after-school programs. In TASC projects where the site coordinator required project staff to submit lesson or activity plans, participants made greater gains in math and reading than in programs where lesson plans were not required (effect size of 0.14 in Math and 0.17 in English Language Arts). The evaluators speculated that the preparation and review of written plans occurred mainly in projects where student learning was a high priority (Reisner, et al., 2004).

It is likely that the quality of planning is more important than the regularity in which it occurs. Planning that does not take into account what is known about developmental processes or flex to the needs of individual children could still result in deleterious outcomes for children, even if planning is frequent and consistent.

We found more evidence to support generic planning in the family day care setting. The learning environment of family day care tends to be weaker in comparison to that of centers. The National Study of Child Care for Low-Income Families (Layzer & Goodson, 2006) and the Observational Study of Early Childhood Programs (Bronson, Tivnan, & Seppanen, 1996) observed low-income preschool children in centers spending a little more than 30% of their time in goal-directed activities. Preschool children in family care, however, spent only 18% of their time in similar activities and were much more involved with television or routine or informal activities, and more than 60% of the homes studied had no learning activities during the morning hours (Layzer & Goodson, 2006). Creative play with objects and adult use of language to teach or to elicit complex responses from children were rare. Providers in the Study of Children in Family Child Care and Relative Care (n=226) who planned activities were more likely to be rated as sensitive and observed to be responsive to children's needs that those without planned activities. Levels of sensitivity or warmth were positively associated with children's secure attachments to their providers, and responsiveness was associated with more complex play and more play with objects. (Ellen Galinsky, et al., 1994).

At BB Level 2, providers must demonstrate how the daily plans address all areas of development as defined in the Arkansas Early Childhood Education Framework or the Arkansas Framework for Infant and Toddler Care, or in the case of the school-age standards, the Developmental Assets concept (2.C.2). This standard begins to approach better practices in early and school-age care. Many states have developed similar frameworks and have incorporated them into state rating systems. These frameworks are often influenced by NAEYC's guidance tool, *Developmentally Appropriate Practice in Early Childhood Programs*

(DAP, Copple & Bredekamp, 2009). DAP is research-based individualized instruction appropriate to the child's age and developmental stage. It establishes suitable goals to stimulate a child's progress in all major developmental domains. Children who attend programs considered to use DAP develop more appropriate social skills and better academic scores (Huffman & Speer, 2000; Marcon, 1999), demonstrate fewer antisocial behaviors, and have higher expectations for success into their teens and early adulthood (Schweinhart & Weikart, 1997). DAP practices are woven into Arkansas Frameworks, so we would anticipate child outcomes would improve if programs implement the recommended practices. However, if written documentation is all that is required for BB, the indicator will be less reliable than a process measure that observes or verifies the use of Frameworks in the presence of children.

The Developmental Assets used for the school-age standards are described as a synthesis of the relationships, opportunities, and personal qualities that young people need to avoid risks and to thrive. Based on the Search Institute's research, the more assets youth have, the less likely they will engage in high risk behaviors such as problem alcohol use and violent behavior or have school problems.[38] The Harvard Family Research Project affirmed that quality standards and measures should have an asset-based youth development approach (Westmoreland & Little, 2006). Similar to our findings with the SF model, although the Developmental Assets are based on research in youth development, we did not find studies validating the implementation of this particular model with documented child outcomes.

BB Level 3 providers in centers and in homes advance beyond generic planning to a more formal method of planning, the written curriculum. There is much more evidence that curriculum use is an indicator of overall quality with a stronger relationship to child outcomes than generic planning. The 2006 evaluation of Pennsylvania's quality rating system states, "Child care centers who reported that they used a standardized curriculum (N=126) had significantly higher ECERS scores on all scales with the exception of the Interaction scale"(Barnard, et al., 2006, p.17). Family child care homes that used a curriculum also scored significantly higher on all FDCRS scales (mean score 4.96 with curriculum versus 4.34 without). These results confirmed findings from a prior study in Pennsylvania in which use of a curriculum related to overall quality ($F=28$, $p < .0001$) (Fiene et al., 2002).

The Midwest Child Care Research Study team surveyed and observed a random sample of 2,000 infant/toddler and preschool child care providers in centers and in family day care in Iowa, Kansas, Missouri, and Nebraska. Based on their findings, the team developed an index of 14 features of child care associated with global quality as measured by ECERS-R, ITERS, FDCRS, and CIS (Helen H. Raikes et al., 2006). Although child outcomes were not studied, use of curriculum predicted the difference between poor and minimal quality for all types of care. This result coupled with the fact that DCCECE offers curriculum technical assistance and provides the Adventures in Learning[39] curriculum free of charge, suggest that standards promoting curriculum use are best placed at lower rather than higher levels of a quality rating system.

Relying on program self-reported planning and use of curriculum will be less accurate than using independent observations of curriculum use. High proportions of programs participating in Missouri's pilot reported use of formal curriculum, but observations revealed great variation in implementation

[38] http://www.search-institute.org/content/what-are-developmental-assets
[39] http://www.arkansas.gov/childcare/adventures/

(Thornburg et al., in press). This find prompted the state to incorporate independent observations of curriculum using a measure that has better validity than provider report, ECERS-E (K Sylva, Siraj-Blatchford, & Taggart, 2003), an extension of ECERS-R. It rates a program's pedagogy as well as the curriculum and program resources and activities that support child development using four curricular subscales: Literacy, Mathematics, Science and Environment, and Diversity. ECERS-E assessors are instructed to give credit only when observations on the day of a visit provide evidence that the planned activities have taken place. Teachers' lesson plans do not provide sufficient evidence "as there is no way of knowing how fully plans are adhered to and how they will be interpreted," (p.17). The predictive validity of this measure was supported by the Effective Provision of Preschool Education (EPPE) project that studied nearly 3,000 children in the UK (Kathy Sylva et al., 2006). Controlling for child, parent, family, preschool and home characteristics, ECERS-E scores predicted improvements in pre-reading scores, early number concepts, and non-verbal reasoning between the ages of 3 and 5.

There is evidence that curricula facilitate children's development in different ways. Most curricular comparisons studies are small, and outcomes differ from one curriculum to the next because each has different goals (National Research Council, 2001). Chambers, Cheung, and Slavin (2006) used a quantitative synthesis approach in their review of preschool instructional methods. Programs that adopted a direct instruction curriculum or a cognitive-developmental curriculum generated more positive outcomes among at-risk 3-5 year-olds than programs using a traditional nursery school approach. A summary of their findings (Table 6-A) shows that short-term cognitive outcomes were greater in direct instruction programs, but cognitive developmental programs produced long-term social and short-term educational benefits.

Table 6-A Comparative Review of Curricular Approaches

Program Type	**maturational/developmental interaction**	**academic/direct instruction**	**cognitive developmental**
Goals	Enhance self-esteem, independence, social development	promote cognitive development	promote cognitive and social development
Characteristics	Unstructured play with minimal teacher interaction	highly scripted program where activities are initiated by the teacher	planned activities often initiated by the child but done in tandem with teacher
Example	traditional nursery schools	Demonstration and Research Center for Early Education (DARCEE)	High/Scope, Dialogic Reading, Montessori, Curiosity Corner
Outcomes	did not compare favorably to other program types on any measure	produced more immediate, short-term cognitive outcomes than other program types	produced more short-term educational and long-term social adjustment outcomes than other program types

The authors of this synthesis note that Project Approach, Reggio Emilia, and Creative Curriculum did not appear in their review because studies on programs using these curricula did not match their requirements for rigorous design or program criteria for inclusion (occurring in a pre-kindergarten group setting, meeting for at least 12 weeks for three hours per week, and clearly articulating the details of the intervention).

In the case of Reggio Emilia, this exclusion is not an accident. The Reggio Emilia community resists being categorized as a curriculum because it indicates rigidity or a fixed plan in opposition to the idea of dynamism or emergent practices (Goffin, 2000). As New writes, "American educators have been perplexed at the lack of empirically derived data with which to validate Reggio Emilia's practices, and yet Reggio Emilians are persistent in their refusal to participate in this positivist tradition that has played such a strong role in determining U.S. educational policies and practices. Rather, the reliance, in Reggio Emilia, on teacher reflection and documentation of their work with young children as their primary means of research—a tradition that has often been criticized for its socio-political orientation" (1998, p. 278).

The Creative Curriculum, on the other hand, does have child outcomes data available. Children who received The Creative Curriculum for Early Childhood in the Sure Start programs for low income military families made statistically significant gains in pre-math and emergent literacy skills within the school year (Abbott-Shim, 2000). In a comparison study, use of integrated curriculum, most notably High/Scope curriculum in Head Start programs, resulted in statistically significant gains in letter identification, more cooperative behaviors, and fewer behavioral problems (Nicholas Zill et al., 2003). Although The Creative Curriculum also provided children with increases in some areas, the increases were not as strong.

Conclusions that we can draw from curriculum validation studies are limited by the lack of control groups in many studies and by inadequate observation and description of the curricula and how they were implemented. For example, many studies do not report on how extensively teachers were trained, how much time children were exposed to the curriculum, and whether teachers presented the curriculum with fidelity to the original model.

Looking at curriculum and structure in after-school programming, some studies have reported good results with the use of curricula, but with the caveat that well-trained staff and coordination with the school day are essential. The TASC evaluations looked at whether staff communicated goals and expectations or organized activities well. They found that quality programs consistently offered participants opportunities for skill building and mastery. The 10 high-performing TASC programs used formal curricula for literacy activities and for strengthening peer relationships and teaching conflict resolution. They provided intensive and ongoing training specific to the curricula used (Birmingham, et al., 2005).

BB school-age standards require that curriculum plans be linked to Arkansas Department of Education K-12 Frameworks. The intent is for school-age programs to plan activities that complement what is being taught during the school day. We found support for this standard, but with an additional requirement – that after-school programs build relationships with the schools they support. In discussions of quality programming, we found that more attention was paid to coordination with the curricula being used at the school than with the use of specific curricula. The TASC evaluators considered "intentional relationship building" to be a strong component of quality programs, and those programs worked closely with the schools throughout the year. The evaluators also noted that some of these programs aligned their after-school activities with curricula used by participants' schools (Birmingham, et al., 2005). However, fidelity to an academically-focused, school-adapted curriculum may be challenged by the limited amount of time children spend in after school settings (SEDL Research Consortium, 2008).

The National Partnership review found that in quality after-school programs, academic activities are linked to state standards and to school-day expectations, particularly in literacy, math, and science.

Quality programs set specific goals for academic achievement based on students' academic data, and they communicated with school staff regularly through sharing goals and providing frequent progress reports (Jordan, et al., 2009). The evidence is clear that when building a quality after-school program, strong connections with schools are essential (Beckett, et al., 2001; Birmingham, et al., 2005; Bodilly & Beckett, 2005; D. Vandell, Reisner, & Pierce, 2007). After-school programs are not intended to be school, but they are intended to complement the school day, and an ongoing partnership can strengthen both, particularly if improvement in academic performance is a desired outcome.

6.2 PORTFOLIOS

The portfolio, a collection of a child's work or a teacher's documentation of a child's behaviors and activities over time, potentially enables providers to assess learning within the context of class curriculum and to make goals for individual students. For this reason, portfolio use is required for Level 3 providers (3.C.2). The existing literature on portfolio assessment is primarily definitional and propositional. Numerous models for portfolio assessment have been presented, but there is little empirical evidence to show that portfolio use independent of an evidence-based curriculum improves quality of instruction, assessment, or child outcomes. Ideally, portfolios focus on change, individualized instruction, teacher and child reflection, and sharing information with others (Gullo, 2006), areas where standardized testing falls short. Providers using DAP "document children's learning and development, including in written notes, photographs, audio recordings, and work samples. They use this information both in shaping their teaching moment by moment and in planning learning experiences," but there may also be cases where, "Assessment results (observation notes, work samples, etc.) go straight into a folder and are filed away. They are not reflected on to inform teachers how to help or challenge individual children" (Copple & Bredekamp, 2009, p. 180).

Implementation of system-wide portfolios for school-age children has been problematic. For instance, the Rochester, New York school system implemented a portfolio system for language arts in kindergarten to second grade. Their validation study showed that teachers could reliably evaluate portfolios for classroom purposes, but there were problems with using external raters and with using the assessments for accountability purposes (Supovitz, MacGowan, & Slattery, 1997). More success has been shown in the use of observational checklists, such as Meisel's Work Sampling System. The version adapted for 3- and 4-year-olds, Work Sampling for Head Start, was shown to be a reliable predictor of early mathematics and reading performance (Meisels, Xue, & Shamblott, 2008).

Likewise, we did not find much discussion of the use of portfolios in the after-school literature. In one report, portfolios are mentioned as a possible source of evaluation data (Committee on After-School Research and Practice, 2005). In the UK, portfolio assessments are used as part of a standards-based recognition scheme to identify high quality out-of-school hours learning programs and to reward them with a special quality credential (Fordham, Boyd, & Apicella, 2004).

Within the school-age education and child development literatures, portfolios have primarily been used as an assessment tool for students with disabilities, and this has been the practice in Arkansas. Legislation that requires states to set up alternate assessment plans for students with disabilities, including the Individuals with Disabilities Education Act and No Child Left Behind Act, led to this use of portfolios as an

alternative to standardized testing for students with disabilities in the school setting. Kentucky was the first state to develop and implement a state-wide system using portfolios and a structured scoring rubric to assess students with disabilities. Other states have followed suit, though there is wide variation in practice across states and lack of evidence of the technical quality of alternative assessment compared to conventional assessment tools (i.e. standardized tests). Instructions for the use of portfolios in this manner for Arkansas school children are detailed and specific.[40] BB exempts after-school programs from developmental assessment, as this process is already occurring through a student's home school, so requiring portfolio assessment may be redundant as well as time consuming for staff. For portfolio assessment to be useful, staff would need to be trained in appropriate contents, how to develop portfolio tasks, and how to apply criteria for assessing students' work (Herman & Winters, 1994).

6.3 INTEREST CENTERS

Classroom materials and activities may vary in their effect on language, cognition and socio-emotional development, and it is a reasonable to expect a greater variety of materials and activities to stimulate child interest and development in a wider range of domains. BB programs must have a minimum of two clearly defined interest centers for Levels 2 and 3 (2.C.1; 3.C.1). The literature on interest centers is similar to the literature on portfolios in that there are many proposed models but little evidence of their affect on child outcomes.

The NCCSS team found that when classrooms received high ratings in activities as measured by ECERS and ITERS, the children (n=414) were likely to orient themselves to adults and peers. Social orientations to adults and peers, in turn, predicted greater competence in peer socialization (C Howes, et al., 1992). This study contained a generous sample size, which is not typically the case in studies examining play.

There is a small body of evidence suggesting that the structure, specifically the number and physical arrangement, of interest centers is indicative of global quality, and to a lesser degree, more competent play and social skills. Children observed in rooms with partitioned play areas had better peer & verbal interactions, fantasy play, and associative-cooperative play (Field, 1980). Observations of 4- and 5- year-olds in two centers and one family day care showed that higher child-to-activity-area ratios related to more off-task behavior and lower likelihood of constructive play (Kantrowitz & Evans, 2004). However, the small sample size of this study limits conclusions. The 2003 Oklahoma QRIS validation study, which included a larger sample, found that the presence of five interest in centers was a significant predictor of ERS quality, but child outcomes were not examined (Norris, et al., 2003). The most convincing evidence comes from the nationally representative NICHD investigations. Children in family and center-based care had better language comprehension and short-term memory when their caregiving environments had a greater variety of well-organized materials (NICHD Early Child Care Research Network, 2003).

In a different vein, play research suggests that what is contained in interest centers and, more importantly, how the materials and activities are presented by teachers is more predictive of child outcomes than the number or physical layout of centers. Outcomes are most often in the social-

[40] See http://www.arkansased.org/testing/pdf/assessment/th_alt_gr3-8_11_061410.pdf.

emotional realm but also occasionally in cognition, as in a study where Kontos and Wilcox-Herzog (1997) found that opportunities for free play with art, blocks, and dramatic play activities, classified as "high-yield" activities, were positively correlated with cognitive competence among 114 4-year-old children. Results were complicated by the presence of a teacher, which was a negative predictor of child cognitive competence. Social competence, on the other hand, was only predicted by peer and teacher presence, not by the activities. Studying the effect of similar activities, but analyzing on the classroom rather than the individual level, Bronson and colleagues (1996) found that the amount of time children spent in block play related to the proportion of tasks that children successfully completed; dramatic play did not relate to any child outcomes. Creative play activity was significantly positively associated with cognitive activity among European-American infant-toddler children in nonsubsidized center groups. Positive associations with other groups did not reach significance (C. Howes & Smith, 1995).

We did not find a discussion of interest centers in the school-age literature. After-school programs are located in a variety of settings, including child care centers, schools, and community programs. The literature does give credence to variety, with a focus on engaging youth so that they are responsive to staff interventions and skill-building opportunities.

One team found that among 18 quality indicators, providing a sufficient variety of activities was one of only three indicators to have strong support in the research literature (Bodilly & Beckett, 2005). Adequate space, age-appropriate activities and materials, and a sufficient variety of activities were part of a set of 20 good practices identified in a Rand study (Beckett, et al., 2001). Offering youth choice within a variety of well-implemented activities was found to be related to a program's ability to attract and retain participants (Goldsmith, Arbreton, & Bradshaw, 2004). Like many after-school programs, the high-quality TASC programs did not focus primarily on academics, but offered diverse enrichment opportunities. For many participants, this was their first exposure to learning in areas such as dance, music, art, and organized sports (Birmingham, et al., 2005). And a review of 53 quality after-school programs in 33 states found that programs where students had more successful academic outcomes maintain a balance between academic programming and a wide variety of enrichment activities (Huang, Cho, Mostafavi, & Nam, 2008; Jordan, et al., 2009).

A study of after-school programs in the Madison, Wisconsin area found that when programs offered a variety of different activities, staff were observed to have more frequent positive interactions with the children, and activities appeared to be more age appropriate. The positive ratings and interactions were not found to be associated with the presence of any single activity, but with the fact that children were offered a variety of different types of activities (R. Rosenthal & Vandell, 1996).

6.4 PHYSICAL ACTIVITIES

Item 2.C.3, "Staff plan and implement daily developmentally appropriate physical activities for all children," is one of the few process measures within BB, and it may have the potential to improve child outcomes in all child care settings by influencing long-term habits and potentially decreasing the likelihood of being overweight among enrolled children.

For most BB standards, we find less data for school-age care than for young children, but this standard is an exception. Eleven out of 14 published studies analyzing data from about 58,000 students found that

regular participation in physical activity is associated with improved academic performance (National Institute on Out-of-School Time, 2009). Higher levels of learning gains by participants were reported for TASC projects that offered extensive opportunities for fitness, sports, and recreation (Reisner, et al., 2004).

Moderate-to-vigorous physical activity (MVPA) decreases dramatically from age 9 to age 15 (Nader, Bradley, Houts, McRitchie, & O'Brien, 2008). After-school programs for middle-school students that promoted physical activity could potentially have a positive impact on the health of those students. Given the high obesity rate and the fact that regular physical activity is associated with real health benefits (Centers for Disease Control and Prevention, 1997; U.S. Department of Health and Human Services, 2001b), the need for increased physical activity for children is clear, but physical education programs conducted with elementary children have yielded varying degrees of success (Stone, McKenzie, Welk, & Booth, 1998; Wallhead & Buckworth, 2004).

Less is known about physical activity and fitness or health outcomes in early childhood. It is presumed that young children who are more active will have better cardiovascular health, will be leaner, will have stronger bone mass, and that habits established in childhood will track into adulthood (Boreham & Riddoch, 2001). Burdette and Whitaker propose that we remember that physical activity for young children is play, and that this play likely contributes to socialization, cognitive function and attention (2005). However, empirical evidence in the realm of early childhood fitness is sparse compared to the literature for adults and school-age children. Although the medical field is clear that children need more physical activity, there is not yet clear consensus on how much more is needed to improve long-term health and habits (Timmons, Naylor, & Pfeiffer, 2007).

We tend to assume otherwise, but young children have been observed to be highly sedentary even during free play (Timmons, et al., 2007) and exhibit inadequate rates of MVPA, which can lead to being overweight. “Overweight preschoolers are at significantly increased risk for child and adolescent obesity, and they are more likely than their nonoverweight counterparts to experience significant short- and long-term health problems such as hyperlipidemia, hypertension, insulin resistance, respiratory problems, and orthopedic complications. In addition, the adverse social consequences of childhood obesity might have long-lasting effects on psychological well-being and economic mobility” (Trost, Fees, & Dzewaltowski, 2008, p. 88). A review of interventions to increase physical activity and to improve fitness outcomes among young children presents mixed findings (Timmons, et al., 2007). The common dosage among interventions was 20 minutes per day, 3 days per week. On the whole, there was little change in overweight status, but there was tentative evidence that these programs improved other aspects of health, such as bone density and motor skills.

Outcomes are in part a product of the quality of the training teachers receive. For instance, one intervention sought to increase MVPA by incorporating physical activity into existing curricula. The first four weeks of implementation produced no increase in MVPA, which the authors attributed to teachers having only been given one training session. After teachers were given additional training that incorporated on-site evaluation and individualized feedback, MVPA outcomes improved in the last four weeks of the study (Trost, et al., 2008).

The type of play environment and materials provided for children can have an effect on the amount of physical activity performed. Children are most active outdoors (Burdette & Whitaker, 2005) and on larger playgrounds with less fixed and more portable playground equipment (Marsha Dowda, et al., 2009).

Choosing the right activities or the right kind of equipment will play a role in having more active healthy children, but raising a center's global quality and teacher qualifications is likely to have an effect as well. Dowda's team observed that when ECERS quality is higher or when teachers have more college education, children have more MVPA (2004).

6.5 KEY STATES COMPARISON

BB has more overlap with other states in curriculum use than the other components of the Learning Environment. Many states use the same standards for early childhood and school-age care, and this area is no exception. When specific standards for school-age care are contained, they are included in the summary. Otherwise, early childhood and school-age standards should be assumed to be the same.

6.5.1 DAILY PLANNING

BB has little overlap with other state quality rating standards in the areas of daily schedules and written planning (1.C.1 and 1.C.2). Oklahoma has comparable standards in that rated centers must have a daily schedule that includes a balance and variety of activities, and classrooms must follow weekly developmentally appropriate lesson plans. Pennsylvania requires programs to maintain learning standards for a Star 1 rating and to apply learning standards in planning for Star 2.

6.5.2 CURRICULUM

More of the key states include provisions for curriculum than the other components of the BB Learning Environment category.

- **Missouri** has the most stringent standards for curriculum. Its quality rating category, Intentional Teaching, employs a standardized measure, ECERS-E, an extension of ECERS that includes four subscales addressing curriculum (K Sylva, et al., 2003). The pilot rating system did not include this scale, but rather only asked whether programs had a written curriculum. The state adopted ECERS-E later as a more discriminating way to assess the quality of the curriculum and whether it was being used well (Thornburg, et al., in press). Because designers did not find a comparable standardized measure addressing curriculum in infant-toddler and school-age programs, they developed separate observations tools for measuring curriculum provision in these programs.

  The Business/Administrative Practices category additionally requires either A.) staff training and incorporation of Missouri Standards into lesson plans or B.) individualized goals for children.

- **North Carolina** centers can implement approved developmentally appropriate curriculum to earn an additional quality point.

- **Ohio** Step 2 programs utilize a written, evidence-based comprehensive curriculum aligned with state guidelines and content standards to address broad range of developmental areas.

- **Pennsylvania** Star 1 sites maintain copies of state learning standards. For Star 2, standards are used to plan and document learning. For Star 3, a formal curriculum is required. For Star 4, sites must crosswalk the curriculum and assessment tools with the state learning standards.

6.5.3 INTEREST CENTERS

Two states, North Carolina and Oklahoma, have standards for interest centers. In North Carolina, centers earn a quality point for having an enhanced space, which may include an increased number of activity areas. Oklahoma centers must have five interest areas for children 2 years and older (blocks, dramatic play, manipulative play, art, and book/reading) to attain 1 Star and two additional areas (math and science/nature) to attain 2 or 3 Stars; two learning centers must be available outdoors. As described earlier, the presence of five interest centers was a predictor of global quality in Oklahoma's system evaluation. For Oklahoma family day care, the term "interest area" is not used, but programs must still provide access to the five categories of activities required for 1-Star center ratings in order to receive a 2- or 3-Star family rating.

6.5.4 PORTFOLIOS

Pennsylvania is the only state that specifically mentions portfolios for center-based and school-age providers. Family providers in the Star 2 level must establish a system to document observations and the developmental progress of each child, and the portfolio is mentioned as one example of a tracking method. Portfolios are not mentioned in the state's center standards, but there are clear provisions for tracking student progress.

Three of the key comparison states have developed provisions that address emerging best practices in assessment of child development.

- **Ohio** Step 2 for family- and center-based care requires all children except school-age receive a developmental screening within 60 days of enrollment. Referrals, if needed, are completed within 90 days. Results are formally communicated with families. Step 3 children are assessed systematically utilizing formal and informal methods to inform intentional teaching and the sharing of progress with families.

- **Pennsylvania** has a Child Observation category for home and center-based care. Family providers must establish a system for observations for 2 Stars. For Star 3 in family care and Star 2 in center care, observations are required within 45 days of enrollment and once per year thereafter. Higher levels require more frequent observations and parent conferences with results used for individualized planning and referrals. Centers at Star 3 must also report child outcomes using a formal work sampling system.

- **Missouri** uses the ECERS-E (Kathy Sylva, et al., 2006) to determine how staff members plan their activities and instruction to meet the needs of individual children.

6.5.5 PHYSICAL ACTIVITIES

No key comparison state matches were found for the planning and implementation of developmentally appropriate physical activity.

6.6 CROSSWALKS

6.6.1 HEAD START

Head Start requires "planning for routines and transitions so that they occur in a timely, predictable and unrushed manner according to each child's needs" (45 CFR 1304.21) and requires the implementation of curricula that meet specific criteria (45 CFR 1304.22). Portfolio use is not specifically required, although it may occur as part of a chosen curriculum. Centers are required to provide indoor and outdoor space and equipment and adult guidance to promote active play and the development of gross and fine motor skills (45 CFR 1304.21).

6.6.2 NAEYC

NAEYC accreditation standards extensively exceed BB and ERS where the quality of the learning environment is concerned. Standard 3.D. specifies how to organize daily schedules, but it does not specify that they be written or posted. "The curriculum guides the development of a daily schedule that is predictable but flexible and responsive to individual needs of children" (2.A.7). NAEYC does not mandate a number of interest centers but rather provides extensive guidelines for a rich variety of materials and equipment (2.A., 3A, 9.A.04). Standards in 3.G.10 specify how teachers should use learning centers.

Portfolios are not a part of NAEYC accreditation standards, but they are mentioned as part of DAP practice in other NAEYC publications. Providers using DAP practices "document children's learning and development, including in written notes, photographs, audio recordings, and work samples. They use this information both in shaping their teaching moment by moment and in planning learning experiences."

Standard 2.A. outlines essential characteristics for curriculum and scheduling. Subsequent sections of this section detail how opportunities of learning must be presented in these developmental areas: social–emotional, physical, language, early literacy, early mathematics, science, creative expression and appreciation for the arts, health and safety, and social studies.

NAEYC standards related to physical activity are 5.A.06, specifying outdoor play, and 2.C.04, indicating that programs should provide opportunities to engage in large motor activities.

6.6.3 NAFCC

NAFCC accreditation standards emphasize a well-organized environment (2.6) with a variety of learning materials to promote child learning. Lists of materials for motor development, literacy, art, math, science, dramatic play are suggested (2.23-2.43). Other items specify types of activities that make use of these materials and in some cases recommend frequency or duration of use (3.52-3.78).

Similar to BB, family teachers are expected to maintain a consistent sequence of daily events (3.17). The standards do not specify written planning or curriculum use but do emphasize intentional, developmentally appropriate and individualized activity (1.12, 3.1, 3.3-3.15). Programs must show that they engage in series of activities that support social-emotional development (1.20-1.22, 3.36-3.51), physical development (3.52-3.54), cognitive development (3.55-3.58), language and communication (3.59-3.63), literacy (3.64-3.66), math and science (3.67-3.69), and creativity (3.70-3.78).

NAFCC does not have provisions for portfolios but expects teachers to gather information and observe each child to set developmental goals and to customize environments and activities (3.4). "The provider plans some activities building on the needs and interests of the children. She is flexible in adapting the plans" (3.5).

6.6.4 COA

COA after-school programming standards, Section ASP-PS 5, cover the daily schedule and activities of a school-age program. Section 5.01 provides guidelines for a daily schedule that is flexible, provides stability without being rigid, and facilitates transitions smoothly. The COA focuses on a wide variety of activities (5.02) that are appropriate for the age, abilities, and interests of participants (5.03) yet offer freedom of choice among program activities and the right to opt out of a particular activity (5.04). The standards do not discuss curriculum or portfolios. Sections ASP-PS 6 and 7 cover the indoor and outdoor environment with guidelines for a program setting that is welcoming and engaging. Interest centers are not discussed, but the space must be arranged so that program activities can be well accommodated, easily accessible to children, and suitable for a wide variety of activities. Standard 7.02 requires that youth have frequent, regular opportunities to participate in outdoor activities.

6.7 CONCLUSIONS

Better Beginnings designers intended each level of the Learning Environment category to indicate increasingly complex levels of intentionality, starting with routines and working up to daily planning. Although a quality rating system should address levels of quality separate from and above that expressed in a state's minimum licensing regulations, there is some redundant language in regard to routines at the lowest Better Beginnings Level. We suggest clarification for these standards.

With the exception of curriculum use, item 3.C.3, we found scarce evidence that child outcomes are tied to items for the Learning Environment section. Absent of a curriculum, general planning was not associated with enhanced child outcomes in center-based care, but one study suggested that family care teachers who make daily plans have better interactions, which are, in turn, associated with more secure attachments and increased complex play. For school-age care, one study found that daily plans may be associated with better academic outcome outcomes.

The number of interest centers in a program has been linked to global quality but not to child outcomes. However, there is evidence that the kind of materials placed in the interest centers and the manner in which the teacher guides children to use these materials are related to interactions and to

child cognitive and social competence. For school-age children, given that materials are adequate and accessible to youth, a sufficient variety of activities may be more salient to outcomes.

Portfolios may be used for assessment to determine atypical development and/or to individualize care and instruction. Validation studies have identified problems with using portfolios for assessment purposes. We did not find studies examining whether the use of portfolios to individualize instruction helps children. We recommend that standard 3.C.2 define the intent of portfolio use. As currently written, this item may not discriminate between programs that use portfolios with a developmentally appropriate intent from those arbitrarily collecting products or recording behaviors without further reflection. For school-age children, the Better Beginnings intent should also be clarified. If portfolios are to be considered a tool for developmental assessment, this standard contradicts the exemption of school-age programs from developmental assessment. If the intent is for portfolios to track individual progress and plan further programming, then training will be necessary to insure staff intentionality and consistency with a method that will be unfamiliar to many school-age providers.

If programs are to be observed or inspected for compliance in planning and implementing developmentally appropriate physical activities, we would classify this measure as one of the few process measures in the system, with greater predictive power for outcomes. However, we found no information as to how this item will be verified in the Better Beginnings regulations or Toolkit materials. The school-age literature presents strong evidence that physical activity in care is associated with better outcomes. Literature related to physical activities and fitness for ECE is sparse and suggests that physical activity will have to increase above the amounts typically introduced in interventions and/or be combined with dietary education and parental outreach to curb the current trajectories for children's overweight status.

The item of the Learning Environment section with the strongest observed ties to child outcomes is 3.C.3, which verifies the presence of a current written curriculum plan. Use of a curriculum has been linked to global quality and to child outcomes more than any other item in the Learning Environment section. The instructional method chosen will affect children in different ways, with the traditional nursery school approach being the least likely to produce academic, cognitive, and social outcomes. In light of evidence that curriculum use is a distinguishing characteristic between poor and minimal care and a lack of evidence that items in the lower Better Beginnings Levels have connections to child outcomes, we suggest that curriculum is foundational to quality and is misplaced at the highest Level. For school-age children, we would add that coordination with participants' schools is equally important to outcomes. DCCECE does have pre-approved and recommended curriculum, such as Creative Curriculum, High/Scope, and Reggio Emilia and DCCECE technical assistants specializing in curriculum, should guide individual programs in curriculum choice and implementation to maximize child outcomes.

7 ENVIRONMENTAL ASSESSMENT

Of all BB components, Environmental Assessment stands out as the strongest because it utilizes the Environment Rating Scales (ERS):

- *Early Childhood Environment Scale, Revised Edition* (ECERS-R;Harms, Clifford, & Cryer, 1998),
- *Infant/Toddler Rating Scale, Revised Edition* (ITERS-R; Harms, Cryer, & Clifford, 2003),
- *Family Child Care Environment Rating Scale, Revised Edition* (Harms, et al., 2007),
- and *School-Age Care Environment Rating Scale* (SACERS; Harms, Vineberg Jacobs, & Romano White, 1996).

These instruments measure the quality of care provided for children in a variety of settings, and have become the most widely used quality measures in ECE practice and research in the past 30 years. Empirical evidence has validated the relationship of ERS quality to child outcomes in child care research around the world, although in this section we explain that findings are not always consistent and are often modest in strength. Use of these scales lends strength to the QRIS in that they rely more on information collected by an independent observer than on teacher and administrator reports.

Each ERS is global measure combining items rating structural aspects of the program—for instance, the physical layout of the space or staff qualifications—with observations of processes that directly involve children, such as instructional activities or discipline methods.

The most commonly used ERS, ECERS-R, is a standardized instrument designed to assess the quality of center-based environments serving children between the ages of 2.5 to 5 years. It contains 7 subscales with 43 items. ECERS-R has a 7-point rating scale: 1 (adequate), 3 (minimal), 5 (good), and 7 (excellent). "Levels of program quality are based on current definitions of best practice and on research relating practice to child outcomes. The focus is on the needs of children and how to meet those needs to the best of our current understanding" (p.2).

The precursor to ECERS-R, ECERS, had a good history of predictive validity. It was initially published in 1980. Over the next 15 years, the ECE field expanded its definition of quality to incorporate inclusion of disabilities and cultural diversity. The ERS authors incorporated these elements into the revised scale and refined their measurements based on what they had learned through the development of three other sister scales, *Family Day Care Rating Scale* (FDCRS; R. M. Clifford, Harms, & Harms, 1989), ITERS, and SACERS. The revision, ECERS-R, was field tested in 1997 with results that "revealed quite acceptable levels of inter-rater agreement at the three levels of scoring-indicators, items, and total score" (Harms, et al., 1998, p.3). However, they did not report on the validity of the new version but rather relied on its relation to the original version's validity. Sakai and colleagues later confirmed similar correlations and distributions between the two versions, while pointing out that both measures still fall short in addressing staff turnover and culturally sensitive practices (2003).

Perlman, Zellman and Lee (2004) attempted to replicate the ECERS-R psychometric properties presented by the scale's authors using data collected from 326 classrooms. Their factor analysis detected only one global factor, not the seven that the scale's authors found, and they found items had high correlations. The team then hypothesized that a shorter version of the scale might be used with similar results. They

randomly selected 3 subsets of 12 items from the original scales and found that the shortened version produced similar results as the long version. They also examined correlations between the full scale and regulatable features of the classrooms. The only relationship they found was between ECERS-R total scores and teachers' years of experience. There were no relationships found between ECERS-R and teachers' level of education or staff-child ratios, two regulatable features that are more often associated with better child outcomes. These results were consistent with prior analysis conducted by Scarr and colleagues (1994) using ECERS and ITERS observations in more than 300 classrooms. This study also found only one factor of quality and, likewise, got good results from shortened versions of the scales, but, unlike Perlman, found that the only characteristic related to quality was teacher wage.

There is considerable variation in how ECERS-R scores are used. For instance, researchers and policy-makers alter ECERS-R, usually in the interest of reducing observation time and cost. Ohio's quality rating system and some ECE studies omitted the Provisions for Parents of Staff in their overall score calculations (Donna M. Bryant, Maxwell, & Burchinal, 1999; Mashburn, 2008), as such items were presumed unlikely to affect child outcomes (Lower & Cassidy, 2007; Mashburn, 2008). Another common variation is to combine ERS scores with other measures to produce a composite score, as in a longitudinal study by Deater-Deckard, Pinkerton, and Scarr (1996) or as in the CQO Study (Helburn, 1995).

7.1 CENTER-BASED ERS RELATED TO CHILD OUTCOMES

ECE investigations have identified an array of statistically significant but small to modest relationships between ERS center quality and child development in the following domains:

- **language development and communication** (Donna M. Bryant et al., 2003; MR Burchinal, Roberts, Nabors, & Bryant, 1996; M. R. Burchinal, Peisner-Feinberg, Bryant, & Clifford, 2000; M. R. Burchinal & Roberts, 2000; L. Dunn, Beach, & Kontos, 1994; Carollee Howes et al., 2008; Kathleen McCartney, 1984; K. McCartney, Scarr, S., Phillips, D., & Grajek, S., 1985; Whitebook & et al., 1989);
- **cognitive development** (MR Burchinal, et al., 1996; M. R. Burchinal & Roberts, 2000);
- **literacy or pre-reading skills** (Mashburn, 2008; ES Peisner-Feinberg & Burchinal, 1997),
- **social-emotional competence, secure attachments and/or positive affect** (Hagekull & Bohlin, 1995; Hestenes, Kontos, & Bryan, 1993; Lamb, Hwang, Broberg, & Bookstein, 1988; K. McCartney, Scarr, S., Phillips, D., & Grajek, S., 1985; ES Peisner-Feinberg & Burchinal, 1997; Whitebook & et al., 1989);
- **counting and applied math** (Donna M. Bryant, et al., 2003; M. R. Burchinal, et al., 2000); and
- **moderate/vigorous physical activity** (Marsha Dowda, et al., 2009; M. Dowda, et al., 2004).

At-risk children are usually found to be more sensitive to the effects of global quality (M. R. Burchinal, et al., 2000; M. R. Burchinal & Roberts, 2000; Helburn, 1995; Mashburn, 2008; ES Peisner-Feinberg & Burchinal, 1997; E. Peisner-Feinberg et al., 2000).

Two of the largest national studies of child care, the National Child Care Staffing Study (NCCSS) and the Cost, Quality, and Outcomes Study (CQO), incorporated ECERS and identified associations between quality

and children's developmental outcomes. A third study, by the National Institute of Child Health and Development (NICHD), is one of the most methodologically sound longitudinal studies of child care across the U.S. to make associations between quality and outcomes. As such, we do discuss some NICHD findings in this section. However, center quality was measured with the Observational Record of the Caregiving Environment (ORCE; NICHD Early Child Care Research Network, 1996), not with ERS, and the reader should be mindful of the distinction when attempting to draw conclusions about the measurement of quality or associated outcomes.

The NCCSS (Whitebook & et al., 1989) selected center-based programs from five cities across the country—Atlanta, Boston, Detroit, Phoenix, and Seattle. The goals of this study were to observe the characteristics of child care staff and their relation to the quality of care they provided. Fewer resources were allocated to investigating child development, so outcomes were collected only from the children in Atlanta. The NCCSS team claimed to be the first to conduct an independent psychometric analysis of ECERS and two factors emerged: developmentally appropriate activity and appropriate caregiving. Other teams corroborated the presence of two factors (M. R. Burchinal, et al., 2002; M. R. Burchinal & Roberts, 2000; E. S. Peisner-Feinberg et al., 2001; Robert Pianta, et al., 2005).

Regarding child outcomes, NCCSS results showed that children in centers with lower quality and higher staff turnover were not as developed in language and social skills. One third of the sample fell below 3 in developmentally appropriate activity; two-thirds fell below a 4. Only 12% achieved 5 or above. The quality of the environment predicted the quality of interactions. Receptive vocabulary, secure attachments, and social competence related to the appropriate caregiving dimension of the ECERS. The developmentally appropriate activity dimension of ECERS was not a strong predictor of the outcomes tested. These outcomes results were not deemed representative of child care across the U.S. because Georgia had the least stringent state regulations compared to Federal Interagency Day Care Requirements (FIDCR) and the lowest quality of care of the five states in the study. None of the children in Georgia attended centers meeting all three FIDCR standards related to staff training, ratios, and group size, which were variables in the study. However, as
Table 7-A shows, Arkansas current minimum licensing standards are not far off what was seen in Georgia at that time. Although only two states included had better training requirements, others had better teacher-infant ratios (AZ 1:5, MI and WA 1:4, and MA 2:7) in 1989 than Arkansas has today.

Table 7-A Standards Used in the NCCSS Compared to Arkansas at Present

FIDCR 1989	Georgia 1989	Arkansas Present
1:3 for ages 0-2 1:4 for ages 3-6 Caregivers who do not have a nationally recognized credential must be regularly participating in specialized training.	1:7 for infants 1:10 for toddlers 1:15 up to 1:18 for preschoolers No preservice requirement or hourly annual requirement for training	1:6 for ages 0-1.5 1:9: for ages 1.5-3 1:12 for ages 2.5-3 years 1:15 for age 4. 10 hours annual training

Few teams have been able to track the effect of child care quality on later development among large samples of children. An exception is the CQO study (Helburn, 1995), which found that child care quality had modest effects on children from a wide range of backgrounds through second grade. In the first wave of research (ES Peisner-Feinberg & Burchinal, 1997), the team selected four states in regions that varied economically and in stringency of regulatory oversight of child care. Their initial study sample included 521 preschool rooms and 228 infant-toddler rooms. Outcomes were collected from a smaller sample. Because there were high correlations between ECERS, The Arnett Caregiver Interaction Scale (CIS; Arnett, 1989) and UCLA Early Childhood Observation Form (ECOF; Stipek, Daniels, Galluzzo, & Milburn, 1992) measures (.74-.91), the research team developed a composite quality index. Controlling for family and child characteristics, the team found statistically significant but modest correlations between quality and a number of outcomes. ECERS ratings most strongly related to children's receptive language scores. "Examination of the regression of coefficients holds that, holding all else constant, a 2-point increase in the observed quality index yields more than a 4-point increase in language scores" (p. 473). Quality was also associated with reading, math, cognitive/attention, and sociability. Children of all backgrounds were affected. In fact, the results indicated that having a more advantaged family did not shield children from the effects of poor quality. Ratings of teacher-child interactions were a stronger predictor of social-emotional outcomes than ECERS quality.

The CQO follow-up study (E. Peisner-Feinberg, et al., 2000) monitored the progress of 733 children from age 4 to age 8 to discern how earlier classroom practices related to children's later language and academic skills. Closeness of teacher-child relationship related to cognitive and social skills. Children from families considered at risk experienced greater positive effects than other children. The observed quality in the children's next-to-last year of day care was a mean of 4.26 on the ECERS; using a shortened version of ECERS in the last year of care, the mean was 4.05. Children who were in higher quality child care had higher language and math scores through kindergarten. After kindergarten, the association faded for language but remained strong for math through second grade, especially for children who had mothers with less education. Quality of day care had no effect on letter-word recognition skills.

There is debate about sustained effects of quality. Longitudinal studies of model preschool interventions and large-scale programs, such as Head Start, tend to demonstrate that cognitive gains are more often short-lived, but that academic and social advantages persist into adulthood, even without sustained attendance in high-quality school-age programs. (Barnett, 1995; Garces, Thomas, & Currie, 2002; Ramey, et al., 2000; Arthur J. Reynolds, et al., 2001; Schweinhart & Weikart, 1997; Yoshikawa, 1995). More research is now being conducted to determine the role that that ECE quality plays in affecting these long-term outcomes. For example, an analysis of NICHD participants through the age of 15 indicates that quality of the ECE experience in all types of care will affect later outcomes (D. Vandell, Burchinal, Steinberg, & Vandergrift, 2010). The quality of a child's care at 4.5 years was a significant predictor of cognitive and academic performance at age 15, and the effect size (d=.09) was the same at both periods.

Most of the associations between ERS quality and outcomes are small to modest, and while investigators sometimes collect a battery of developmental outcomes, not all statistically significant associations are for the same developmental outcomes. For instance, Dunn and colleagues (1994) found that ECERS predicted children's language abilities but found no associations between quality and measures of intelligence. In a different study quality predicted sociability but did not predict either language or cognitive development after controls for ages and backgrounds were introduced (S. J. Kontos, 1991). In this case, family background emerged as a better predictor of these outcomes.

On the other hand, Chin-Quee and Scarr (1994) tracked children of the Bermuda Study through fourth grade and failed to find a significant relationship between child care quality and long-term social competence or academic achievement. The Three-State study (Deater-Deckard, et al., 1996; K McCartney et al., 1997) developed a composite measure of quality combining ECERS/ITERS, the Assessment Profile, caregiver training, education and wages. Controlling for a variety of family and child characteristics, quality at the first collection point was not a predictor of children's sociability or behavior four years later. Results of Three-State study were perhaps weakened by moderate inter-rater agreement, .58 for ECERS and .55 for ITERS. Neither study accounted for changes in quality of care that may have been experienced between the initial observation and follow-up (D. Vandell & Wolfe, 2000).

7.1.1 CONSEQUENCES OF LOW QUALITY

Studies attempting to identify the quality of care across the U.S. have found anywhere from 10-20% of the programs falling below the threshold of custodial care (Ellen Galinsky, et al., 1994; Helburn, 1995; Whitebook & et al., 1989). The quality of care for infants is the least researched but has been found to be of worse quality with estimates of inadequate care ranging 25-40%. (E. Galinsky, et al., 1995; Helburn, 1995; Whitebook & et al., 1989). One study, for instance, found that the majority of infant classrooms observed in Georgia and Virginia scored below 3 on ITERS (S. Scarr, et al., 1994). NICHD presented a more positive picture of infant care, but the authors point out this may be attributed to the ORCE and HOME measurements being used rather than ERS and to the selection of states with more stringent regulations for the sample (NICHD Early Child Care Research Network, 1996). It is generally agreed that ECE research does not address the full spectrum of quality in U.S. child care. Low quality providers are less likely to agree to be study participants and to be observed, so unfortunately, rates of low quality care are probably higher than those detected by research.

Low quality non-maternal care has serious consequences: it disrupts physiological processes and impedes cognitive and social-emotional development and academic readiness. One window into the effects of inadequate attention or stimulation from caregivers is the examination of cortisol levels in children in child care. Cortisol is a hormone that helps increase blood sugar and aids metabolism. High levels of cortisol suppress the immune system, physical growth, and reproductive hormones. In parental care, cortisol levels decrease from morning to afternoon, but analysis of children's salivary cortisol levels suggest this pattern can be disrupted in child care. In center-based full-day care, Watamura's team (2003) found reverse patterns wherein cortisol increased from morning to afternoon in 35% of infants and 71% toddlers in center-based care. Dettling and colleagues (2000) also revealed this reversal among children in family-based child care. In both cases, social stimulation appeared to be significantly related to cortisol production. In the Watamura study, children who played with peers retained the same patterns that they experienced at home. In the Dettling study, children in care that measured above the median split on a quality index measure of caregiver's focused attention and stimulation had no change in cortisol levels between home and child care, while those in care below median quality experienced the reversed pattern. A more recent longitudinal study indicates that the reversed patterns may not be experienced as much by older children and that they may diminish as the child gains more experience in child care (Ouellet-Morin et al., 2010).

Animals have been found to be resilient to short periods of cortisol elevation, but prolonged periods lower their tolerance to stressors and establish long-term patterns where stress and anxiety are triggered

more easily (Makino, Gold, & Schulkin, 1994). Animal studies also reveal how neurochemical systems react to physical environments. For example, rats raised in complex environments with play objects receive great benefits to brain development and function compared to rats in an empty cage. (Black, Jones, Nelson, & Greenough, 1998; Shonkoff & Phillips, 2000).

Large ECE studies demonstrate that low quality care also impedes other aspects of child development. For example, Burchinal and colleagues (2000) combined data from three large projects—CQO, the North Carolina Head Start Partnership Study, and the Public Preschool Evaluation. Controlling for poverty, ethnicity and gender, the analysis showed that children in poor quality care, as measured by ECERS, tested almost a full standard deviation lower on a language measure and nearly a third of a standard deviation lower in math and reading tests. Children's social, language and academic development were all related to quality with only small differences found for at-risk children. In other words, quality affected all children. These studies did not incorporate random assignment, but when combined offer sufficient power from which to draw conclusions.

Smart Start North Carolina tracked the progress of children from 110 preschool programs in 9 areas of development from 1994 to 1999. After controlling for the effects of gender, ethnicity, and poverty, Bryant and colleagues (2003) found that receptive language, print awareness, book knowledge, applied math, and counting one-to-one were all significantly positively related to quality. Higher percentages of children in medium-quality classrooms and even more in low-quality classes scored low on receptive language and in applied math problems.

Percentage of NC Smart Children with Poor Scores (n=512)			
ECERS	*High Quality*	*Medium Quality*	*Low Quality*
PPVT Language	24%	30%	55%
Applied Math	21%	28%	45%
Source: (Donna M. Bryant, et al., 2003)			

One of the striking finds from the Missouri QRS evaluation (Thornburg, et al., 2009) is that in centers rated in low to mid tiers, all children lost social skills and children in poverty lost vocabulary between fall and spring. Children in poverty experienced fall-to-spring social and emotional gains only in high tiers of quality. This is cause for concern in Arkansas because BB standards are less stringent than Missouri's on a number of indicators.

The final risk we found to be associated with low quality care in ECE research is an increase in children's anger and defiance, especially for children who enter child care at a younger age or who have families with risk factors (Hausfather, Toharia, LaRoche, & Engelsmann, 1997). High quality care may buffer children from potential negative effects of earlier entry into child care.

7.1.2 HEAD START RESEARCH RELATED TO QUALITY

Because Head Start has its own complex set of regulations and standards, we reviewed Head Start studies to see how the program fared in terms of ERS quality. Overall the research presents a picture of programs providing care rated in the good range of quality without a great degree of variability, whether in small studies, such as that conducted by Hubbs-Tait's team (2002), where 15 out 16 classrooms observed had scores of good or above, or in much larger studies, such as the longitudinal FACES study (Nicholas Zill, et al., 2003). The relatively limited range, though encouraging, may constrain attempts to draw conclusions about the effect of quality in Head Start related to participant outcomes.

One group reported that ECERS related to cognitive and academic readiness scores but not to social and behavioral skills (D.M. Bryant, Burchinal, Lau, & Sparling, 1994). Children who had more stimulating home environments gained more in problem solving and reasoning when their school environments were of higher quality. The quality of instructional practices, such as providing a variety of activities, related to transference of positive behaviors to home and moderated the influence of maternal depression on children's problem behaviors. This study was not representative of Head Start in general. It had a small sample size (n=145) and, like the others, had a limited range of quality. Scores ranged between 3.5 and 5.4, but only 9% of the sample scored above 5.

The Head Start Impact Study (Puma, Bell, Cook, & Heid, 2010) offers a more representative sample and claims to address all levels of quality provided within Head Start. Almost 5,000 children in 23 states were randomly assigned to either a Head Start program or to other early childhood programs. Seventy percent of the children were in care rated 5 or above on ECERS-R/FDCRS. Of the 3-year-old cohort (n=1,083), only 7% of children who attended an early year of Head Start were in care rated 1-3, and 5.26 % of children who attended 2 years were in care rated 1-3. Other than describing the percentages of children in various levels of quality, this study did not provide information about the impact of quality on child outcomes. Instead, the study focused on the overall impact of participation. The effect of one year of participation in Head Start on development into kindergarten and 1st grade was statistically significant but small. Children who entered at age 4 did better on receptive vocabulary by the end of 1st grade, and those who entered at age 3 tested better in oral comprehension. Evidence showed that the 3-year-old cohort had closer, more positive parent relationships. We can speculate that global quality as measured by the ERS is responsible for positive outcomes, but it is possible that outcomes are buoyed by other factors or processes that are prevalent in Head Start but not included in these measures.

There is less research available on Early Head Start center-based care quality. Regression analyses from the randomized Early Head Start Research and Evaluation project (John M. Love, et al., 2004) detected better developmental outcomes among children given higher quality care. Mean ECERS/ITERS quality significantly predicted two of six developmental outcomes, 24-month mental development and 36-month vocabulary. Early Head Start centers received higher ratings (5.0-5.2) compared to community centers enrolling comparison group children (3.8 ITERS at 14 months to 4.9 on ECERS-R at 36 months).

7.1.2.1 CROSSWALKING ERS WITH CLASS

Arkansas stakeholders have raised the question of whether Head Start programs should be excused from elements of Better Beginnings. Head Start is not using ERS but has adopted the CLASS as its measurement of the classroom environment. There is modest overlap between ECERS-R and the primary environmental

assessment tool used for Head Start, Classroom Assessment Scoring System (CLASS, RC Pianta, et al., 2008). CLASS was designed to measure interactions between teachers and children, the primary mechanisms through which children learn and develop in the classroom, and does not measure the physical environment or provision of materials.

The CLASS is organized into three broad domains: Emotional Support, Classroom Organization, and Instructional Support.

Emotional Support consists of four subscales:

- Positive Climate (emotional connection and respect in the classroom)
- Negative Climate (anger, hostility, aggression)
- Teacher Sensitivity (awareness and responsiveness to students)
- Regard for Student Perspectives (student's interests and motivations are emphasized)

Classroom Organization includes three subscales:

- Behavior Management (monitoring and prevention and redirection of disruptive behavior)
- Productivity (routines and organization of activities)
- Instructional Learning Format (facilitate activities and provision of interesting materials)

Instruction Support includes three subscales:

- Concept Development (promotion of higher-order thinking)
- Quality of Feedback (extension of learning through responsiveness to students' outputs)
- Language Modeling (facilitation and encouragement of language development)

As reported in the CLASS manual, there are moderate correlations between the CLASS and the ECERS-R. The three domains of the CLASS—Emotional Support, Classroom Organization, and Instructional Support—were more strongly correlated with elements of teacher-child interaction than provisions as measured with the ECERS-R. Correlations between domains on the CLASS and ECERS-R ranged from .45 to .63 for teacher-child interactions and .33 to .36 for provisions. A study of quality in publicly supported pre-Kindergarten programs (Mashburn, 2008) reported correlations on the total ECERS-R scale and CLASS Emotional Support of .54 and CLASS Instructional Support of .43. The pre-kindergarten quality study reported that CLASS Instructional Support was a significant predictor of multiple measures of children's language and math skills. The ECERS-R total score only predicted oral and written language once elements of the CLASS were considered. Further, social skills (social competence and problem behaviors) were related to the CLASS Emotional Support domain and not ECERS-R scores.

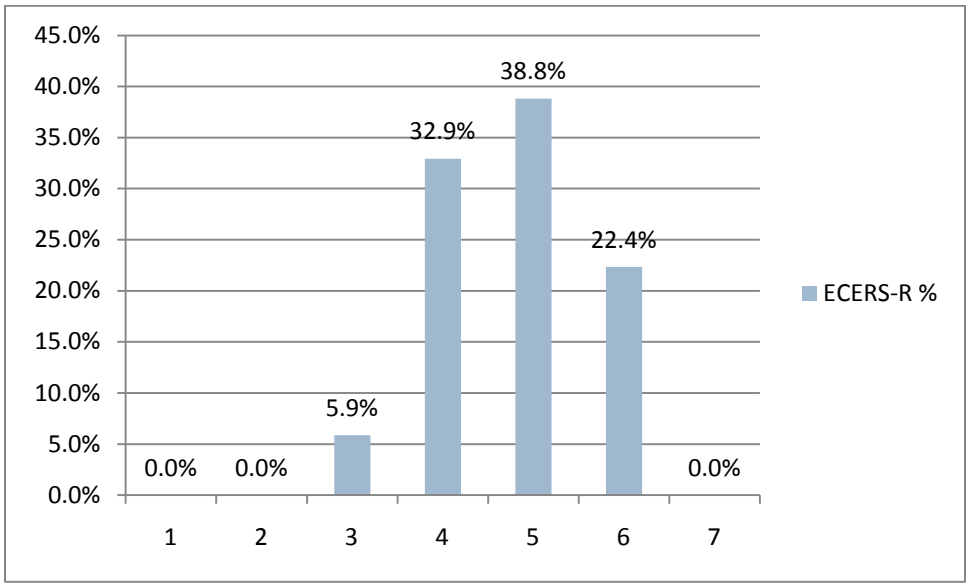
It is clear that elements of instructional and emotional support of students are related to child development. There is minimal overlap with the ECERS-R as the CLASS was not designed to measure the physical environment of the classroom. As expected, our content analysis demonstrates the most overlap

between the ECERS-R Interactions subscale and the CLASS scale, while other areas of the ECERS-R were not fully covered.[41]

As reported in the Quality Thresholds discussion (see 7.2), a recent study by Burchinal and colleagues examined threshold scores on the CLASS (M. Burchinal, Vandergrift, Pianta, & Mashburn, 2010). In this study the team did not find a level of quality at which outcomes quit improving, but they did define a ***lower*** threshold that had to be reached to find an association between outcomes and quality. The "findings indicated that children acquire academic skills only when the minimal standards represented by our cut-off point of above 3.25 on the CLASS Instructional Quality Dimension are met, and that higher quality instruction produces more academic gains" (M. Burchinal, et al., 2010, p. 174).

The Head Start Family and Child Experiences Survey (FACES) 2003 study reported relatively high quality in their national sample of Head Start programs (N. Zill, Sorongon, Kim, Clark, & Woolverton, 2006). Using the ECERS-R, Head Start programs scored 4.8 on average. Given current BB cut scores, only 7.8% of programs assessed during the FACES 2003 study would not be given a Level 3 rating. Using Quality Approved Head Start Centers in the state of Arkansas, we also find programs of at least minimal quality (Table 7-B). The data available to better understand quality in Arkansas was drawn from 85 Head Start classrooms participating in the Quality Approval process. Programs volunteering for Arkansas Quality Approval may not be representative of all Head Start programs in the state, but available data would suggest that the majority of Head Start programs are of sufficient quality to be rated Level 3 per BB standards. Should Arkansas adopt BB Levels beyond Level 3, there is sufficient variability in scores that programs could continue to strive to improve global quality.

Table 7-B: Quality of Head Start Classrooms: Arkansas Quality Approval



Theoretically and empirically, the structural components of the classroom would provide teachers with the materials needed to optimally interact with students. Therefore, the CLASS and ECERS-R complement

[41] See the ECERS-R & CLASS Crosswalk in the Technical Appendices available on the BB website, http://www.arbetterbeginnings.com/.

one another – with the elements assessed with the CLASS building on the elements assessed with the ECERS-R. There is evidence to suggest that Head Start centers would likely have ECERS-R scores in the highest Level of BB (Level 3). Although there is little overlap between the constructs, CLASS scores of at least 3.25 could be used in lieu of ECERS-R scores as a rating of Level 3 quality for Head Start programs in compliance with performance standards when assessments are completed by independent reviewers.[42] If additional levels of quality are adopted as the team hopes, this substitution would need to be reexamined. To confirm, an examination of the scoring of Head Start programs with both assessments is warranted and comparisons of ECERS-R and CLASS scores for Head Start programs in Arkansas should be completed.

7.1.3 SECONDARY ANALYSIS AND META-STUDIES

To observe the possible effects of quality on larger, potentially more representative groups, several research teams have employed secondary analyses and meta-studies. Burchinal and colleagues (2000) conducted secondary analysis combining data from three large studies and found modest but statistically significant evidence that quality is related to social, language and academic skills. The authors cited this as the first study to have enough children of color from low income families to achieve a good power in examining the potential of child care to moderate developmental risk factors. Their analysis suggested that quality is important for all children regardless of social or economic status and that good quality is a protective factor for the language development of low income children of color.

Another combined study examined the development of preschool children in 2,439 preschools in 11 states (Mashburn, et al., 2008). Data from the NCEDL Multi-State Study and the NCEDL-NIEER State-Wide Early Education Program's Study (SWEEP) were used. In addition to controlling for child, family, and state characteristics, the team controlled for children's prior developmental levels, uncommon in previous research but a helpful procedure to eliminate possible bias (Duncan & Gibson-Davis, 2006). In this analysis, ECERS-R only predicted children's expressive language. CLASS, which measures emotional and instructional support in greater detail than ECERS-R correlated with all five measures of child development used in the study. The authors note, "in the past decade, the ECERS–R has been a guide for program development, which has resulted in investments in the features of the environment in early childhood settings that are assessed using the ECERS–R…It may very well be that this process of program development, with the ECERS–R as a target, has indeed produced overall higher levels of program quality that reduces variation in quality between classrooms, thereby attenuating prediction of the ECERS–R to children's development" (p.743).

Secondary data analysis using NICHD, CQO, NCEDL, and two FACES studies, demonstrated that outcomes of low-income children are more closely associated with quality when quality is high (M. Burchinal, et al., in press). Language scores in one study (FACES 1997) and math skills in three studies (NCEDL, NICHD, and FACES 1997) related to quality in the good to high range. Reading outcomes had mixed relationships; one study showed stronger relationships at the good to high range, but another showed relationships at the

[42] Independent review of Head Start programs occur triennially, at which point CLASS assessment will be conducted.

low to average range. Two of the studies showed more negative behavior when quality was good to high. The correlations were modest, but the ECERS scales that had the strongest correlation to outcomes were Interactions first and Program Structure second. This study points us to two key questions for rating system stakeholders: 1) how low can we set the bar to start seeing developmental gains rather than setbacks, and 2) is there high point of quality after which outcomes cease improving?

7.2 QUALITY THRESHOLDS

A recent, largely unanswered question for policy makers is whether there may be asymptotic relationships between quality and outcomes, in particular whether there is an upper-level threshold in quality after which outcomes quit improving so money is not wasted on investments on further improvements of quality (Blau, 2000). This question has become more pressing as more states implement state-run pre-K and quality rating systems. Emerging evidence suggests that care par with minimum licensing standards or even environments with "adequate" ERS ratings are insufficient to produce positive effects on child outcomes. The following studies suggest that the higher the quality of care, the greater the magnitude of effects on child development.

NICHD 2003 and 2006 studies were only able to detect a linear relationship where increases in quality are met with increases in outcomes. Further analysis of NICHD data showed that low income children performed similarly to children with greater income when in higher quality care (Dearing, McCartney, & Taylor, 2009). The authors were careful to distinguish that *higher* quality simply meant the top half of the distribution, which does not mean *high* quality care. Following up on NICHD participants at age 15, Vandell and colleagues (2010) identified a quadratic relationship between quality and developmental outcomes that was not identified in earlier years of the project and thus deemed a "sleeper effect." Controlling for maternal, child, family, and quality characteristics at three points in time, when quality of care at age 4.5 was moderate to high, a greater magnitude of long-term effects was found.

Three recent analyses of the NCEDL Multi-State Study, the NCEDL-NIEER SWEEP Study have also examined this issue (M. Burchinal, et al., 2010; Carollee Howes, et al., 2008; Mashburn, et al., 2008). Two that included the early childhood rating scales examined only linear associations (Carollee Howes, et al., 2008; Mashburn, et al., 2008). One that sought asymptotic relationships only included CLASS (M. Burchinal, et al., 2010). In this study the team did not find a level of quality at which outcomes quit improving, but ***they did define a lower threshold*** that had to be reached to find an association between outcomes and quality. Social outcomes "were more strongly influenced by the quality of teacher-child interactions, but only when teachers are actively and positively engaged with children as indicated when caregivers are rated in the 5-7 range on the CLASS Emotional Support Scale….Similarly, these findings indicated that children acquire academic skills only when the minimal standards represented by our cut-off point of above 3.25 on the CLASS Instructional Quality Dimension are met, and that higher quality instruction produces more academic gains" (p. 174).

Results of the Missouri rating system evaluation (Thornburg, et al., 2009) reflect similar findings. All children in the state's lower tiers lost social skills, and children in poverty lost vocabulary, but those enrolled in upper tiers had gains in both domains. While gains were better in mid tiers than in lower tiers, gains in some developmental domains were only witnessed in the upper two levels of quality.

7.2.1 BB LEVELS: INFANT-TODDLER ENVIRONMENT RATING SCALES

The evaluation team used data from the national evaluation of Early Head Start (EHS), the EHS Research and Evaluation Project (EHSREP), to examine the relationship between ITERS-R and ECERS-R and child outcomes across many ages. The EHSREP included 17 programs across the country and was a rigorous, large-scale, random assignment evaluation. EHS is a two-generation program designed to serve low-income families with children birth to age 3. The 17 programs selected for the EHSRE included all major approaches to early education (distributed across urban and rural settings) reflective of the characteristics of all programs funded in 1995. At study enrollment, respondents were randomly assigned to the program (51%) or comparison (49%) groups. Participants in the comparison group were free to access other services in their communities, which might include other options for enrolling their children in center-based child care. These included programs that were predominantly home-based, programs that were predominantly center-based, and programs that included aspects of both. Details of the program selection and sampling plan are contained in the Final Technical Report (J.M. Love, et al., 2002). In brief, the sample of 3,001 families was approximately 34% African American, 24% Hispanic, 37% European American, and 5% other. Almost half the primary caregivers lacked a high school degree; more than half were unemployed; 36% received TANF/AFDC; 11% did not speak English.

Table 7-C: Infant-Toddler Environment Ratings at 14 Months of Age: Better Beginnings Levels and Child Outcomes at Ages 1 and 2

Construct		14 Month			24 Month			Difference 14 to 24 Months		
		N	Mean	SD	N	Mean	SD	N	Mean	SD
BAYLEY MDI Score; 14M^{ns}, 24M^{ns}, Change^{ns}	Level 1	42	100.14	10.15	36	89.61	13.66	27	-11.48	14.35
	Level 2	56	99.04	9.19	50	91.16	14.08	47	-7.49	10.96
	Level 3	231	99.91	9.66	194	92.60	11.18	179	-7.63	11.80
BBRS: Orientation and Engagement; 14M^{ns}, 24M*, Change*	Level 1	49	3.53	.68	39	3.33	.70	35	-.28	.84
	Level 2	60	3.70	.66	50	3.95	.78	49	.23	.75
	Level 3	239	3.55	.68	214	3.73	.71	206	.23	.92
BBRS: Emotional Regulation; 14M^{ns}, 24M*, Change*	Level 1	48	3.61	.71	38	3.34	.97	33	-.38	1.15
	Level 2	59	3.80	.67	49	3.82	.78	47	.04	.76
	Level 3	239	3.70	.65	214	3.78	.72	206	.13	.85

Note: *p<.05, **p<.01, ns=non-significant

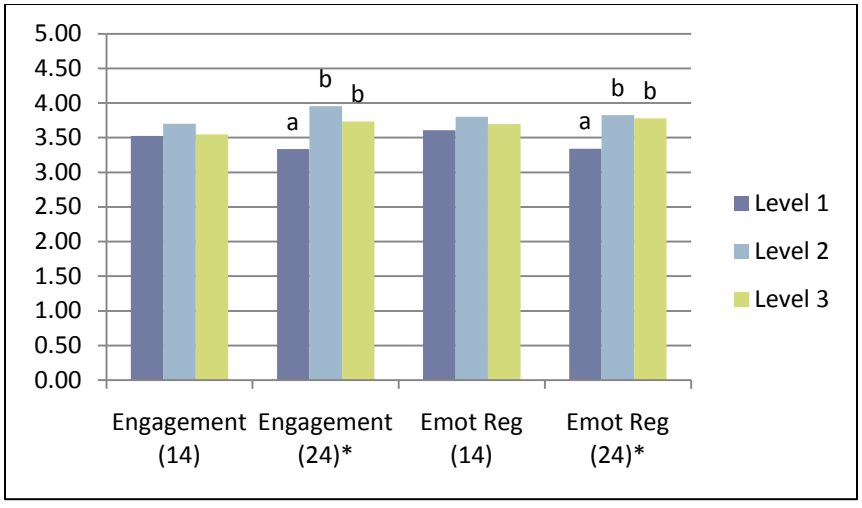
When children were 14 months of age, there were 368 children who had center assessments. The majority of the scores were on the high end of the ITERS, making comparisons on the low end of the scale (and lower Levels of BB) more difficult. Level 1 was defined as scores less than 3, but in actuality, BB does not assess ERS at Level 1. These analyses are based in an assumption that programs in Level 1 would score lower than 3 on the ERS, as is the level required to attain Level 2. The average ITERS scores for children in centers with BB Level 1 were 2.41 (SD=.37, N=55), 3.62 for Level 2 (SD=.24, N=61), and 5.17 (SD=.79, N=252) for Level 3. Data collected for the EHSREP included many quality programs (EHS program children), but also included children from a comparison condition in non-EHS programs in their communities.

At ages 1, 2, and 3, children were administered the Bayley Scales of Infant Development (BSID-II; Bayley, 1993). The BSID-II is a structured assessment administered to toddlers to measure their mental, motor, and behavioral development. Mental Development Index (MDI) scores reflect performance on the cognitive and language portion of the assessment and parallel other IQ scores with a mean of 100 and a standard deviation of 15. Internal consistency for the MDI is high (Cronbach's alpha=0.88; Bayley, 1993). Additional data was available for standardized Bayley Behavior Rating Scales (BBRS), done in conjunction with the BSID-II, which assess children's emotional regulation and children's orientation and engagement, describing children's cooperation, positive affect, and interest in testing materials (Bayley, 1993). Scores are reported on a 5-point Likert scale (α=.92 for the emotional regulation subscale and α=.83 for the orientation and engagement subscale).

Using the child assessment of the MDI and BBRS, we examined the relationship between child care quality with the ITERS collected at 14 months and 14- and 24-month cognitive development (MDI) and engagement in the task and emotion regulation during the task as rated by the assessor. Additionally, we created a change variable meant to document gains and losses on the variables described above. These were created by subtracting age 14-month scores from age 24-month scores, such that higher scores reflect greater gains in skill.

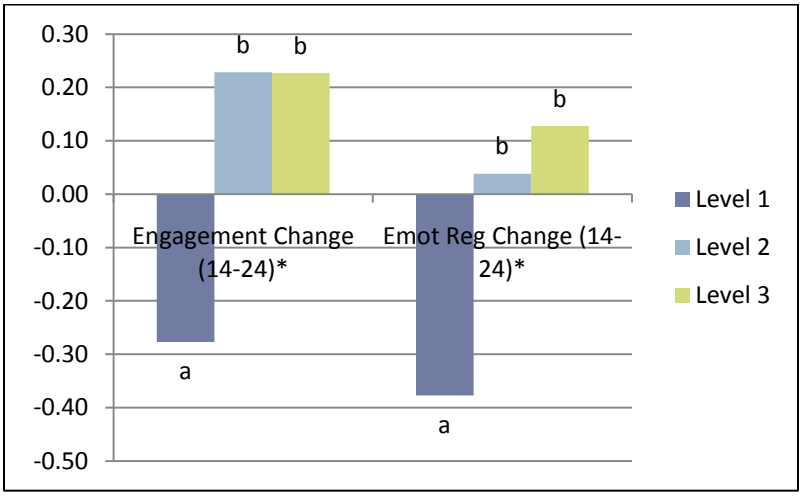
Overall, there were no relationships of the BB Levels with cognitive development as indicated with the MDI, nor the associated change score. There were, however, relationships evident for age 24-month engagement and emotion regulation during the administration of the Bayley, as well as the change scores for these two variables computed from 14 to 24 months of age (Table 7-E).

Table 7-D: Bayley Behavior Rating Scale Scores at 14 and 24 Months of Age: Predicted by Better Beginnings Minimums on ITERS (14 months)



Note: *p<.05, a) 24 month Engagement and Emotion Regulation Scores of Level 1 significantly differ from Levels 2 and 3; Level 1 defined as ITERS < 3.00.

Table 7-E: Changes in Bayley Behavior Rating Scale Scores at 14 and 24 Months of Age: Predicted by Better Beginnings Minimums on ITERS (14 months)



Note: *p<.05; a/b) Level 1 significantly differ from Levels 2 and 3; Level 1 defined as ITERS < 3.00.

The analyses show that infants and toddlers in center-based programs that did not provide at least minimum quality (ITERS scores lower than 3 collected when the child was 14 months) have less optimal engagement and emotion regulation scores at the age of 2 (Table 7-D) and significantly lose skills in both of these areas between the ages of 14 and 24 months (Table 7-E). This finding is concerning as BB Level 2 requires ERS scores of at least 3, and Level 1 requires only a self-assessment. **It is possible that centers with ERS scores lower than Level 2 will apply for Level 1 and earn a quality rating when the potential for harm to children exists.**

We conducted similar analyses to the ones described above for children in child care programs at 24 months of age. Using the EHSREP data, there were 406 children in centers of varying quality (Level 1 M=2.44, SD=.33, N=46; Level 2 M=3.6, SD=.27, N=65; Level 3 M=.53, SD=.76, N=295). Again, a majority of children were in higher quality programs. As with the 14-month ITERS, using the 24-month ITERS to determine BB Levels demonstrates no relationship with cognitive outcomes for children at 24 and 36 months of age. Also, as seen with younger children, there are relationships between BB Levels and child socio-emotional development (BBRS engagement and emotional regulation). At age 3, the Peabody Picture Vocabulary Test–III (PPVT-III; L. M. Dunn & Dunn, 1997) was administered to the children. The PPVT-III is a nationally-normed test that measures children's word knowledge. Children are given a word and then asked to point to one of four pictures that best represents the meaning of the word. Internal consistency estimates for the PPVT-III is > .90. The analyses (in Table 7-F) show that toddlers in center-based programs that did not provide at least minimum quality (ITERS scores collected when the child was 24 months lower than 3; BB Level 1) have significantly lower receptive vocabulary scores at the age of 3 than children in centers with higher ITERS scores in BB Levels 2 and 3 (Table 7-G).

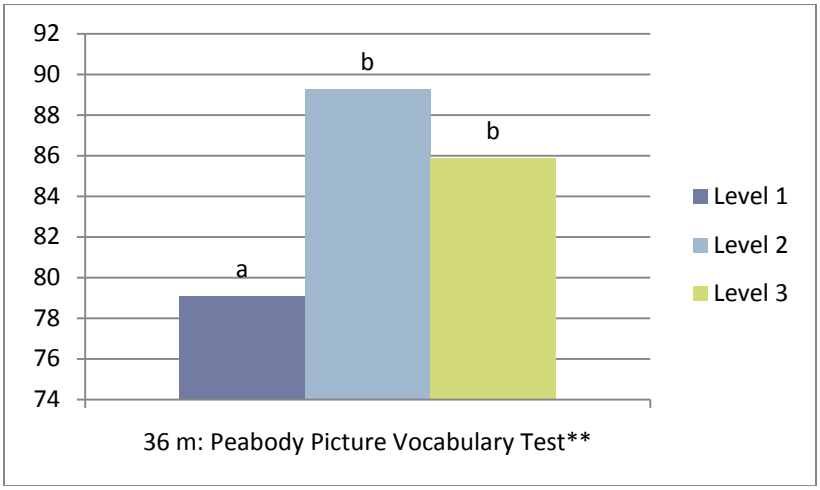
Table 7-F: Infant-Toddler Environment Ratings at 24 Months of Age: BB Levels and Child Outcomes at Ages 2 and 3

Construct		24 Month			36 Month			Difference from 24 to 36 Months		
		N	Mean	SD	N	Mean	SD	N	Mean	SD
BAYLEY MDI Score; 24M[ns], 36M[ns], Change[ns]	Level 1	40	87.70	40	33	90.15	33	29	.10	29
	Level 2	56	91.61	56	48	92.83	48	43	1.80	43
	Level 3	257	92.49	257	233	92.82	233	207	.29	207
Peabody Picture Vocabulary Test**	Level 1	Not Applicable			31	79.10	31	Not Applicable		
	Level 2				52	89.29	52			
	Level 3				226	85.86	226			
BBRS: Orientation and Engagement Scale; 24M*, 36M[ns], Change[ns]	Level 1	46	3.41	46	35	3.75	35	35	.30	35
	Level 2	61	3.75	61	53	4.04	53	51	.26	51
	Level 3	279	3.72	279	243	3.90	243	232	.17	232
BBRS: Emotional Regulation; 24M*, 36M*, Change[ns]	Level 1	46	3.43	46	35	3.76	35	35	.30	35
	Level 2	60	3.59	60	53	3.84	53	50	.20	50
	Level 3	278	3.76	278	240	4.03	240	228	.27	228

Note: *p<.05, **p<.01, ns=Non-significant

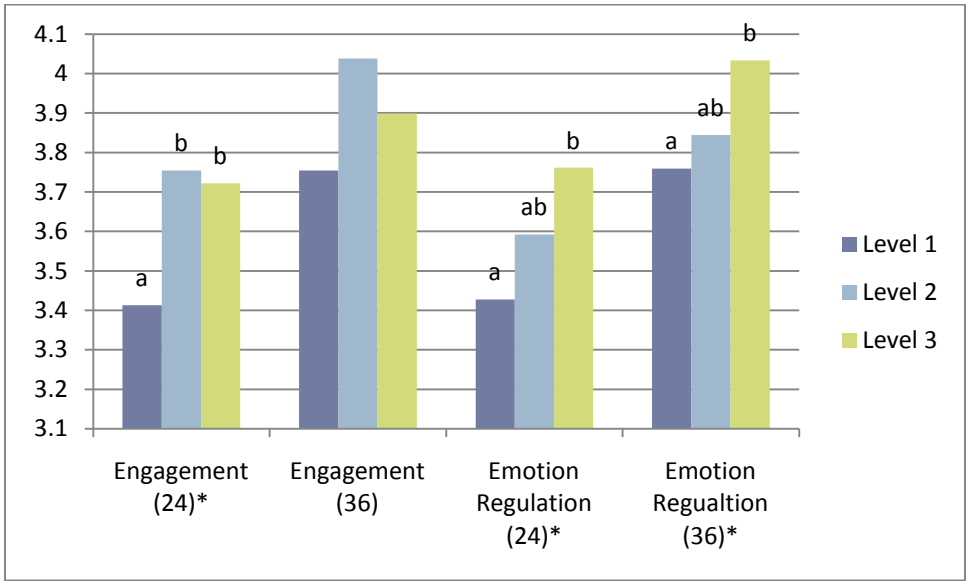
The analyses also show that toddlers in BB Level 1 center-based programs have less optimal engagement scores at the age of 2 (Table 7-F) than children in centers with higher ITERS scores in BB Levels 2 and 3. Further, children in less than minimal quality centers (defined as Level 1) score significantly less in emotional regulation during the Bayley administration at ages 2 and 3 than children in the highest level of quality, BB Level 3 (also shown in Table 7-F). Unlike ITERS scores at 14 months, there were no differences in gains/loss of skills when examining the 24-month ITERS scores.

Table 7-G: Receptive Vocabulary Scores at 36 Months of Age: Predicted by Better Beginnings Minimums on ITERS (24 months)



Note: *p<.05, 36 month PPVT Scores of Level 1 significantly differ from Levels 2 & 3.

Table 7-H: Bayley Behavior Rating Scale Scores at 24 and 36 Months of Age: Predicted by Better Beginnings Minimums on ITERS (24 months)



Note: *p<.05, a) 24 month Engagement Scores of Level 1 significantly differ from Levels 2 and 3; 24 and 36 month Emotion Regulation Scores of Level 1 significantly differ from Levels 3; Level 1 defined as ITERS<3.

As a whole, findings from analyses with the ITERS suggest that infancy and toddlerhood is a time when program quality is especially important. Infants attending programs of less than minimal quality lost emotion regulation skills over the course of a year. In both infancy and toddlerhood, programs of less than minimal quality were related to less optimal child outcomes.

7.2.2 BB LEVELS: EARLY CHILDHOOD ENVIRONMENT RATING SCALES

Data available as part of the EHSREP also includes ratings of the child care environment using the Early Childhood ERS (ECERS). ECERS data is available on child care centers when children were 3 and 5 years of age. We used outcome data collected at ages 3 and 5 to examine whether BB cut scores on the ECERS were useful for predicting child outcomes. Again, as was demonstrated at younger ages, the majority of programs met quality ratings at BB Level 3 (Level 1 M=2.26, SD=.48, N=42; Level 2 M=3.55, SD=.31, N=81; Level 3 M=5.35, SD=.76, N=366).

At age 3, the measures of cognitive and socio-emotional developments that were described above were still in use. At age 5, child assessments included different measurements of child outcomes. As a result, we are not able to compute change scores for many of the child outcomes assessed at these ages. At age 5, children were administered the Letter-Word Identification and Applied Problems Scales from the Woodcock-Johnson-III Tests of Achievement (WJ-III; Woodcock, McGrew, & Mather, 2001)and the Sustained Attention Task from the Leiter International Performance Scale-Revised (Leiter-R; Roid & Miller, 1997). WJ-III is a nationally-normed achievement battery. Letter-Word Identification measures the child's word identification skills. Items for young children primarily involve identifying specific letters of the alphabet. Applied Problems requires the child to analyze and solve math problems. To solve the problems, the child must listen to the problem, recognize the procedure to be followed, and then perform relatively

simple mathematical operations. Also at ages 3 and 5, the Peabody Picture Vocabulary Test–III was administered to the children.

The Leiter-R Sustained Attention subtest for 4- to 5-year-olds is a timed cancellation task. Children are presented with a target figure, such as a flower, and are asked to find and cross out as many of the target figures on the page as possible and to work as fast as they can within the allotted time, which varies by target figure from 30 to 60 seconds. The task measures visual prolonged attention and requires good visual scanning and motoric inhibition. The Sustained Attention subtest results in a scaled score by application of age-appropriate norms. The Leiter-R manual reports an attention sustained internal consistency for 4- to 5-year-olds of .83 (p.157), and test-retest reliability for 6- to 18-year-olds of .85 (Roid & Miller, 1997; p.162). Also at ages 3 and 5, parents reported on children's aggressive behaviors using 19 items from the Child Behavior Checklist (ASEBA; Achenbach & Rescorla, 2000). Parents reported on the frequency of each behavior problem within the past two months using a 3-point scale of 0 (not true), 1 (somewhat or sometimes true), or 2 (very true or often true). Internal reliability estimates for the Child Behavior Checklist aggression scales were high (alpha=0.88).

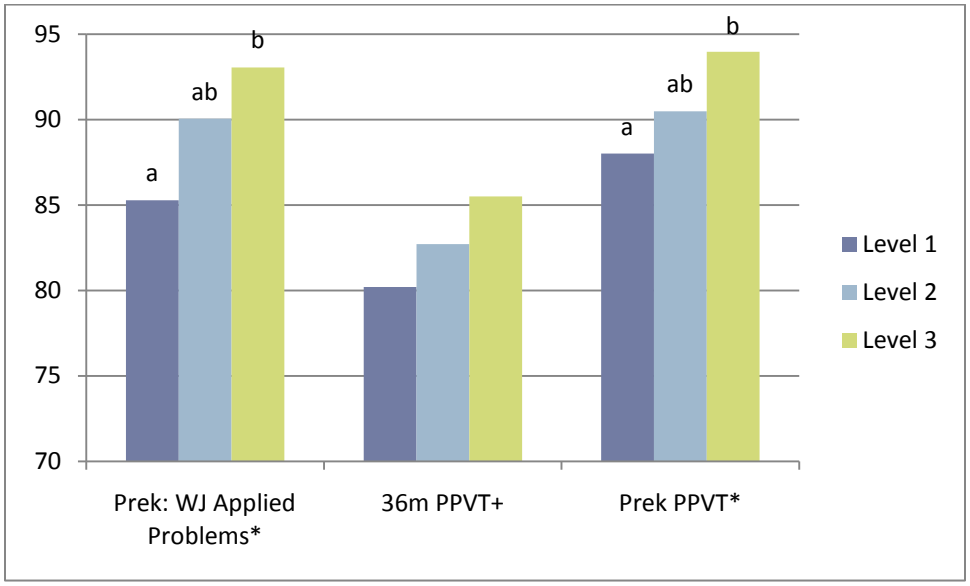
Across the multiple assessments of child cognitive and language development available at the ages 3 and 5 (Table 7-I), there were two significant differences for children across BB Levels. For children's pre-kindergarten (age 5) outcomes, we found that children in Level 1 centers had significantly lower Woodcock-Johnson Applied Problems and PPVT-III scores than children in BB Level 3 programs (Table 7-J). Like the 24-month ITERS scores, the scores of children in the lowest BB Level are worrisome. Both the PPVT and Woodcock-Johnson scales are normed with average scores of 100 and standard deviations of 15. Children in BB Level 1 centers have average scores on math abilities and receptive vocabularies that are (or are nearly) a full standard deviation below the national average.

Table 7-I: Early Childhood Environment Ratings at 36 Months of Age: Better Beginnings Levels and Child Cognitive and Language Outcomes at Ages 3 and 5

Construct		36 Month			Pre-Kindergarten			Difference from 36 Months to Pre-Kindergarten		
		N	Mean	SD	N	Mean	SD	N	Mean	SD
BAYLEY MDI Score; 36M[ns]	Level 1	34	91.21	12.36	Not Applicable			Not Applicable		
	Level 2	68	91.15	11.49						
	Level 3	323	93.13	11.78						
Woodcock Johnson Letter-Word Identification; Pre-K[ns]	Level 1	Not Applicable			36	93.50	13.47	Not Applicable		
	Level 2				73	90.51	11.90			
	Level 3				295	92.26	13.16			
Woodcock Johnson Applied Problems Standard; Pre-K*	Level 1	Not Applicable			36	85.28	20.57	Not Applicable		
	Level 2				73	90.04	17.63			
	Level 3				295	93.04	16.82			
PPVT Std Score, 36M+, Pre-K*, Change[ns]	Level 1	35	80.20	14.74	36	88.00	18.44	30	8.20	11.84
	Level 2	66	82.71	14.38	73	90.47	12.46	61	7.01	14.20
	Level 3	301	85.49	15.21	293	93.96	15.48	259	8.67	15.04

Note: ns=non-significant, +p<.10, *p<.05, **p<.01

Table 7-J: Child Cognitive and Language Development Scores at Ages 3 and 5: Predicted by Better Beginnings Minimums on ECERS (36 months)



Note: *p<.05, Applied Problems and PPVT at 5; Level 1 significantly differs from Levels 3.

For children's social-emotional outcomes, we found no significant differences in child scores across the Levels at ages 3 or 5. Finally, using data collected at the pre-kindergarten follow up of EHS, we examined center quality scores for children at age 5 and found no relationships of BB Levels (1, 2, or 3) with any of the outcomes measured, cognitive or socio-emotional (data not shown).

Head Start FACES provides longitudinal data on a periodic basis on the characteristics, experiences, and outcomes of children and families served by Head Start, as well as the characteristics of the Head Start programs that serve them. Each round of FACES is a study with a nationally representative sample of Head Start children and their families. Data collection for the first FACES cohort began in fall of 1997 and for the second cohort in fall of 2000. Data collection for the most recent FACES cohort began in the fall of 2003. This cohort includes a nationally representative sample of approximately 2,400 newly entering 3- and 4-year-old children and their families from 63 Head Start programs.

The FACES study collected ECERS data from Head Start programs in the fall of 2003 and spring of 2004. We conducted analyses separately for Fall 2003 ECERS scores, using them to predict Fall 2003 and Spring 2004 child outcomes. Again, like the EHSREP data, there are fewer programs at the lower end of quality (Level 1 M=2.67, SD=.31, N=74; Level 2 M=3.6, SD=.27, N=401; Level 3 M=5.02, SD=.68, N=1618). Indeed, there was more variability with the EHSREP data as that study design included a comparison group, but the FACES study included only Head Start programs. Although there are programs with lower quality ratings in FACES, these are still Head Start programs required to meet performance standards. In the FACES 2003 cohort, similar measures were used to assess child cognitive development and school readiness skills as were used in the EHSREP. In addition to the PPVT and Woodcock-Johnson, the FACES study included Story and Print Concepts (modified from Mason & Stewart, 1989), which measures book knowledge, print knowledge and reading comprehension, and the Comprehensive Test of Phonological Processing (PRE-CTOPP; Lonigan, Wagner, Torgeson, & Rashotte, 2002), which assesses phonological awareness, phonological memory, and rapid naming.

Table 7-K: Child Cognitive and Language Development Scores at Ages 3 and 5: Predicted by Better Beginnings Minimums on ECERS (36 months)

		Fall Scores			Spring Scores			Change from Fall to Spring		
Constructs		N	Mean	SD	N	Mean	SD	N	Mean	SD
PPVT Standard Score: Fall**, Spring**, Change+	Level 1	72	82.54	12.24	55	83.52	12.09	53	.20	9.13
	Level 2	389	82.42	10.70	290	83.82	10.44	284	1.65	8.63
	Level 3	1324	85.35	11.95	1005	86.75	12.55	901	2.68	9.50
WJ-III Letter-word: Fall+, Spring*, Change^{ns}	Level 1	73	88.29	16.16	55	93.51	15.28	54	7.02	16.21
	Level 2	387	93.08	18.11	287	99.25	16.36	281	5.60	15.44
	Level 3	1326	92.56	16.49	1001	97.37	15.76	898	5.20	13.80
WJ-III Applied problems: Fall**, Spring**, Change^{ns}	Level 1	73	84.40	17.33	55	86.84	15.38	54	3.05	16.48
	Level 2	386	85.50	17.82	289	88.80	16.72	283	3.11	15.83
	Level 3	1307	89.51	17.45	1001	91.80	16.08	884	3.07	13.80
WJ-III Dictation and Spelling: Fall*, Spring+, Change^{ns}	Level 1	71	90.05	10.24	52	90.75	11.78	51	-.60	11.04
	Level 2	376	94.00	12.02	273	94.77	12.47	260	.77	11.38
	Level 3	1251	93.49	11.19	972	93.97	12.16	821	.45	10.71
Story and print concepts: Fall+, Spring**, Change+	Level 1	74	2.55	1.88	55	4.04	2.63	55	1.39	1.96
	Level 2	399	2.64	1.83	290	4.06	2.21	290	1.46	1.67
	Level 3	1591	2.85	2.00	1112	4.65	2.47	1098	1.70	1.91
PRE-CTOPP: Fall^{ns}, Spring+, Change*	Level 1	73	3.76	1.51	55	4.24	1.50	54	.32	1.31
	Level 2	388	3.67	1.47	289	4.60	1.55	283	.85	1.60
	Level 3	1320	3.78	1.65	1000	4.71	1.77	894	.94	1.77

Note: ns=non-significant, +p<.10, *p<.05, **p<.01

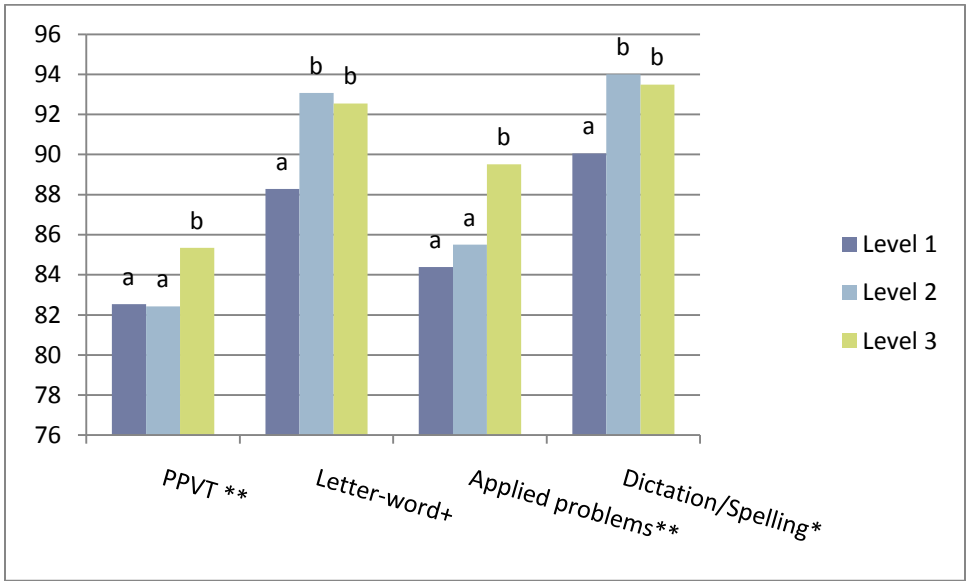
Like the EHSREP data, preschool quality predicts receptive vocabulary and applied math skills in the fall (

Table 7-K). In addition, there was a trend towards prediction of letter-word knowledge and a significant prediction to Woodcock-Johnson Dictation and Spelling scores, which includes writing letters and words from dictation. The findings reported for the fall are still demonstrated with child outcomes in the spring of the same school year (also in

Table 7-K). Across academic skills, children in Level 1 programs score significantly lower than children in higher Levels, although findings differ across constructs in the fall (Table 7-L) and spring (Table 7-M and Table 7-N). Again, scores for some cognitive/academic outcomes for children in Level 1 programs fall lower than one standard deviation below national norms. These are low-income, high-risk children, however, and all scores are lower than national norms for cognitive/academic skills.

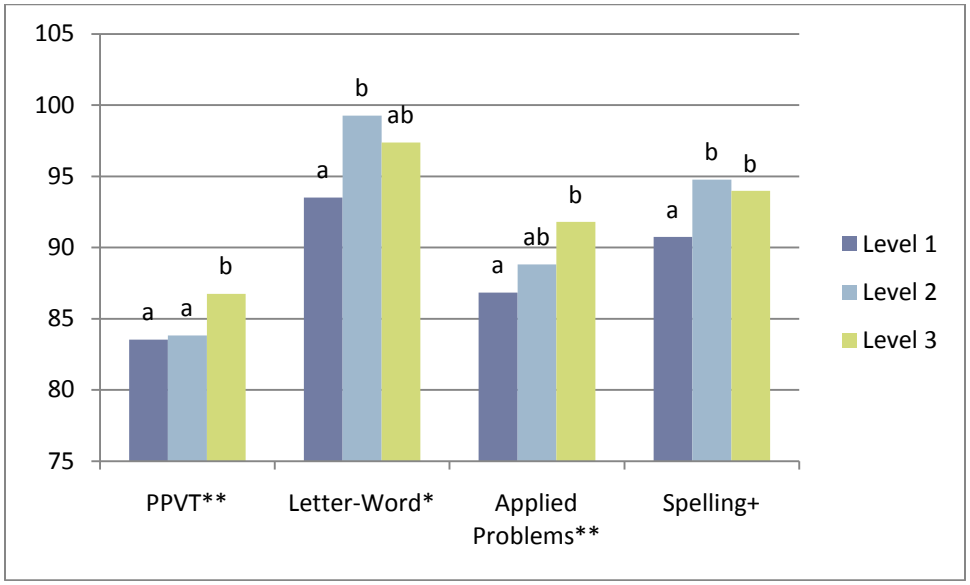
A criterion for eligibility for Head Start is living in a family at 100% of federal poverty. Even at the higher levels of quality, scores are low for many of the academic/cognitive measures included in the Head Start FACES study. However, for most outcomes, children in Level 1, and sometimes in Level 2 BB programs have significantly lower scores than children in the higher quality centers (BB Level 3, or centers with ECERS scores above 4). This is concerning as low-income children are already at a disadvantage, and low quality care has the capacity to further impede development (Table 7-L). Looking at the significant findings, it appears that for math skills, being in the highest quality BB centers (Level 3) is related to higher scores. This is concerning as Level 1 and Level 2 programs do not show divergence in support of math skills. For language skills, however, only children in the lowest quality programs (Level 1) have significantly lower skills.

Table 7-L: Child Cognitive and Language Development Scores in the Fall in Head Start: Predicted by Better Beginnings Minimums on ECERS



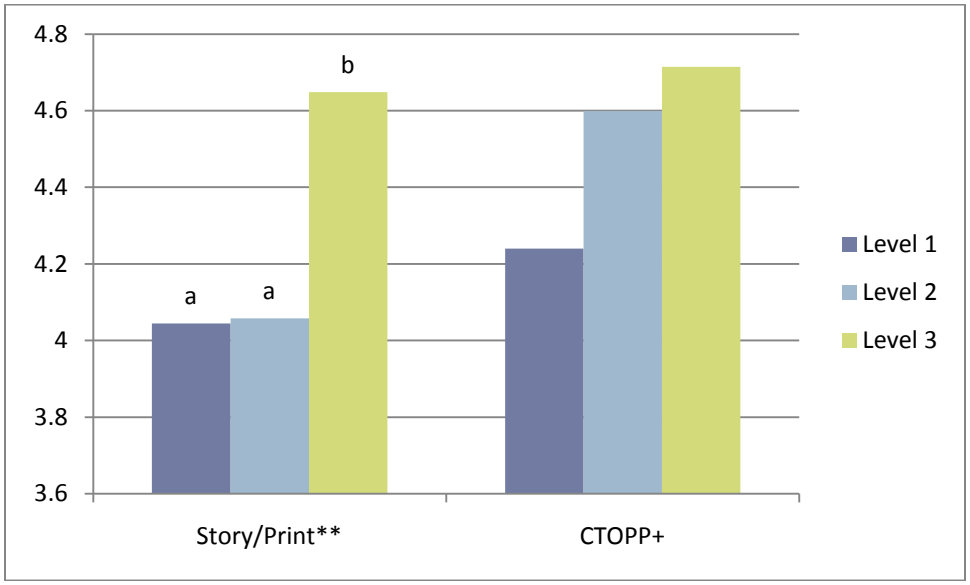
Note: +p<.10, *p<.05, **p<.01; for PPVT and Applied Problems Levels 1 and 2 significantly differ from Level 3; for Letter Word Identification and Dictation/Spelling, Level 1 differs from Levels 2 & 3.

Table 7-M: Child Cognitive and Language Development Scores in the Spring in Head Start: Predicted by Better Beginnings Minimums on ECERS



Note: +p<.10, *p<.05, **p<.01; for PPVT, Levels 1 and 2 differ from 3; for Letter-Word, Level 1 significantly differs from Level 2; for Applied Problems, Level 1 significantly differs from Level 3; for Spelling, Level 1 significantly differs from Levels 2 and 3.

Table 7-N: Child Cognitive and Language Development Scores in the Spring in Head Start: Predicted by Better Beginnings Minimums on ECERS



Note: +p<.10, *p<.05, **p<.01; for Story/Print Concepts, Levels 1 and 2 differ from Level 3.

Using child outcomes data in the fall and the spring, we calculated change scores for cognitive outcomes. These were calculated with fall scores subtracted from spring scores, such that positive scores represent gains, and negative scores represent losses in skills. There were two trend level differences and one significant difference in change scores from fall to spring (Table 7-K). For the PPVT, children in Level 1 programs made almost no change in scores, where children in Level 2 and 3 programs made gains of over 1.5 and 2.5, respectively. The PRE-CTOPP—measuring children's phonological awareness, phonological memory, and rapid naming—showed gains of nearly a full point for children in Level 2 and 3 programs.

The FACES study also included measures of child social development. Measures included teacher reported social skills and classroom conduct problems. Social skills were measured with a 12-item scale assessing frequency with which child engaged in friendly, cooperative, and compliant behavior in class during the past month based on the Social Skills Rating System (Elliot, Gresham, Freeman, & McCloskey, 1988). Classroom Conduct Problems were measured with a 14-item scale assessing frequency with which child engaged in aggressive, hyperactive, or depressed-withdrawn behavior in class during the past month (Achenbach, 1992). Table 7-O shows the means for children in centers with scores paralleling those of the BB Levels.

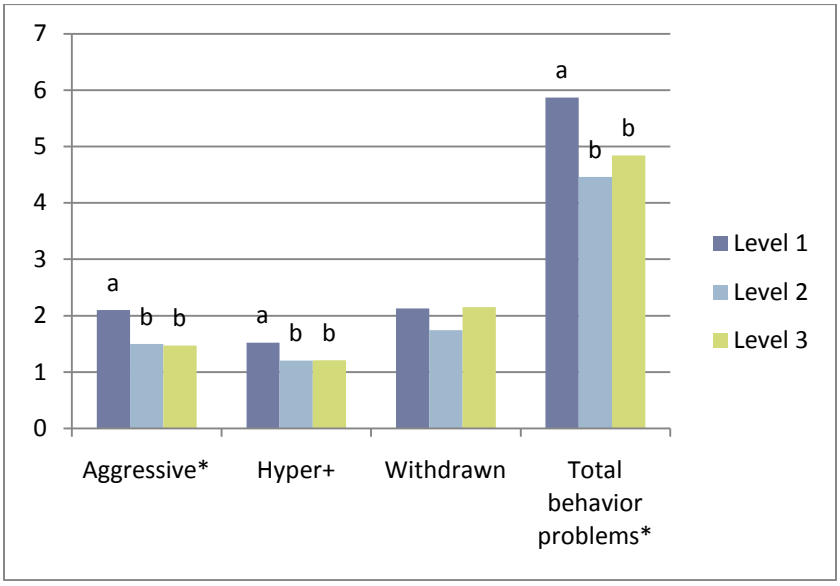
Children in centers with lower ECERS scores are reported by their teachers to have less optimal social development than children in higher ECERS-rated centers in the Fall of the preschool year (Table 7-P). Children in BB Level 1 were reported to be nearly significantly more hyperactive, significantly more aggressive, and to have more problem behaviors overall than children in higher BB Level centers. There were fewer significant differences demonstrated in the spring of the school year for social development (also in Table 7-O) than in the fall. There were not differences in teacher-reported aggression and hyperactivity, as was shown in the fall, but there was a significant difference in overall social skills, with children in the lowest quality programs (Level 1) scoring significantly less than children in Levels 2 and 3 centers (Table 7-O).

Table 7-O: Early Childhood Environment Ratings at in Head Start Programs: Better Beginnings Levels and Child Socio-Emotional Outcomes in the Fall and Spring

Construct		Fall Scores			Spring Scores		
		N	Mean	SD	N	Mean	SD
Social skills score: Fall^{ns}, Spring^{ns}	Level 1	76	15.03	4.31	58	16.21	4.74
	Level 2	395	15.63	4.66	295	17.68	4.83
	Level 3	1646	15.46	4.63	1138	17.54	4.37
Aggressive behavior score: Fall*, Spring^{ns}	Level 1	77	2.10	2.18	57	1.86	2.22
	Level 2	396	1.50	1.95	293	1.35	1.85
	Level 3	1647	1.47	1.90	1142	1.34	1.86
Hyperactive behavior score: Fall+, Spring^{ns}	Level 1	77	1.52	1.51	57	1.04	1.28
	Level 2	398	1.20	1.42	296	1.08	1.50
	Level 3	1647	1.21	1.47	1143	1.02	1.37
Withdrawn behavior score: Fall^{ns}, Spring^{ns}	Level 1	72	2.13	2.16	55	1.93	2.19
	Level 2	398	1.74	2.20	294	1.58	2.07
	Level 3	1639	2.15	2.31	1130	1.79	2.10
Total behavior problems score: Fall*, Spring^{ns}	Level 1	77	5.87	4.77	58	4.93	4.52
	Level 2	399	4.46	4.55	297	4.00	4.54
	Level 3	1660	4.84	4.64	1145	4.16	4.34

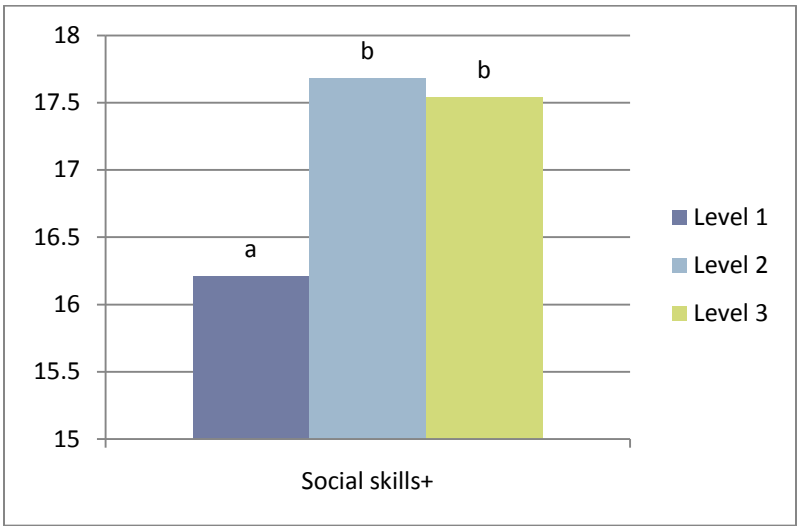
Note: ns=non-significant, +p<.10, *p<.05, **p<.01

Table 7-P: Child Social Development Scores in the Fall in Head Start: Predicted by Better Beginnings Minimums on ECERS in the Fall



Note: +p<.10, *p<.05, **p<.01; for aggressive, hyper, and overall total behaviors Level 1 significantly differs from Levels 2 and 3.

Table 7-Q: Child Social Skill Scores in the Spring in Head Start: Predicted by Better Beginnings Minimums on ECERS in the Fall



Note: +p<.10, *p<.05, **p<.01; Level 1 significantly differs from Levels 2 and 3.

For Better Beginnings, cut scores on the lower end of quality ERS could be problematic. Findings from Infant-Toddler Environment Ratings Scale data (discussed in 7.2.1) showed a loss of emotion regulation and engagement skills for very young children at the lowest ERS levels. Further, as children age, it was evident that programs scoring lower than 3 on the ECERS resulted in children with less optimal language and math skills and socio-emotional development.

7.3 QUALITY IN FAMILY DAY CARE

Much less attention has been dedicated to the observation and analysis of quality in family child care related to child outcomes. Some comparisons of types of care show lower quality scores in family child care than center care (Bigras et al., 2010; B. Fuller, Kagan, Loeb, & Chang, 2004; Susan Kontos, et al., 1994). NICHD (2000) identified slightly higher quality in family day care (11% poor, 30% fair, 43% good, and 16% excellent) than in center-based care, but in general, good quality care is not found to be the norm for children in standard family day care arrangements. In a study of 225 children in the homes of 226 family care providers in California, Texas, and North Carolina, Galinsky and colleagues (1994) found that only 9% of the 225 children were in good quality family care (rated by FDCRS, the ERS version prior to FCCERS). Of the remaining children, 56% were in care rated adequate, and 35% were in care rated harmful. The team observed that the quality of care had ramifications for children's sense of attachment and security. Children who received care in the good to adequate range were more securely attached to their providers. Overall, however, only half the children experienced secure attachments.

Low income families are less likely to receive comparable quality when they choose family care than when they choose center care. Kisker's team observed that programs with a primary goal of school preparation or developmental enhancement charge more than those whose primary goal is to provide care for

working mothers. While subsidies made it possible for low-income children to receive center-based care comparable to that received by higher-income children, this equity was not found in family-based care (Kisker, Hofferth, Phillips, & Farquhar, 1991). The Study of Family Child Care and Relative Care produced similar results: nearly 75% of very low income children were in care of inadequate quality.

Within family day care research, more time has been dedicated to determining the variables that drive global quality, most commonly types of teacher training and education, rather than on which variables specifically improve child outcomes (Barnard, et al., 2006; M. Burchinal, et al., 2002; Clarke-Stewart, et al., 2002; Doherty, Forer, Lero, Goelman, & LaGrange, 2006; Norris, 2001; H. A. Raikes, Raikes, & Wilcox, 2005). A handful of studies have associated FCCERS quality to the following child outcomes:

- **more frequent and/or competent play with adults, peers, and/or objects** (Carollee Howes & Stewart, 1987; S. Kontos, et al., 1995);
- **better infant/caregiver attachments** (Elicker, Fortner-Wood, & Noppe, 1999; Ellen Galinsky, et al., 1994; S. Kontos, et al., 1995);
- **sociability** (Susan Kontos, et al., 1994; M. Rosenthal, 1994);
- and **language development** (Goelman & Pence, 1988; Susan Kontos, 1994);

The greater number of associations made between quality and social-emotional outcomes reflects divergent goals for home-based and center-based care. Whereas parents who choose center care may be more interested in academic progress, parents who choose family day care are likely to be more interested in a warm, home-like environment offering security for their children. Reflecting this partiality, studies of family child care research more frequently incorporate social-emotional measures, and some have not included cognitive measures at all**.**

7.3.1 BB LEVELS: FAMILY CHILD CARE ENVIRONMENTAL RATING SCALES

The evaluation team used data from the national evaluation of Early Head Start (EHS), the EHS Research and Evaluation Project (EHSREP), to examine the relationship between FCCERS and child outcomes across many ages. The EHSREP used FDCRS, the precursor to FCCERS. While retaining basic similarities in format and content that provide continuity between the FDCRS and FCCERS–R, the authors made several changes to bring the scale in line with the other revised editions in the ERS series. Of substance were changes to the wording of items to make them more inclusive for children with special needs and to make them more sensitive to cultural variability. There were also items added and removed. Many FDCRS items changed significantly: Helping children use language; Art; Use of TV, video, and/or computer; Schedule; Adaptations for special needs; and Relationships with parents.

Analyses of the FDCERS may not reflect current scores on the FCCERS-R, but the existing data available on family care is limited. There were far fewer EHSREP children in family care than in center care, which results in a smaller sample from which to draw conclusions. The average FDCERS scores for children in centers at BB Level 1 was 2.39 (SD=.44, N=62), at Level 2 was 3.50 (SD=.28, N=43), and at Level 3 was 4.64 (SD=.60, N=38). Using the child assessment of the MDI and BBRS (described in 7.2.1), we examined the relationship between family child care quality with the FDCERS collected at 14 months and 14- and 24-month cognitive development (MDI), engagement in and emotion regulation during the Bayley tasks, as well as the change variables used in the ITERS/ECERS analysis above. Overall, there was one trend-level

relationship of the BB Levels with cognitive development at 14 months of age, with children in the lowest BB Level having the lowest scores. There were not relationships with scores at 24 months of age nor with the cognitive change score (Table 7-R).

Table 7-R: FDCERS at 14 Months of Age: BB Levels and Child Outcomes at Ages 1 and 2

		14 Month			24 Month			Difference from 14 to 24 Months		
Construct		N	Mean	SD	N	Mean	SD	N	Mean	SD
BAYLEY MDI Score; 14M+, 24M^{ns}, Change^{ns}	Level 1	51	95.75	11.19	47	87.68	12.93	38	-8.08	14.08
	Level 2	38	100.95	9.85	35	89.23	12.40	33	-12.27	15.48
	Level 3	30	98.60	11.03	31	92.10	12.93	23	-6.96	15.01
BBRS: Orientation and Engagement; 14M^{ns}, 24M^{ns}, Change^{ns}	Level 1	60	3.49	.80	60	3.65	.75	52	3.67	.67
	Level 2	38	3.70	.72	38	3.76	.59	34	3.76	.79
	Level 3	36	3.57	.74	35	3.83	.63	32	3.77	.62
BBRS: Emotional Regulation; 14M^{ns}, 24M^{ns}, Change^{ns}	Level 1	52	3.60	.70	51	.16	.80	51	-.11	.85
	Level 2	32	3.66	.95	32	-.05	.79	31	-.10	.67
	Level 3	32	3.91	.69	30	.20	.68	29	-.00	.94

Note: ns=non-significant, +p<.10, *p<.05, **p<.01

Table 7-S: FDCERS at 24 Months of Age: Better Beginnings Levels and Child Outcomes at Ages 2 and 3

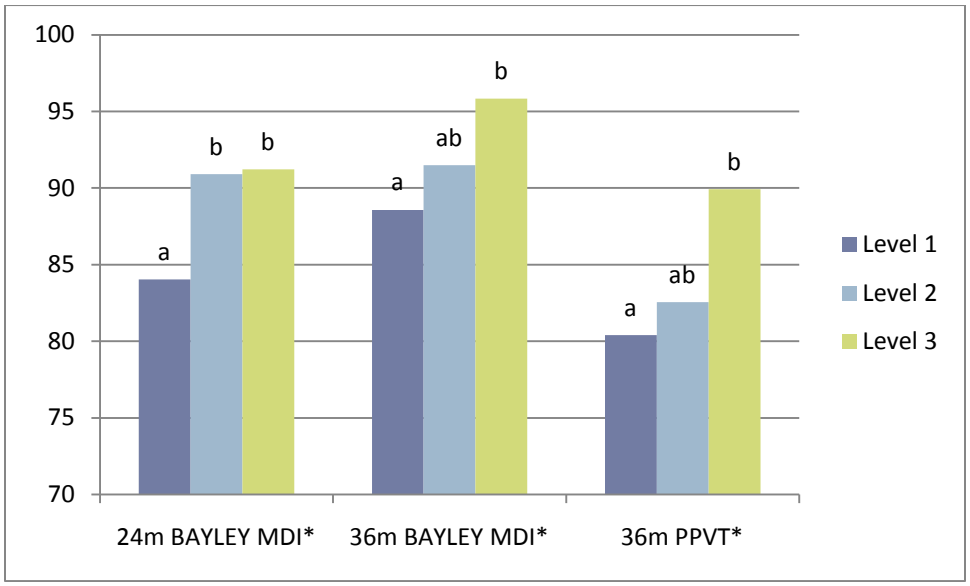
Construct		24 Month			36 Month			Difference from 24 to 36 Months		
		N	Mean	SD	N	Mean	SD	N	Mean	SD
BAYLEY MDI Score; 24M*, 36M*, Change^{ns}	Level 1	43	84.05	13.26	33	88.58	9.94	29	2.83	11.13
	Level 2	33	90.91	11.93	29	91.48	11.38	28	1.54	10.88
	Level 3	48	91.23	14.22	44	95.84	13.40	40	4.98	13.50
BBRS: Orientation and Engagement Scale; 24M*, 36M+, Change^{ns}	Level 1	49	3.41	.82	37	3.57	.54	35	.11	.84
	Level 2	32	3.78	.60	29	3.83	.78	27	.09	.78
	Level 3	53	3.78	.76	45	3.87	.64	44	.07	.84
BBRS: Emotional Regulation, 24M+, 36M^{ns}, Change^{ns}	Level 1	48	3.47	.86	37	3.80	.64	35	.37	1.00
	Level 2	31	3.83	.55	28	3.91	.75	25	.11	.88
	Level 3	52	3.76	.83	43	3.88	.86	41	.12	.91
36m PPVT STANDARD SCORE*	Level 1	Not Applicable			32	80.41	10.49	Not Applicable		
	Level 2				23	82.57	15.07			
	Level 3				28	89.93	16.41			
36m CBCL Aggressive Behavior	Level 1	Not Applicable			39	10.27	7.09	Not Applicable		
	Level 2				27	11.22	7.41			
	Level 3				52	11.31	7.03			

Note: ns=non-significant, +p<.10, *p<.05, **p<.01

We conducted a similar analysis with FDCERS scores collected when children were 2 years of age. Here we find more significant predictions to children's cognitive and social development (Table 7-S). Overall, we find lower cognitive development scores at ages 2 and 3, lower scores in how the child is engaged in

the Bayley task, and lower receptive language scores (using the PPVT) at age 3. Cognitive and language scores are presented in Table 7-T for ease of comparison.

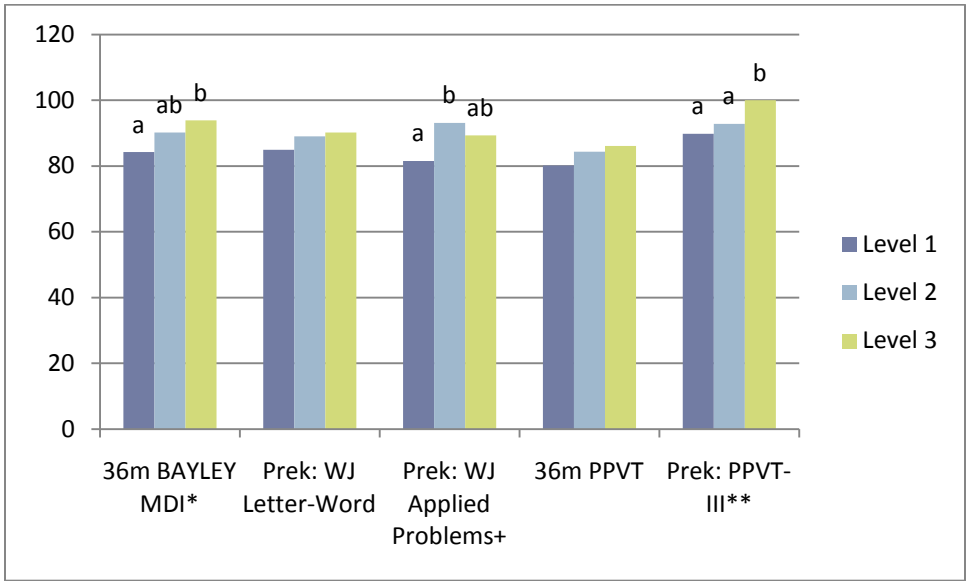
Table 7-T: Child Cognitive and Language Development: Predicted by Better Beginnings Minimums on FDCERS



Note: +p<.10, *p<.05, **p<.01; 24 Mo MDI, Level 1 significantly differs from Levels 2 and 3; 36 Mo MDI and PPVT, Level 1 significantly differs from Level 3.

There was also 36-month FDCERS data available in the EHSREP data set (Table 7-V). Again, there were noted differences in cognitive and language skills (Table 7-U).

Table 7-U: Child Cognitive and Language Development: Predicted by Better Beginnings Minimums on FDCERS at age 3



Note: +p<.10, *p<.05, **p<.01; 36 Mo MDI, Level 1 significantly differs from Level 3; Pre-K PPVT, Levels 1 and 2 significantly differ from Level 3.

Table 7-V: Family Day Care Environment Ratings at 36 Months of Age: Better Beginnings Levels and Child Outcomes at Ages 3 and 5

		36 Month			Pre-Kindergarten			Difference from 36 Months to Pre-K		
		N	Mean	SD	N	Mean	SD	N	Mean	SD
BAYLEY MDI; 36M*	Level 1	28	84.21	9.28	Not Applicable			Not Applicable		
	Level 2	31	90.16	12.77						
	Level 3	29	93.86	12.98						
WJ Letter-Word Identification; Pre-K^{ns}	Level 1	Not Applicable			28	84.93	12.29	Not Applicable		
	Level 2				28	89.04	15.96			
	Level 3				29	90.14	11.02			
WJ Applied Problems; Pre-K+	Level 1	Not Applicable			28	81.57	18.18	Not Applicable		
	Level 2				28	93.11	16.69			
	Level 3				29	89.31	23.71			
Leiter-R Sustained Attention; Pre-K^{ns}	Level 1	Not Applicable			26	11.31	2.99	Not Applicable		
	Level 2				27	10.59	2.75			
	Level 3				34	10.97	3.31			
Orientation/Engagement; BBRS@36M* and Leiter @Pre-K^{ns}	Level 1	32	3.46	.72	28	91.79	8.66	Not Applicable; Measures not Identical		
	Level 2	32	3.92	.56	26	93.00	9.53			
	Level 3	35	3.51	.75	33	94.97	7.23			
Emotion Regulation; BBRS@36M^{ns} and Leiter @Pre-K^{ns}	Level 1	32	3.81	.78	28	89.36	9.20	Not Applicable; Measures not Identical		
	Level 2	31	3.97	.62	27	90.89	12.97			
	Level 3	34	4.05	.69	33	93.09	8.07			
PPVT; 36M^{ns}, Pre-K**, Change^{ns}	Level 1	25	80.16	13.43	27	89.78	12.52	22	11.27	10.26
	Level 2	27	84.33	15.36	27	92.81	10.37	22	11.86	11.73
	Level 3	17	86.12	14.89	26	100.04	13.68	13	13.85	15.58
CBCL Aggressive; 36M*, Pre-K^{ns}, Change*	Level 1	31	15.39	6.54	30	10.73	5.61	28	-3.82	4.96
	Level 2	32	11.06	5.52	29	9.28	5.65	28	-1.29	6.04
	Level 3	39	11.68	6.71	36	11.60	8.44	34	.32	6.67

Note: ns=non-significant, +p<.10, *p<.05, **p<.01

As we saw with the ITERS/ECERS analysis, the FDCERS scores that are used for BB suggest that cognitive and language skills are negatively impacted when children are in centers at the lowest levels of quality. Again, although BB does not require Level 1 accreditation programs to complete anything more than a self-assessment on the corresponding ERS, it stands to reason that programs who apply for Level 2 that do not qualify will be given Level 1 status. There were relatively few children in this larger study who attended family child care programs, and these results should be validated in a larger sample.

Children in family child care centers with scores lower than 3 had significantly lower cognitive, math, and language skills that were in the same ranges seen for children in center-based early childhood programs. What was most worrisome is children in the lowest quality centers scored more than one standard deviation below the national average in cognitive, math, and language scores.

7.4 QUALITY IN SCHOOL-AGE PROGRAMS

There is limited but growing evidence that quality school-age programs can complement in-school learning and yield a range of positive impacts on children's development. With the increase in demand and funding for after-school programs, the demand for accountability and verified outcomes has increased. Findings from evaluations of after-school programs are mixed.

Durlak and Weissberg (2007) reviewed studies of 73 after-school programs with experimental or quasi-experimental designs of programs with the goals to improve participants' personal and social development. The authors grouped program effects into three broad areas: 1) school performance (performance on achievement tests, school grades, and school attendance); 2) social behaviors (positive social skills, problem behaviors, and drug use); and 3) attitudes and beliefs (bonding to school and self-esteem). They reported positive average effects in every category with the exception of school attendance. The review also examined the manner in which programs were conducted using a sequenced set of activities to achieve their goals, active learning techniques at least in part focused on personal or social development, and had explicit objectives for personal and/or social skills were the programs with the strongest findings.

On the other hand, the large-scale evaluation of 21st Century Community Learning Center after-school programs (James-Burdumy et al., 2005) identified no impact on participants' homework completion, on academic achievement (course grades in math, English, science, and social studies), nor on parental involvement. Higher levels of behavior problems for specific populations were reported. This evaluation did not measure program quality and reported considerable variation in the way programs were implemented.

Clearly, measuring quality in the program setting is important, but the current research literature is insufficient to help us understand how, why, and when program goals, characteristics, and quality produce positive outcomes for youth (Bodilly & Beckett, 2005; Miller, 2003). Our evaluation reviewed evidence available for two quality measures used for school-age BB standards, SACERS and the Youth Program Quality Assessment (Youth PQAAdams, Brickman, & McMahon, 2005b). A review by Yohalem and Wilson-Ahlstrom (2009) asserts that SACERS and Youth PQA share a common core including relationships, the environment, engagement, social norms, skill-building opportunities, and routines and structure. However, the two measures' abilities to predict child outcomes have not been proved.

7.4.1 SCHOOL–AGE CARE ENVIRONMENT RATING SCALE (SACERS)

The SACERS (Harms, et al., 1996) was designed to measure global quality in center-based settings for children ages 5 to 12. The scale includes 43 items categorized into six subscales: 1) Space and Furnishings, 2) Health and Safety, 3) Activities, 4) Interactions, 5) Program Structure, and 6) Staff Development. There is also a set of six supplementary items for programs serving school-age children with special needs. Like the other ERS, scores on each of the items are completed on a 7-point scale; with 1 representing "inadequate" care that compromises a child's development, 3 representing "minimal" or custodial levels of care, 5 representing "good" care that is developmentally appropriate, and 7 representing "excellent" care that "expands children's experiences, extends their learning, and provides warm and caring support" (Harms, et al., 1996, p. 1).

The SACERS authors state that the instrument has good internal consistency reliability and inter-rater agreement (Harms, et al., 1996). Data used for measures of internal consistency, reliability, and inter-rater agreement were drawn from a study of 24 centers in Canada that were observed by two independent observers. Internal consistency reliability estimates for each of the subscales were moderate to high, and Cronbach's Alpha coefficients ranged from .67 to .94. Internal consistency for the full scale was high (.95). Inter-rater agreement estimates (weighted Kappas) for each of the subscales were good to excellent, ranging from .79 to .91, with .83 for the total SACERS score.

A recent longitudinal study by Pierce, Bolt, & Vandell (2010) demonstrates relationships between observed after-school program quality and outcomes in middle childhood. The observational measure of quality was not the SACERS, but three features of quality that were found significantly related to similar constructs on the SACERS: positive child-staff relationships, available activities, and programming flexibility. Correlations were significant and ranged from .58 to .89. For children in Grade 2, positive staff-child relationships were found related to reading and math grades, controlling for children's scores in Grade 1. For children in Grade 3, positive staff-child relationships continued to predict grades for reading, above and beyond the child's grades in Grade 2. Another feature of program quality that was found related to child outcomes (in Grade 3 only) was availability of activities. In Grade 3, programs with greater availability of diverse age-appropriate activities had children with positive changes in math grades and work habits, controlling for the child's performance in Grade 2.

7.4.2 YOUTH PROGRAM QUALITY ASSESSMENT (YOUTH PQA)

The second measure of quality permitted for BB assessment is The Youth PQA Form A that consists of four subscales: Safe Environment, Supportive Environment, Interaction, and Engagement, and is scored by observation.[43] For an assessment tool to be useful in measuring diverse after-school programs, it must be based on a definition of quality as a "common, credible unit of study across programs and offerings" (Smith, Devaney, Akiva, & Sugar, 2009). Form A examines the quality of staff performances at the point of service, in particular the processes of staff-child and peer interactions, during program offerings as its unit of study.[44]

The tool's validation study was a four-year effort to develop and validate a quality assessment tool that could be used in a variety of youth after-school settings and involved 59 Michigan organizations, most providing daily or weekly after-school programs, serving 1,635 youth. The first draft of the Youth PQA was finished in 2002, and the validation study continued through 2005. The instrument was revised during these years, based on feedback from users, and in 2004 two major rewrites were completed. In 2005, the instrument was revised for use with Kindergarten to fourth grade (Younger Youth PQA; Adams, Brickman, & McMahon, 2005a) with about 80% identical content to the original Youth PQA, which is used for grades 4 through 12 (Smith, et al., 2009).

The Youth Development Strategies, Inc. (YDSI) Youth Survey was administered as part of the Youth PQA validation study, and trained external data collectors were used (Smith, et al., 2009). For each program, three observations were conducted for three different program offerings; the scores for the three

[43] A second part of these instruments that measures structural elements of quality, Form B (Adams, et al., 2005c), is not used for BB. Refer to 4.1.5 for further description.
[44] BB uses the terminology "classroom/program space" rather than "point of service."

offerings were then averaged to create a score that depicted overall quality. Program offerings were selected randomly for observation, and the evaluators noted that quality varied dramatically within organizations (Smith & Hohmann, 2005). All of the existing literature that we found relating to the Youth PQA as a valid and reliable instrument was based on the Youth PQA rather than the version for younger youth.

To assess internal consistency, Cronbach's alpha was calculated for the seven Youth PQA subscales. Subscales II, III, IV, V, and VI demonstrate acceptable levels of internal consistency (ranging from .64 to .85). However, Subscale I that measures the safety of the environment and Subscale VII that measures access have low internal consistency (.38 and .45, respectively) in the unacceptable range. Intraclass correlation coefficients for paired raters (N = 48 paired ratings) on all of the observational subscales except Safe Environment and Youth Centered Policies and Practices were in the acceptable ranges.

There are three ways to use the Youth PQA, each with its particular uses and strengths. Lower-stakes program self-assessment provides "rough data to get staff thinking and discussing program quality in the context of best practice," while higher-stakes external assessment provides "precise data for internal and external audiences for evaluation, monitoring, accountability, improvement, and reporting" (Smith, et al., 2009). A third method combines these two approaches.

Using trained outside observers who complete the Youth PQA with a random selection of several program offerings in one organization will result in scores with the highest levels of reliability and validity, though at a higher cost to the program than self-assessment. Two major limitations of self-assessment are 1) data is likely to be biased in a positive direction, and 2) given the team approach of the self-assessment, it does not produce a "consistent, definable unit of study, that is, a single staff performance during a defined offering" (Smith, et al., 2009). Thus, programs conducting self-assessments cannot realistically be compared. The strength of self-assessment is the dialogue that it creates among staff (Smith & Hohmann, 2005). Staff trained in administering the Youth PQA develop common understandings of program quality (Smith, et al., 2009).

The developers and evaluators of the Youth PQA found that programs using this instrument most successfully began with self-assessment, to contribute to the formation of a team culture and to increase the comfort level of staff with the instrument and the process, so that more objective external assessment of staff performance against quality standards could follow (Smith, et al., 2009).

In 2008 the Arkansas Out-of-School Network partnered with the David P. Weikart Center for Youth Program Quality to undertake a pilot project with a group of Arkansas after-school programs. The project provided training and technical assistance in quality assessment and improvement planning using the Youth PQA. The goals were to provide these programs with a data-driven continuous improvement structure, to provide an intensive experience that could lead to "visible change" in staff and program quality, and to learn about how the use of this assessment tool fits within the context of Arkansas after-school programming. The evaluation of this pilot project revealed high participation and intensity and improved perception of staff skills and program quality.

7.4.2.1 CROSSWALK FOR YOUTH PQA AND SACERS

In lieu of the SACERS, BB allows programs to choose to use Youth PQA and the Younger Youth PQA for school-age programs. In a review of 10 assessment tools, The Forum for Youth Investment found that only two – the SACERS and the Youth PQA – covered all their selected purpose categories: improvement, monitoring/accreditation, and research/evaluation. Observation was the primary data collection method for each of these tools. Three of the instruments also use document review, which the SACERS and the Youth PQA do not. The authors found the Youth PQA to be the strongest of the 10 in terms of technical properties (Yohalem, et al., 2009; see figure 4, p. 14). The PQA reviewed in this report was the youth version for grades 4-12.

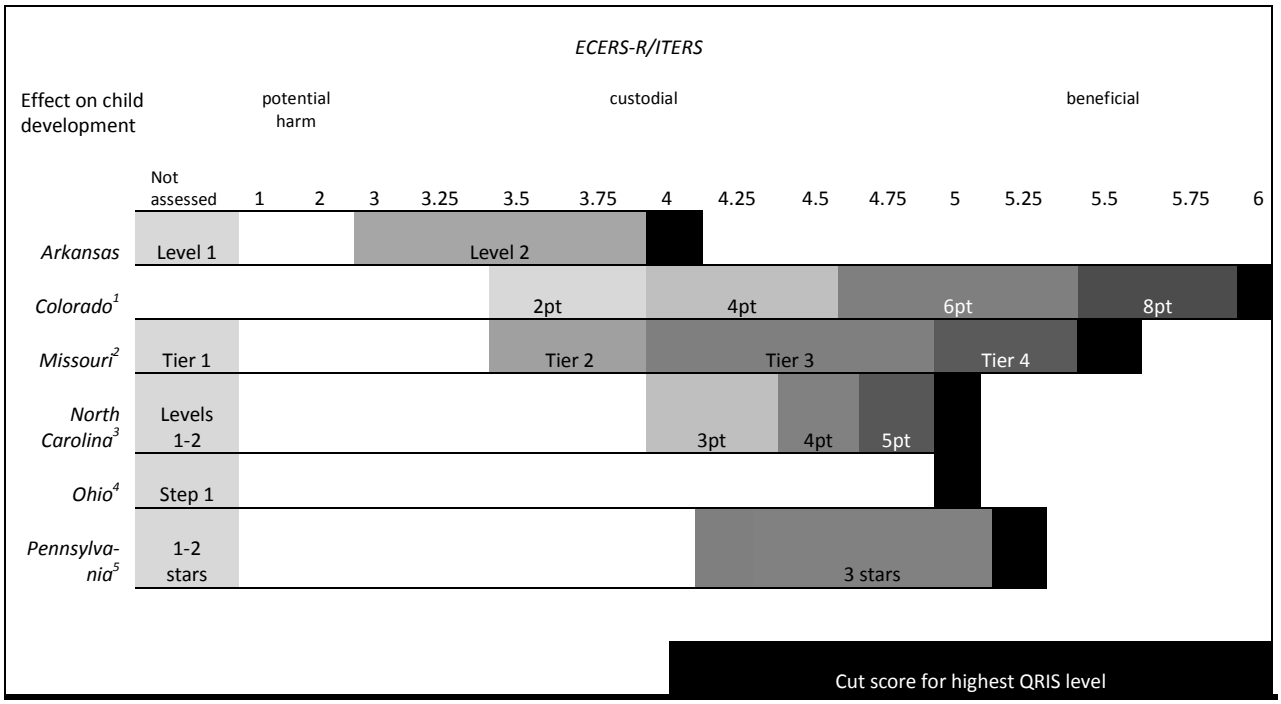
Examining areas of overlap for the SACERS and Form A of the Youth and Younger Youth PQA suggests moderate overlap between the instruments (see Appendix 5 in the *Evaluating Arkansas' Path to Better Child Outcomes: Technical Appendices* available online at www.arbetterbeginnings.com). The SACERS is a global measure of program quality that includes many elements of the provisions available within the care environment, while the Form A of the Youth and Younger Youth PQA focuses more attention on the process components of youth programs with greater emphasis on staff-youth interactions, the socio-emotional climate of the program, and youth engagement.

Given that the instruments were designed to measure different components of programming, it is unknown how the use of equivalent scoring criteria for BB ratings will function during implementation of the QRIS. While there is some overlap in content of the school-age care environment with the two assessments, the research team could identify no existing data that would support the use of identical scoring criteria for the assessments, Furthermore, existing data of SACERS, Youth PQA, and Younger Youth PQA were not available, so it is unknown how the current BB Levels may work for school-age children.

7.5 KEY STATES COMPARISON

Out of 17 statewide quality rating systems, 13 use the ERS but with considerable variation in timing and frequency of assessments, in percentage of classrooms/groups observed, in how scores are averaged, and in where thresholds are set for each level. Four of the five key comparison states make use of the original ERS. Oklahoma, the exception, uses its own checklist adapted from the ERS. As illustrated in Table 7-W, Arkansas is similar to all key states except Colorado in allowing self-assessment for entry into the lowest rating system level. However, Arkansas diverges when it comes to scores required for entry into upper levels. All key states have upper levels that require scores in the good range. In contrast, Arkansas accepts scores in the adequate range, lower than other states, for mid and high levels. Each of the five key states has conducted evaluations of their rating systems, although the research questions varied and most did not investigate child outcomes.

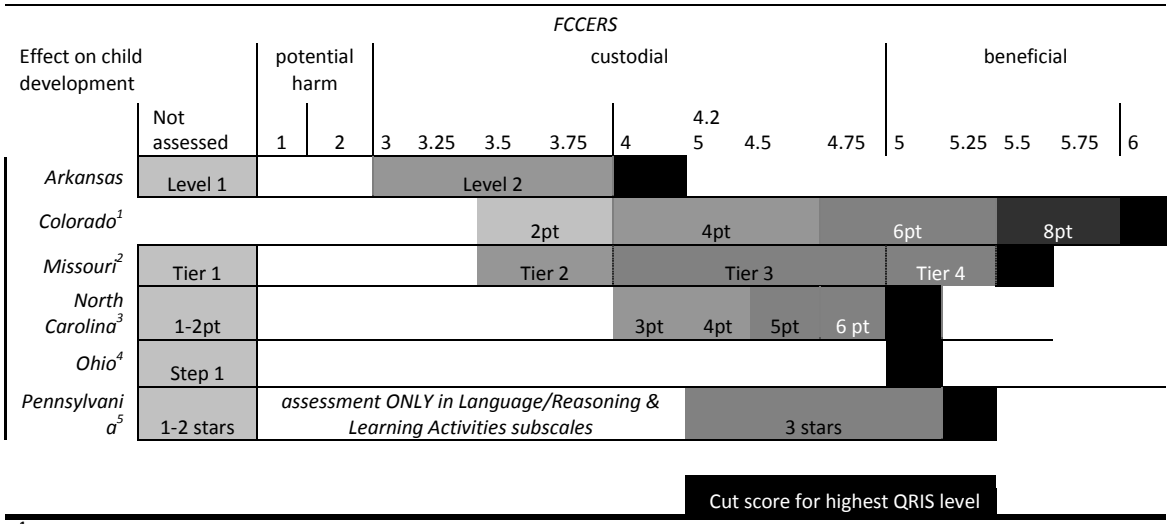
Table 7-W States ERS Score Requirements for Centers



[1]**CO**: Centers score at least 3.5 for 2 points, 4 for 4 points, 4.7 for 6 points, 5.5 for 8 points, or 6 for 10 points.
[2]**MO**: Tier 2 centers average 3.5; no room score less than 3. Tier 3 centers average of 4; no room score less than 3.5. Tier 4 centers average 5; no room score less than 4.5. Tier 5 centers average 5.5.
[3]**NC:** Meet minimum licensing for 1-2 points. For 3 points, the lowest classroom scores at least 4. For 4points, centers average 4.5 with no score less than 4.0. For 5 points, centers average 4.75 with no score less than 4.0. For 6points, centers average 5 with no score lower than 4.0. For 7points, the lowest classroom score must be at least 5.
[4]**OH**: Prefers that centers meet recommended ratios, but will allow ERS scores of 5 or above as an alternative for Step 2 & 3.
[5]**PA**: For 2 stars, self assessment is done. For 3 stars, centers average 4.25 with no scoreless than 3.5. For 4 stars, centers average 5.25 with no score less than 4.25.
OK: does not use the full ERS to conduct evaluations and is therefore not included. They use checklists adapted from ERS to conduct evaluations. No cut scores for each level are designated.

Distribution of state ERS scores is similar in family day care rating requirements (Table 7-X). Like BB, most states do not require FCCERS assessment in the bottom level of quality, and Arkansas accepts lower scores for entry into upper levels than all other comparison states.

Table 7-X State ERS Score Requirements for Family Day Care



[1]**CO**: 3.5-3.99 for 2 pts, 4-4.69 for 4 pts, 4.7-5.49 for 6pts, 5.5-5.99 for 8 pts, 6-7 for 10 pts.
[2]**MO:** Tier 1 minimum license, Tier 2 3.5 and above, Tier 3 4.0 and above, Tier 4 5.0 and above, Tier 5 5.5 and above.
[3]**NC:** 3 pts for 4.0, 4 pts for 4.25, 5 pts for 4.5, 6 pts for 4.75, and 7 pts for 5.0.
[4]**OH**: For Tier 2 & 3 providers have a choice of meeting recommended ratios, participating in the NAFCC accreditation process, or achieving ERS overall score of 5 with no score less than 4 on each subscale.
[5]**PA**: Programs must average 4.25 on subscales for 3 start, average 5.25 for 4 stars.
OK: does not use the full ERS to conduct evaluations and is therefore not included. They use checklists adapted from ERS to conduct evaluations. No cut scores for each level are designated.

All of the states use SACERS as a measure of program quality. As seen with ECERS-R and ITERS, BB awards quality ratings when SACERS scores are lower than comparison states.

- **Colorado:** Colorado uses a point system. Two points are awarded for ERS scores from 3.5 to 4; 4 points are awarded for scores from 4 to 4.75; 6 points for scores from 4.75 to 5.5; 8 points for scores from 5.5 to 6; and 10 points for programs scoring 6 or higher.
- **North Carolina**: The lowest score listed to be eligible to achieve points toward a star rating is 4.0. To earn 3-7 points toward total rating, programs must score between 4.0 and 5.0 for each group assessed.
- **Ohio**: A score 5 may be used as an alternate pathway to 2- and 3-star ratings in lieu of meeting ratio requirements.
- **Oklahoma**: A SACERS assessment is completed within one year of receiving 2-star status and once every three years thereafter. No minimum scores are designated for star ratings. Maintaining SACERS 5.0 score is incorporated as part of Master Teacher qualifications for 2 and 3 stars.
- **Pennsylvania**: The average assessed SACERS score must be 4.25 with no individual class under 3.5 for 3 stars. For 4 stars, the average assessed score is 5.25 with no class under 4.25.
- **Missouri**: An average score of 3.5 with no group below 3 is required for Tier 2, and an average score of 4.0 with no group below 3.5 is required for Tier 3. An average score of 5 with no group below 4.5 is required for Tier 4. An average score of 5.5 is required for Tier 5.

7.5.1 COLORADO

Colorado, the first state to implement and conduct validity testing of a quality rating system, did not find a relation between the state's star ratings, which use ECERS and ITERS, and child outcomes (G Zellman, et al., 2008). Unfortunately, conclusions of this study are severely limited by methodological concerns and very high attrition within the sample. Only 7% of the children initially enrolled in the study remained at the final stage of the project.

7.5.2 MISSOURI

Missouri has eight rating categories. Five categories are given 5 points each. Along with two categories that address staff and administrator training, Learning Environment receives 10 points. Missouri rates half of its classrooms using the appropriate ERS. Tier 1 is minimum licensing and formal assessments are not conducted. To enter Tier 2, a program must receive an average ERS score of 3.5 with no class receiving below a 3.0, and to enter Tier 2 a program must score an average of 4.0 with no room rated less than 3.5. Children in 4- and 5-star programs showed statistically significant gains in social and behavioral skills, motivation, self-control, and positive adult relationships. **The validation study of the rating system throughout its first year detected significant gains in social and behavioral skills and motivation among children in 3-star programs compared to children in 1- and 2-star programs (Thornburg, et al., 2009). Children in the state's Tier 1 and Tier 2 lost social and behavioral skills within the school year. Children in poverty gained vocabulary while when in medium and high quality programs but lost vocabulary in low quality programs. Children demonstrated gains in social and emotional development only in high quality programs.**

7.5.3 OHIO

Ohio incorporated the ERSs into its original system design, but the "Parents and Staff" subscale was not used because of perceived lack of relevance and the additional time and expense required to observe those indicators. ECERS-R scores from the rating system pilot clustered around 5, "good," with little difference between each of the levels. The state eliminated ERS as a key component of its center-based rating but continues to accept an ERS score of 5 or above as an alternative to meeting ratio and group size requirements in Steps 2 and 3.

When the rating system was rolled out state-wide, the team collected parent, teacher, and director surveys and conducted a battery of assessments on 138 children in 28 randomly selected classrooms. Results, though limited by the small sample size, showed significant differences between the rating levels. **"Controlling for differences in socio-economic status and family chaos, children in step level two and three programs were marked higher on measures of problem solving than were children in step level one programs. There was no difference in problem solving skills as reported by parents for children between step level two and step level three programs…children in step level one and two programs had lower independence than children in step level three. In terms of self regulation, children in step level one had lower adjusted mean scores than children in either step level two or three" (Ohio Collaborative, 2009, p.17).**

7.5.4 OKLAHOMA

Oklahoma uses a checklist that is adapted from ERS. Like Arkansas, Oklahoma uses self-assessment for its lower levels. Within one year of achieving 2-star status and thereafter every three years, the center is assessed using an adapted version of the ERS.

Oklahoma's star system validation study (Norris, et al., 2003) did not observe child outcomes, rather the study's goals were to determine whether global quality had increased as a result of the system and to discern which program and staff characteristics, such as ratios or teacher-child interactions, stood out as predictors of global quality. Eighty percent of centers (n=336) had ECERS-R ratings of 5 or higher, an increase from prior data. In infant-toddler classrooms 3-star/Accredited centers with average ITERS scores of 5.91 had higher sensitivity scores than 1-star centers with average 4.69 ITERS. Results of regression analysis "indicated that each set of predictors, subsidy density, licensing variables, and the 2-Star criteria, had a statistically significant influence on child care quality. The 2-Star criteria were the most powerful predictors, accounting for 29% of the variance in composite quality. Of the 2-Star criteria, the most important were Master teacher-child ratio, parent involvement, and clearly defined learning centers" (p.4).

7.5.5 PENNSYLVANIA

Pennsylvania's Keystone STARS Quality Rating System uses self assessment for lower levels. Pennsylvania requires formal assessments with the result of at least 4.25 among all classrooms with no score below 3.5 to achieve 3 Stars and an average score of 5.25 with no score below 4.25 for 4 stars.

The state's 2006 evaluation of child care did not collect child outcomes. The key research questions were whether the state had reversed a downward trend in global quality of care that occurred from 1996 to 2002 and whether there were significant differences in global quality from one level to the next in their rating system.

Four years after their rating system was introduced, the evaluators collected data from 356 centers, some in the rating system, others not. In the seven years prior to the rating system, quality had dropped from 4.5 to 3.9. The state confirmed that in 2006 the decline had ended. States that were not a part of the rating system were scoring an average of 3.94. Those participating in the program had scores ranging from an average of 4.11 at the initial "Start with Stars" level up to 5.42 for the Star 4 level.

7.6 CONCLUSIONS

The ERS for all three types of care are valid, reliable global measures, assessing both structural and process program features. The center-based tools demonstrate a modest ability to predict outcomes related to child communication, language development, literacy, cognitive development, social-emotional attachments, counting and applied math, and physical activity. In Head Start centers, ERS scores are usually in the mid-to-high range, typically exceeding the highest level of Better Beginnings. The formal assessment of process components of care completed by Head Start reviewers could be substituted.

Low ERS quality in center-based care has been associated with elevated stress, anger and defiance, and setbacks in vocabulary and applied math development. Studies in family-based care have identified relationships between ERS quality and social-emotional and language development. Recent large multi-state studies with sophisticated controls indicate that the higher the quality, the greater the effect on child outcomes. Better Beginnings is similar to other comparison state rating systems in permitting self-assessment for entry-level quality ratings but diverges from other states by permitting scores in the adequate range for upper tiers of quality. Among other state quality rating systems that include the ERS suite, the lowest average ERS score for center-based care typically recognized is either 3.0-3.75 (N=13) or 4.0-4.5 (N=6). The highest average ERS score recognized is typically in the 5.0-5.5 range (N=13) or higher (N=4). This is similar for family child care programs, with the lowest average ERS score recognized is usually in the 3.0-3.5 range (8) or in the 4.0-4.5 range (7), and the highest average ERS score is in the 5.0- 5.5 range for most (12) family child care programs (Kathryn Tout, et al., 2010). The highest range for Better Beginnings is 4, which represents a significant divergence from what is seen in the majority of rating systems across the nation. Further, the Missouri rating system evaluation found that all children lost social skills and children in poverty lost vocabulary when enrolled in centers with lower quality ratings (Missouri Tiers 1 and 2) which are comparable to Better Beginnings Levels 1 and 2 in criteria for environmental ratings.

Overall, the findings from studies of school-age programs provide evidence that positive impacts on child outcomes can be achieved. There is less evidence that quality as measured by the SACERS is related to more optimal development, because the bulk of the extant literature compares children with and without after-school program experiences and focuses less on variations in quality on child outcomes. The Pierce, Bolt, and Vandell (2010) study does provide evidence that elements of program quality (correlated with items measured with the SACERS) are related to growth in language, math, and social skills. In lieu of the SACERS, the BB standards allow programs to choose to use Youth PQA and the Younger Youth PQA for school-age programs. The SACERS is a global measure of program quality which includes many elements of the provisions available within the care environment, while the Form A of the Youth and Younger Youth PQA focuses more attention on the process components of youth programs with greater emphasis on staff-youth interactions, the socio-emotional climate of the program, and youth engagement. The instruments were designed to measure different components of programming and it is unknown how the use of equivalent scoring criteria for Better Beginnings' ratings will function during state-wide implementation. Existing data of SACERS, Youth PQA, and Younger Youth PQA were not available, so it is unknown how the current Better Beginnings Levels may work for school-age children.

7.7 INTERACTIONS AND OVERALL PROGRAM QUALITY

Because interaction measures appear to have stronger associations with child outcomes than global quality measures, we investigated whether children in higher Levels of BB (Levels 2 and 3) but with Interactions scores lower than the BB cut score were functioning differently from children in centers where Interaction scores more closely approximated the overall ERS.

Data from the Head Start FACES 2003 cohort were used, as they are the most recently collected data that are publicly available, and measurement constructs most closely approximate those used in BB. Further, there was a great number of children assessed in the FACES study in center-based care (N=1778). Using this data, however, limits the findings to children in preschool programs.

The Early Childhood Environment Rating Scale – Revised (ECERS-R) was collected in each classroom for children in the Head Start FACES study. Because data were collected in Head Start programs, there is more variability at the high end of quality; there were 63 children in Level 1 programs (ECERS-R scores below 3), 308 children in Level 2 programs (ECERS-R scores between 3 and 4), and 1,207 children in Level 3 programs (ECERS-R scores over 4). The ECERS-R Interactions subscale includes five items that measure supervision of gross motor activities, general supervision of children, discipline, staff-child interactions, and interactions among children. Using cut scores on the Interactions subscale that parallel those of BB, we were able to examine BB Level and Interaction scores together. We were most interested in understanding if overall quality (Total ECERS-R score) solely predicted child outcomes or if Interactions scores were related to child outcomes beyond the overall program quality score.

When comparing children attending Level 2 centers with Interactions scores lower than the minimum cut score for the overall ECERS-R (score less than 3) to those in centers where Interactions were at or above the minimum score, there were significant differences. Children in Level 2 centers with lower interaction scores had significantly lower letter-word identification scores, and trends toward significantly lower story and print concepts scores. Furthermore, although at trend level, these children also differed in social outcomes; with higher levels of teacher-reported aggressive and hyper behavior (Table 7-Y). Although not significant, across each cognitive and social outcome measured, children in centers where interactions were lower than the cut score for the overall ECERS-R for centers to achieve the Level 2 rating had less optimal outcomes.

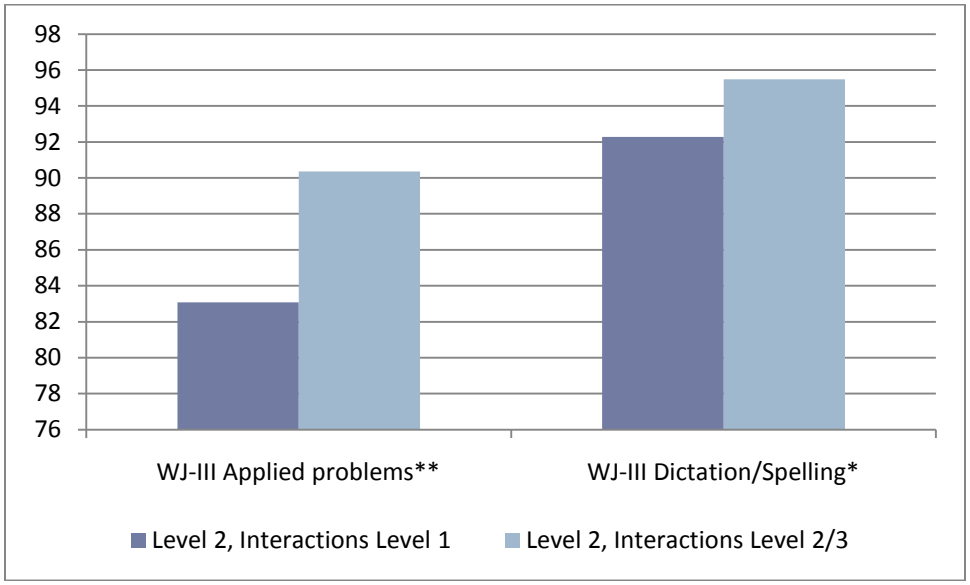
Table 7-Y Level 2 Comparison of ECERS-R and Interactions Subscale Scores on Child Outcomes in Fall and Spring

Constructs		Fall Scores			Spring Scores		
		N	Mean	SD	N	Mean	SD
PPVT Standard Score: Fall[ns], Spring[ns]	Level 2, Interactions Level 1	64	80.99	10.62	63	83.22	9.34
	Level 2, Interactions Level 2/3	234	82.81	11.18	227	83.99	10.74
WJ-III Letter-word Identification: Fall*, Spring[ns]	Level 2, Interactions Level 1	64	88.45	17.03	61	98.08	17.55
	Level 2, Interactions Level 2/3	233	94.65	17.74	226	99.56	16.05
WJ-III Applied problems standard score: Fall[ns], Spring**	Level 2, Interactions Level 1	64	82.70	17.54	62	83.08	16.89
	Level 2, Interactions Level 2/3	233	86.75	18.28	227	90.36	16.36
WJ-III Dictation/Spelling standard score: Fall[ns], Spring*	Level 2, Interactions Level 1	63	92.11	12.19	61	92.28	10.15
	Level 2, Interactions Level 2/3	227	94.58	12.38	212	95.48	13.00
Story and print concepts: Fall+, Spring+	Level 2, Interactions Level 1	63	3.58	2.02	63	3.58	2.02
	Level 2, Interactions Level 2/3	227	4.19	2.24	227	4.19	2.24
Pre-CTOPP: Fall[ns], Spring*	Level 2, Interactions Level 1	63	3.58	1.51	62	4.19	1.52
	Level 2, Interactions Level 2/3	234	3.82	1.48	227	4.71	1.54
Social skills score: Fall[ns], Spring[ns]	Level 2, Interactions Level 1	70	15.17	4.498	65	17.65	4.73
	Level 2, Interactions Level 2/3	227	16.04	4.610	230	17.69	4.87
Teacher reported aggressive behavior: Fall+, Spring**	Level 2, Interactions Level 1	69	1.65	1.939	64	1.94	2.08
	Level 2, Interactions Level 2/3	230	1.23	1.759	229	1.19	1.75
Teacher reported hyper behavior: Fall+, Spring[ns]	Level 2, Interactions Level 1	69	1.39	1.467	66	1.21	1.47
	Level 2, Interactions Level 2/3	231	1.06	1.361	230	1.04	1.51
Teacher reported withdrawn behavior: Fall[ns], Spring[ns]	Level 2, Interactions Level 1	69	1.93	2.110	66	1.58	1.65
	Level 2, Interactions Level 2/3	231	1.73	2.309	228	1.58	2.18
Teacher reported total behavior problems: Fall[ns], Spring[ns]	Level 2, Interactions Level 1	69	4.97	4.482	67	4.66	3.96
	Level 2, Interactions Level 2/3	232	4.04	4.506	230	3.80	4.68

Note: ns=non-significant, +p<.10, *p<.05, **p<.01

We conducted similar analyses using the Level 2 centers examining child outcomes in the spring of the school year. Again, across most outcomes measured, children in programs where interaction scores on the ECERS-R met the critical threshold for the rating score of the full ECERS-R scale (Table 7-Z). There were several significant differences in scores for children as well. Children in centers where both ECERS-R and Interaction scores both met or exceeded the cut score to achieve a BB Level 2 rating had significantly higher math skills (applied problems), spelling skills with dictation, phonemic processing (Pre-CTOPP), and significantly less aggressive behavior per teacher report. There was also a trend in the direction of better story and print concepts as well. Table 7-Z provides graphical representation of significant findings for the Woodcock-Johnson subscales; Applied Problems and Dictation and Spelling.

Table 7-Z: Comparison of ECERS-R and Interactions Subscale Scores on Child Woodcock-Johnson Outcomes in Spring



Note: +p<.10, *p<.05, **p<.01

We also examined children in BB Level 3 programs, comparing children in centers where interactions subscale scores did and did not meet the minimum threshold for a rating of Level 3 (score of 4 or higher) on the ECERS-R. There are few children who are in higher quality centers where interaction scores are lower (N=46) compared with over 1,000 children (N=1161) where interactions are higher. We are providing descriptive statistics and results from means comparisons, but we interpret findings with a note of caution as there are relatively few cases to reliably interpret comparisons (Table 7-AA). All of the significant or trend level differences were in social outcomes, with significantly lower overall social skills and significantly more aggressive behavior reported for children in Level 3 centers with Interactions scores lower than 4 (the minimum ECERS-R score for the Level 3 rating). We conducted similar analyses using the Level 3 centers examining child outcomes in the spring of the school year (Table 7-AA). Again, significant differences were demonstrated for social outcomes. Children had less optimal scores in social skills and aggressive behaviors in programs where interactions did not meet the minimum for the rating.

Programs with Interaction subscale scores that do not meet the minimum criterion score for the overall ERS needed to achieve a particular BB Level may result in less optimal outcomes for children. There is some evidence that cognitive skills are impacted. We also see consistency across findings for some elements of social development (such as aggressive behavior). These findings highlight the relative importance of interactions beyond the general quality of a preschool program.

Table 7-AA: Level 3 Comparison of ECERS-R and Interactions Subscale Scores on Child Outcomes in Fall and Spring

Constructs		Fall Scores			Spring Scores		
		N	Mean	SD	N	Mean	SD
PPVT Standard Score: Fall^{ns}, Spring^{ns}	Level 3, Interaction Lower	37	86.41	9.37	38	83.57	14.95
	Level 3, Interaction Level 3	937	85.59	12.51	967	86.88	12.44
WJ-III Letter-word Id: Fall^{ns}, Spring^{ns}	Level 3, Interaction Lower	37	90.37	16.08	38	94.37	17.03
	Level 3, Interaction Level 3	938	92.67	16.28	963	97.49	15.71
WJ-III Applied problems: Fall^{ns}, Spring^{+}	Level 3, Interaction Lower	36	88.38	17.22	38	85.97	20.21
	Level 3, Interaction Level 3	926	89.45	17.34	963	92.02	15.87
WJ-III Dictation & Spelling: Fall^{ns}, Spring^{ns}	Level 3, Interaction Lower	35	91.68	10.16	37	94.51	10.74
	Level 3, Interaction Level 3	880	93.49	11.45	935	93.95	12.22
Story & print concepts: Fall^{ns}, Spring^{ns}	Level 3, Interaction Lower	40	4.17	2.63	40	4.17	2.63
	Level 3, Interaction Level 3	1072	4.67	2.46	1072	4.67	2.46
Pre-CTOPP: Fall^{ns}, Spring^{ns}	Level 3, Interaction Lower	37	4.06	1.71	37	4.49	1.73
	Level 3, Interaction Level 3	932	3.88	1.66	963	4.72	1.77
Social skills score: Fall*, Spring*	Level 3, Interaction Lower	42	14.00	4.72	41	15.93	5.25
	Level 3, Interaction Level 3	1142	15.72	4.66	1097	17.60	4.32
Teacher report aggressive behavior: Fall*, Spring*	Level 3, Interaction Lower	40	2.20	2.32	41	2.02	2.19
	Level 3, Interaction Level 3	1147	1.38	1.88	1101	1.31	1.84
Teacher reported hyper behavior: Fall^{ns}, Spring+	Level 3, Interaction Lower	41	1.61	1.86	41	1.51	1.65
	Level 3, Interaction Level 3	1148	1.14	1.44	1102	1.00	1.36
Teacher report withdrawn behavior: Fall^{ns}, Spring^{ns}	Level 3, Interaction Lower	39	2.13	2.18	38	1.61	2.03
	Level 3, Interaction Level 3	1141	2.06	2.25	1092	1.80	2.10
Teacher report total problems: Fall^{ns}, Spring^{ns}	Level 3, Interaction Lower	41	5.83	5.44	41	5.24	4.59
	Level 3, Interaction Level 3	1155	4.58	4.61	1104	4.12	4.32

Note: ns=non-significant, +p<.10, *p<.05, **p<.01

While it is clear from the extant literature that teacher-child interactions are a potential mechanism through which quality programs and teachers' training and education help achieve impacts on children's outcomes, the addition of a measure of teacher-child interactions to the BB system would be costly and onerous – and could potentially lead to provider resentment (Zellman & Perlman, 2008). Ideally, data collected as part of the already implemented BB process will be used to document interactions. These findings could be used to support changes to the BB standards wherein programs not only meet an overall minimum ERS score to achieve a particular Level, but also meet the same minimum for the Interactions subscale. Further, support for programs in the area of teacher-child interactions would be warranted when scores on the Interaction subscale of the ECERS-R are substantially lower than the overall score.

7.8 RECOMMENDATIONS FOR CONTINUOUS QUALITY IMPROVEMENT

There was much discussion in the RAND evaluation, *Child-Care Quality Rating and Improvement Systems in Five Pioneer States: Implementation Issues and Lessons Learned* about piloting and making adjustments to new QRIS/QRS. Zellman and Perlman (2008) suggest that states limit changes to the system after implementation statewide and recommend piloting of the system as the first step to understanding what changes may need to be made. They warn that changes, including raising the bar to prevent provider complacency, may create confusion for parents and may undermine their trust in the system.

The authors recommend that states "set up a system of continuous quality improvement with clear incentives for improvement and a substantial number of rungs to climb." Many rungs encourages programs to strive for the highest quality without imposing new requirements and prevents providers from shifting to a "maintenance" mode in which they no longer strive to improve. Including many rungs also makes progress more attainable at the lower quality Levels, thereby facilitating provider engagement.

In Arkansas, the QRIS includes some of the issues that Zellman and Perlman discuss in the RAND evaluation; namely, the system was developed with lower Levels that encourage provider participation, but not higher Levels that will promote centers to continue to strive to make improvements. The highest BB Level sets a relatively low standard for providers to attain (with minimum scores on the ERS at our highest level being 4; between minimal and good quality), and there are not incentives for programs to improve beyond that level. Further, BB Level 1 requires no actual assessment on the part of the state. Other states have included similar levels but have been clear that they are a means to assist programs in preparation for QRIS ratings. (For example, Ohio's Step Up to Quality initially included a level similar to BB Level 1 but determined in piloting that it would be considered a Level 0 "Getting Ready" step during statewide implementation.)

Findings from the validation efforts we undertook gave cause for concern. Children in programs with low ERS scores (below 3) lost key academic skills in toddlerhood and had scores on standardized measures of cognitive and academic development in preschool far below the national norm for children of the same age. Giving a quality accreditation to a program that has the potential to yield negative impacts for children is unsettling. However, changing or renumbering the existing BB standards would prove problematic when resentment for the system already exists. Therefore, we recommend additional Levels to BB to encourage quality providers to continue striving to provide optimal educational and care opportunities for our state's children.

The UAMS evaluation team has used existing data to help guide decisions around the inclusion of additional rungs to the BB quality ladder. While we recognize that there are other elements of the current standards that cannot be evaluated, we believe that demonstrating additional variance in child outcomes attributable to structural characteristics of the programs is a helpful step in making recommendations. Cut scores of the key comparison states (Table 7-W) and research literature on the relationship between program quality scores and children's outcomes were used to inform analyses (Mashburn, 2008). We used data collected as part of the Head Start FACES 2003 study to evaluate additional cut scores on the ECERS that could be employed at higher BB Levels.

Current BB standards include three Levels: Level 1 includes a self-assessment and centers with ERS scores lower than 3 that do not qualify for Level 2; Level 2 includes centers with ERS scores between 3 and 4; and Level 3 includes centers with ERS scores greater than 4. The UAMS evaluation team proposes revision to Level 3 and the addition of two Levels. Revised Level 3 would include centers with ERS scores between 4 and 5. The added Level 4 would include centers with scores between 5 and 5.5. The added Level 5 would include centers with scores of 5.5 and greater. The score of 5.5 was chosen because it is a common cut score for the upper levels of our comparison states. Data from the Head Start FACES 2003 evaluation resulted in the greatest number of children in programs with scores falling at Level 3 (Table 7-BB).

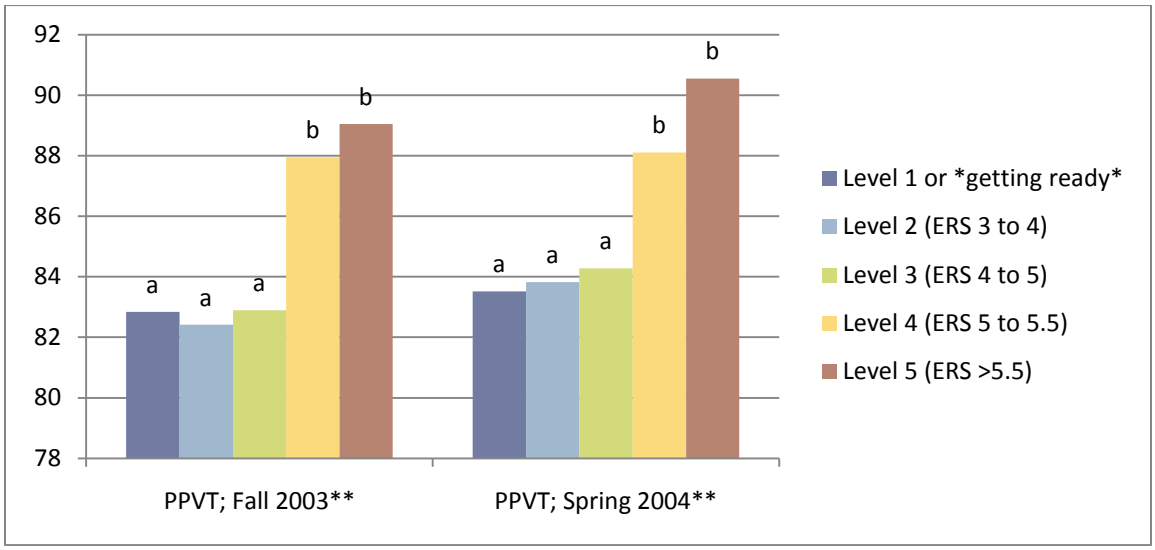
Table 7-BB Proposed Better Beginnings Levels (based on ECERS Scores in Head Start Programs)

		Fall 2003			Spring 2004		
		N	Mean	SD	N	Mean	SD
PPVT Standard Score; Fall 2003** and Spring 2004**	Level 1 or *getting	61	82.84	12.72	55	83.52	12.09
	Level 2 (ERS 3 to 4)	298	82.42	11.07	290	83.82	10.44
	Level 3 (ERS 4 to 5)	517	82.90	11.83	552	84.28	12.22
	Level 4 (ERS 5 to 5.5)	142	87.95	10.53	144	88.11	12.25
	Level 5 (ERS >5.5)	315	89.04	13.05	309	90.55	12.25
WJ-III Letter-word standard score; Fall 2003 and Spring 2004+	Level 1 or *getting	62	87.65	15.99	55	93.51	15.28
	Level 2 (ERS 3 to 4)	297	93.31	17.74	287	99.25	16.36
	Level 3 (ERS 4 to 5)	518	92.53	17.16	550	97.20	16.24
	Level 4 (ERS 5 to 5.5)	143	91.72	15.27	144	95.96	17.11
	Level 5 (ERS >5.5)	314	93.06	15.19	307	98.34	14.15
WJ-III Applied problems standard score; Fall 2003** and Spring 2004**	Level 1 or *getting	62	83.11	17.39	55	86.84	15.38
	Level 2 (ERS 3 to 4)	297	85.88	18.17	289	88.80	16.72
	Level 3 (ERS 4 to 5)	514	87.41	17.66	552	89.55	16.35
	Level 4 (ERS 5 to 5.5)	139	89.53	16.08	142	93.57	15.13
	Level 5 (ERS >5.5)	309	92.68	16.85	307	95.01	15.42
WJ-III Dictation - Spelling standard score; Fall 2003 and Spring 2004+	Level 1 or *getting	62	90.26	10.58	52	90.75	11.78
	Level 2 (ERS 3 to 4)	290	94.04	12.36	273	94.77	12.47
	Level 3 (ERS 4 to 5)	481	93.77	11.38	535	94.53	11.91
	Level 4 (ERS 5 to 5.5)	134	92.99	10.86	139	92.57	11.67
	Level 5 (ERS >5.5)	300	93.06	11.70	298	93.62	12.80
Story and print concepts IRT true-score; Fall 2003** and Spring 2004**	Level 1 or *getting	63	2.58	1.92	55	4.04	2.63
	Level 2 (ERS 3 to 4)	306	2.58	1.67	290	4.06	2.21
	Level 3 (ERS 4 to 5)	669	2.56	1.76	631	4.27	2.34
	Level 4 (ERS 5 to 5.5)	173	3.17	2.23	161	4.49	2.34
	Level 5 (ERS >5.5)	346	3.61	2.24	320	5.48	2.58
CTOPP Elision IRT true-score; Fall 2003** and Spring 2004**	Level 1 or *getting	62	3.87	1.54	55	4.24	1.50
	Level 2 (ERS 3 to 4)	297	3.77	1.49	289	4.60	1.55
	Level 3 (ERS 4 to 5)	514	3.70	1.53	549	4.46	1.56
	Level 4 (ERS 5 to 5.5)	143	4.08	1.88	142	4.62	1.82
	Level 5 (ERS >5.5)	312	4.10	1.75	309	5.21	1.97

Note: ns=non-significant, +p<.10, *p<.05, **p<.01

Across the fall and the spring, there were significant differences seen for children in proposed Level 4 and Level 5 programs on receptive vocabulary (Table 7-CC). There were not significant changes in receptive vocabulary.

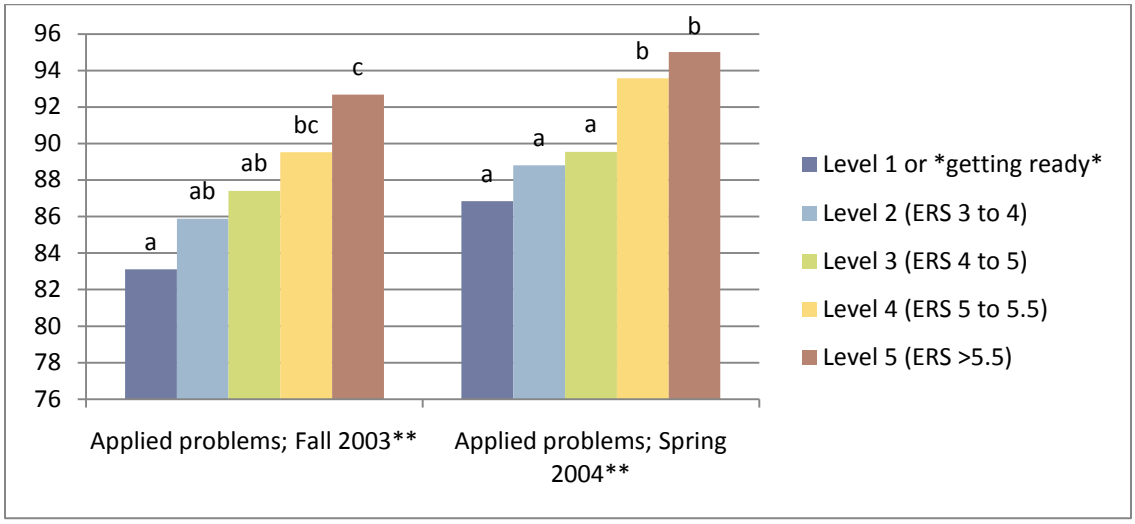
Table 7-CC Children's Receptive Vocabulary by Proposed Better Beginnings Standards



Note: +p<.10, *p<.05, **p<.01; Levels 1, 2, 3 significantly differ from Levels 4 & 5

Across the fall and the spring, there were also significant differences seen for children in proposed Level 4 and Level 5 programs on applied problems scores, which reflect math ability (Table 7-DD). Again, there were not significant changes in math skills across the school year, but children in the higher Levels made higher gains in skills on average.

Table 7-DD Children's Applied Problems Scores by Proposed Better Beginnings Standards

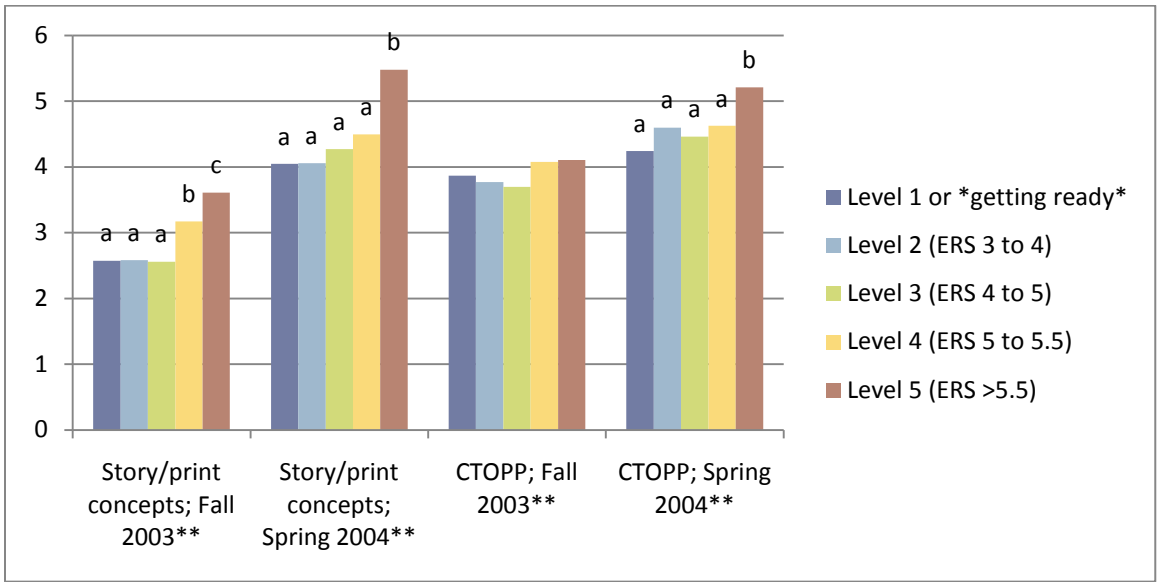


Note: +p<.10, *p<.05, **p<.01; in Fall 2003, Levels 1 significantly differ from Levels 4 & 5; in Spring 2004, Levels 1, 2, 3 significantly differs from Levels 4 and 5.

Findings from receptive vocabulary and applied problems demonstrate scores falling at or below one standard deviation lower than what is expected from national norms where children in higher quality programs are faring significantly better, although not quite approximating national norms. All of the children in the sample would be from low-income families. (Income at 100% of federal poverty is an

eligibility criterion for Head Start and Early Head Start.) Scores that are closer to approximating national norms are promising suggestions that quality programs can help reduce the poverty gap.

Table 7-EE Children's Story and Print Concepts Scores by Proposed Better Beginnings Standards



Note: +p<.10, *p<.05, **p<.01; Story/Print Concepts in Fall 2003, Levels 1 significantly differ from Levels 4 & 5; for Story/Print Concepts and Phonemic Processing in Spring 2004, Levels 1, 2, 3 significantly differs from Level 5.

There were also differences in children's knowledge of story and print concepts and phonemic processing in the fall and spring of the school year (Table 7-EE). We see significant gains in both areas where children in the lowest Level gained significantly fewer skills over the course of the school year than children in the higher quality programs.

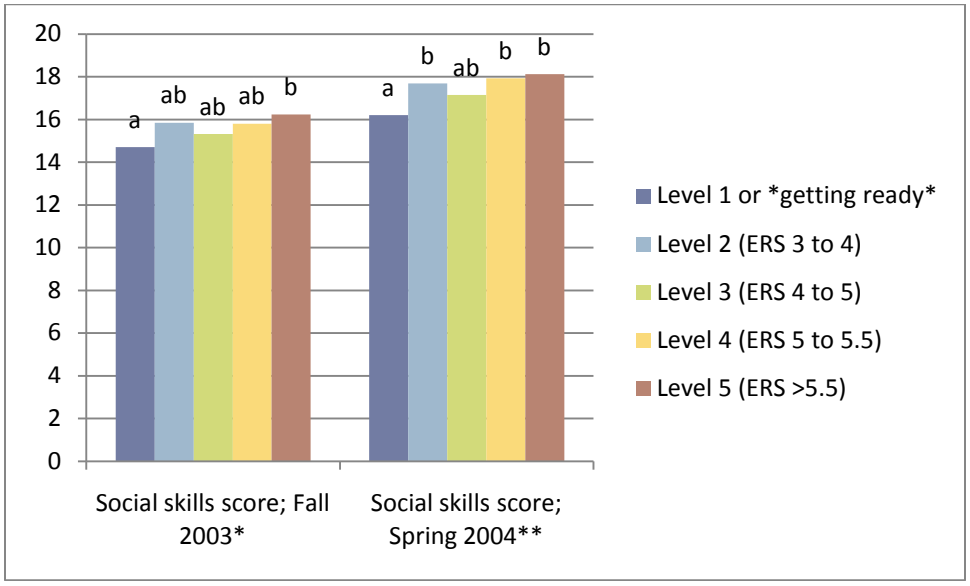
Table 7-FF Proposed Better Beginnings Levels (based on ECERS Scores in Head Start Programs)

Construct		Fall 2003			Spring 2004		
		N	Mean	SD	N	Mean	SD
Social skills score; Fall 2003* and Spring 2004**	Level 1 or *getting ready*	63	14.71	4.23	58	16.21	4.74
	Level 2 (ERS 3 to 4)	297	15.84	4.59	295	17.68	4.83
	Level 3 (ERS 4 to 5)	667	15.32	4.61	639	17.14	4.38
	Level 4 (ERS 5 to 5.5)	174	15.80	5.09	168	17.92	4.26
	Level 5 (ERS >5.5)	343	16.23	4.51	331	18.12	4.33
Aggressive behavior score; Fall 2003* and Spring 2004	Level 1 or *getting ready*	63	2.03	2.15	57	1.86	2.22
	Level 2 (ERS 3 to 4)	299	1.33	1.81	293	1.35	1.85
	Level 3 (ERS 4 to 5)	667	1.47	1.97	641	1.39	1.93
	Level 4 (ERS 5 to 5.5)	174	1.47	1.81	169	1.21	1.65
	Level 5 (ERS >5.5)	346	1.26	1.81	332	1.30	1.80
Behavior problems score; Fall 2003 and Spring 2004	Level 1 or *getting ready*	63	5.86	4.77	58	4.93	4.52
	Level 2 (ERS 3 to 4)	301	4.26	4.51	297	4.00	4.54
	Level 3 (ERS 4 to 5)	674	4.72	4.73	643	4.23	4.41
	Level 4 (ERS 5 to 5.5)	175	4.42	4.40	169	3.59	3.97
	Level 5 (ERS >5.5)	347	4.54	4.61	333	4.30	4.37

Note: ns=non-significant, +p<.10, *p<.05, **p<.01

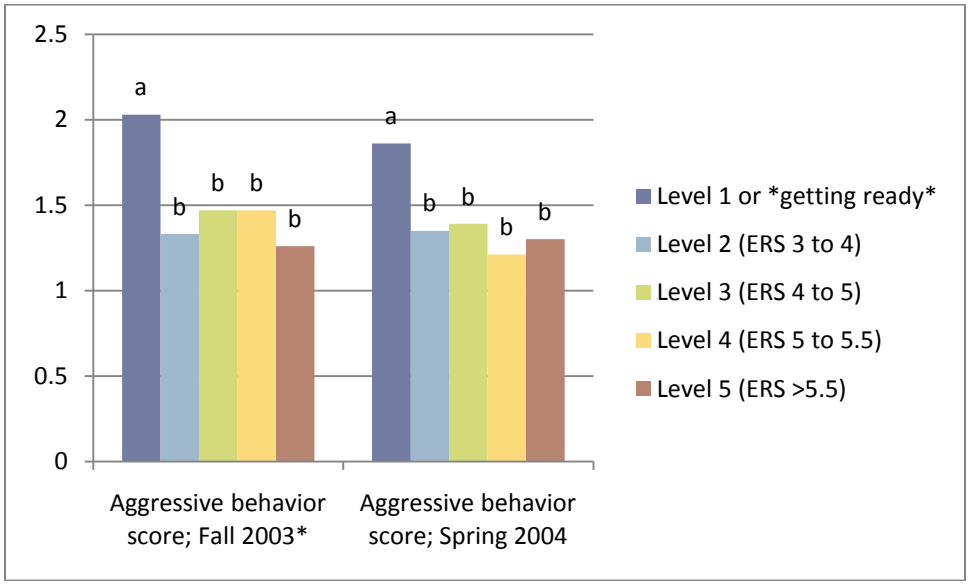
Children's social-emotional development was also assessed during the Head Start FACES study (Table 7-FF). For overall social skills, there were significant differences in Levels during fall and spring of the school year. Children in centers with the lowest quality (Level 1, "getting ready" whose scores fall below 3 on the ECERS-R) had the lowest reported level of social skills (Table 7-GG). There were no significant group differences in changes in social skills from fall to spring. Further, teacher-reported aggressive behaviors were significantly higher in fall and spring for children in Level 1 programs than for children in any other group (Table 7-HH). Overall, there were not significant differences in total behavior problems, although means were the highest across all groups among children in Level 1 programs.

Table 7-GG Children's Social Skills Scores by Proposed Better Beginnings Standards



Note: *p<.05, **p<.01; In Fall, Level 1 differs from 5; in Spring, Level 1 differs from 2, 4 & 5.

Table 7-HH Children's Aggression Scores by Proposed Better Beginnings Standards



Note: +p<.10, *p<.05, **p<.01; Levels 1 significantly differs from all other Levels.

7.9 CONCLUSIONS

Although our validation efforts are limited by the data that are publicly available, and there were many elements of Better Beginnings scoring that could not be replicated with existing data, we believe that there are some broad conclusions that can be drawn. It is clear that there are elements of the current standards that reflect lower levels of quality and may result in less optimal outcomes for children.

For Better Beginnings, cut scores on the Environmental Rating Scales (ERS) could be problematic. Findings from Infant-Toddler Environment Ratings Scale data collected as part of the national evaluation of Early Head Start (1 of 17 sites included children in Arkansas) show a loss of emotion regulation and engagement skills for very young children at the lowest ERS levels. Further, as children age, it was evident that programs scoring lower than 3 on the Early Childhood Environment Rating Scale (ECERS) resulted in children with less optimal language and math skills and socio-emotional development. These findings were mimicked with data from Early Head Start family child care programs. Children in family child care centers with scores lower than 3 had significantly lower cognitive, math, and language skills. The score ranges were the same as those seen in center-based early childhood programs. Most worrisome, children in the lowest quality centers, regardless of program type, scored more than one standard deviation below the national average in cognitive, math, and language scores.

Findings from analyses examining quality at higher levels of ECERS scores than are currently used in Better Beginnings showed group level differences that support the notion of encouraging programs to strive for the highest quality environments possible. Children in the highest Level programs (the proposed Levels 4 and 5 programs with respective cuts on ECERS scores of 5 and 5.5) showed higher cognitive and academic skill scores than children in lower Levels. Indeed, children in Level 5 programs fared even better than children in Level 4 programs in their awareness of print materials and phonemic knowledge. Analyses showed that children in the lowest quality centers fared the worst in cognitive, academic, and social skills.

There are many elements of Better Beginnings that the evaluation team was unable to articulate with existing data, and we believe further validation of the system with ongoing collection of child outcomes data is warranted. That said, analyses examining current cut scores on the Environmental Rating Scales for Infant/Toddler and Early Childhood programs would lead us to caution accrediting programs with a quality rating when at least minimum scores on Environmental Rating Scales (3 or higher) have not been met. We acknowledge that 1) centers should be given credit for volunteering to engage in quality improvement efforts, and 2) now that Better Beginnings has been implemented state-wide, changes to the qualifications for Levels could prove problematic. Therefore, we recommend Better Beginnings Level 1 be considered a "Getting Ready Level" to invite program participation, but also to communicate to parents that programs have not been assessed. We also recommend the development of Levels beyond the current highest Level of Better Beginnings to provide impetus for programs to continue to striving for program improvements that are demonstrated to promote the most optimal development for our children.

8 CHILD HEALTH & DEVELOPMENT

BB intends to improve child health and development through sharing information with parents—a structural measure—and documentation of the implementation of medical and educational plans, which is more proximal to the child. We divide this section into two discussions. The first conveys findings related to the topics of information to be shared with families. The second relates to standard 1.E.3, the documentation and implementation of medical and educational plans. Individualization of medical or educational care should be considered a part of basic care, but the subject is under-researched. The most relevant literature concentrates on whether the inclusion of children with disabilities or developmental delays into regular care is beneficial. Although our evaluation team did not find studies tracking the implementation of individualized plans for care in relation to child outcomes, this standard offers a pathway into a wider discussion of screening and assessment for individual needs and the ability of other measures within BB to address the quality of care provided for children with disabilities or delays.

8.1 SHARING INFORMATION

BB breaks out child health and development information shared with families across the three Levels:

- Level 1 programs share information on ARKids First (1.E.1) and on child development and health (1.E.2).
- Level 2 programs share information regarding medical homes (2.E.1) and on the stages of child development (2.E.2).
- Level 3 programs share information on nutrition and physical activity for children (3.E.1).

The rationale behind these standards is that the "child's early experiences set the stage for lifelong habits and behaviors. The combination of inadequate nutrition with limited physical activity has serious long-term consequences" (BB Guide), and increased parental knowledge of child development and health practices will lead to improved parent-child interactions and health routines in the home. Even when children are in pediatric care, children's developmental issues are often not adequately addressed by medical providers (Schuster, Duan, Regalado, & Klein, 2000). Because child care providers have daily contact with parents and children, they are in a unique position to fill this gap.

Items within BB that mandate sharing certain types of information with parents are structural measures. They do not address processes directly involving children in the center; rather they seek to enhance the conditions in which children are situated, in this case to enhance parenting and parental oversight of child health and development.

8.1.1 INFO ON HEALTH INSURANCE AND MEDICAL HOMES

ARKids First provides comprehensive medical coverage to children whose families might not be able to afford regular health insurance. Level 1 BB providers will help increase awareness of the program by documenting the distribution of ARKids First information to families of uninsured children (1.E.1). There is evidence that enrollment in State Children's Health Insurance Programs (CHIPS) increases use of the

medical home model and adherence to recommended schedules for immunizations and well-child visits (Kempe et al., 2000). In turn, parents have reported higher satisfaction and improved health, especially for children with severe conditions, when the medical home model is implemented (Palfrey et al., 2004). The medical home model provides children medical care that is accessible, continuous, comprehensive, coordinated, family-centered, compassionate, and culturally effective (Medical Home Initiatives for Children with Special Needs Project Advisory Committee, 2002). Use of the medical home reduces the need for emergency care that is more expensive, less consistent, and lower quality (Medical Home Initiatives for Children with Special Needs Project Advisory Committee, 2002; Starfield & Shi, 2004). We found no analysis on enrollment rates or child outcomes resulting from child care program distribution of information about or applications to enroll in State CHIPS.

8.1.2 INFO ON CHILD DEVELOPMENT AND HEALTH

Although there is some evidence to indicate children's home environments can be influenced by information parents obtain from the child care center (Powell, 1978) and that parenting practices may be improved as parents observe teacher-child interactions (Belsky, Steinberg, & Walker, 1982; C. Howes, 1987) positive and productive teacher-parent relationships do not come naturally or easily. Educators have little preparation in establishing partnerships with parents that provide opportunities to engage in productive discussions able to impact parenting practices, the home environment, and children's outcomes in meaningful ways (Knopf & Swick, 2007; Lightfoot, 2003). Some models of parent-teacher communication/education are in the process of being validated. They usually involve a comprehensive curriculum, training for teachers and recommended guidelines for when and how to present information to parents. Head Start is a leader in promoting parent-teacher communication and child medical care. Head Start programs offer a comprehensive and involved approach to parent education and arrangement for medical and developmental consultation and care. Evaluations show enrollment improves in participation in medical care and related health outcomes, but the effects of parent-teacher communication have not been isolated in the program's evaluations.

As previously discussed in the Strengthening Families section (4.3), brief parenting intervention models for the classroom, such as TIPS, appear to be a feasible way to enhance parent-teacher communication and to share parenting information (P. A. Bokony, McKelvey, Swindle, Patrick, & Shaw-Bailey, in preparation; McKelvey, et al., 2010). For more evidence, we turned to pediatrics literature associating sharing of information with improved parent and child outcomes. There we find that *anticipatory guidance*, the conveyance of information that doctors and nurses perceive may be important to the health of the child or of concern to the parent, can produce positive effects.

In two pediatric studies, mothers reported that anticipatory guidance increased their knowledge of developmental stages and the frequency of activities and sensitive, development-enhancing interactions with their children (Chamberlin & Szumowski, 1980; Chamberlin, Szumowski, & Zastowny, 1979). No significant differences in child development were found in families receiving more education during the visits. Casey and Whitt (1980) found infants to be more advanced in vocalization when mothers received anticipatory guidance, but no significant differences were detected for other measures.

BB distinguishes between information on child development and health on a lower Level and stages of development on a higher Level. Although the intent of the BB designers may have been to divide the workload for providers, we found one study suggesting this may not be a meaningful split in terms of outcomes. In comparing the effects of presenting information about nutrition, safety and sleep to parents

with and without explanation of affective, cognitive, and physical developmental stages, Dworkin and colleagues (1987) found that adding the developmental basis garnered no significant effects. One limitation of the literature on anticipatory guidance is that the content provided to parents varied (P Dworkin, 2000).

As implied by examples provided in the BB Toolkit, communication with parents is most likely to occur through print media, such as bulletin boards or handouts. In the medical field, adult patients are more satisfied with care when provided written information (Aizpuru, 1993), and we would anticipate the same is true for parents regarding their child care providers. Pediatric studies have demonstrated the ability of handouts to improve

- vaccination rates (Cates, 1990),
- compliance with treatment recommendations (Finney, Friman, Rapoff, & Christophersen, 1985),
- reduction of unnecessary telephone calls and the need for treatment visits (R. Casey et al., 1984; Long, Rickert, & Ashcraft, 1993; Roberts, Imrey, Turner, Hosokawa, & Alster, 1983), and
- parents' reported confidence in their knowledge of child development and greater willingness to ask their primary care provider questions after receipt of handouts addressing common parental concerns and providing recommendations for age-appropriate activities (Frankenburg & Thornton, 1989).

It stands to reason that talking to parents about the topic at hand in tandem with providing written information will produce greater results than providing written information alone. Several teams researching medical practices have confirmed that a dual approach optimizes outcomes (Li et al., 1984; McIntosh, Clark, & Howatt, 1994; Wall, Severson, Andrews, Lichtenstein, & Zoref, 1995).

8.1.3 INFO ON NUTRITION AND PHYSICAL ACTIVITY

Sharing nutritional information with families may be beneficial in a number of ways. Children need support from their parents to implement behavior changes, and changes last longer if interventions are aimed at family attitudes and habits rather than individuals. In a study of third-grade children who received either a school-based or home-based nutritional training program, students in the home-based program reported more behavior change, had reduced the amount of fat in their diets, and had more of the encouraged foods in their kitchens (C. L. Perry et al., 1988).

8.2 INDIVIDUALIZED PLANS

Apart from items contained in PAS and ECERS-R, which will be discussed here, the only item in BB to address individualized medical and educational planning and care for children is 1.E.3, "Any medical and educational care plans involving a child are written and on file, and implementation is documented while maintaining confidentiality." The promotion of high quality care for all children is crucial, but for children with developmental delays and disabilities high quality care is especially important. In addition to elements of environments that are considered necessary for promoting optimal outcomes in children with

typical developmental patterns, those with delays and disabilities need additional supports to achieve their optimal potential.

There are primarily two systems serving young children with delays and/or disabilities: 1) general ECE programs, and 2) programs that provide specialized services and supports for young children with disabilities. General ECE programs are those that are typically available within the community and include state and federally funded programs, such as Head Start. Head Start includes eligibility criteria for children with developmental delays and/or disabilities, and programs (including Early Head Start) must make 10% of their funded enrollment opportunities available for children with disabilities (U.S. Department of Health and Human Services, 1998).

Inclusion is a term used to describe ECE programs serving both children with developmental delays and/or disabilities and typically developing children. "Early childhood inclusion embodies the values, policies, and practices that support the right of every infant and young child and his or her family, regardless of ability, to participate in a broad range of activities and contexts as full members of families, communities and society" (DEC/NAEYC, 2009, p.1). The Individuals with Disabilities Education Act (IDEA), as amended in 2004, does not require inclusion but requires that children with disabilities be educated in the "least restrictive environment appropriate" to meet their needs.

There is empirical evidence to suggest that including children with delays/disabilities in general ECE programs may lead to more optimal outcomes than serving them in segregated programs. A review study reported that 11 of 16 programs that examined children's social development found more positive outcomes for children in integrated rather than segregated settings (V. Buysse & Bailey, 1993). Although findings were not always consistent, and there were threats to validity of the extant studies (children were not randomly assigned to setting), the authors concluded that inclusion in "integrated settings may be socially beneficial for some preschoolers with disabilities" (V. Buysse & Bailey, 1993, p. 457).

The second type of setting for children with delays and/or disabilities includes specialized services provided in a wide array of settings, including segregated care and education settings. In the state of Arkansas, there are multiple types of programs that provide specialized services across varied service-delivery models. In addition to special education services, programs include Rehabilitative Services for Persons with Mental Illness (RSPMI), Child Health Management Clinics (CHMS), and Developmental Day Treatment Clinic Services (DDTCS).

In Arkansas, Rehabilitative Services for Persons with Mental Illness (RSPMI) are often delivered in day-treatment and outpatient settings for children and families. Services typically include individual, family, group, and play therapy sessions, as well as psychological testing, diagnostic services, psychiatric evaluations, and speech therapy services. Child Health Management Clinics (CHMS) provide both developmental and medically focused treatment. Services are available in a day school setting and include physician and nursing services, physical therapy, occupational therapy, speech therapy, nutrition, early childhood developmental teaching, and psychological services. Finally, Developmental Day Treatment Clinic Services (DDTCS) provide clinic-based services to children with developmental disabilities. The core services include Early Intervention and/or preschool services for children birth to 5 years of age.

8.2.1 QUALITY ASSESSMENT FOR PROGRAMS SERVING CHILDREN WITH DISABILITIES

Spiker, Hebbeler, and Barton (in press) provide a detailed commentary of the current state of quality assessment for programs serving children with disabilities. The authors report that assessment of the quality of care for children with disabilities has mostly been conducted with the ERS assessment tools used in general education and care settings, and there are few studies on the quality of programs serving young children with disabilities.

There is one existing study examining child care quality specifically for children with developmental delays and/or disabilities receiving care in inclusive versus segregated care environments. LaParo, Sexton, and Snyder (1998) compared 29 segregated preschool special education classrooms and 29 inclusive child care programs. This study found similar ECERS scores in both settings. The average ECERS score of the segregated programs was 4.68 (range 3.19-5.59) and 4.77 in the inclusive programs (range 3.05-5.97).

There are also studies that examine the quality of care for children with disabilities compared to general ECE classrooms that do not serve children with special needs. One study compared quality of care in segregated early childhood special education classrooms serving children with developmental delays or disabilities to child care settings with typically developing children and reported lower ECERS scores for segregated settings (Bailey, Harms, & Clifford, 1982).

Other studies have examined general ECE settings, comparing those that are inclusive to those that are not inclusive. Across those studies higher quality has been found in settings that are inclusive of children with developmental delays and disabilities. For example, Buysse, Wesley, Bryant, and Gardner (1999) reported ANCOVA results predicting ECERS mean scores and demonstrated that program type (inclusive versus non-inclusive program) significantly predicted ECERS scores even when including teacher education, experience in the field, and knowledge of typical child development. Overall mean ECERS scores for the 62 inclusive programs was 4.44 (range 2.47-6.31) and for the 118 non-inclusive programs was 4.15 (range 2.81-5.97). Another recent investigation of the quality of care for young children also found significantly higher scores on ECERS-R for inclusive classrooms (M=4.74, SD=0.90) compared with classrooms without children with disabilities (M=3.93, SD=0.95), even after controlling for director and teacher education status, center accreditation, and services offered by the program (Grisham-Brown, Cox, Gravil, & Missall, 2010). Similar findings were also reported in infant-toddler care settings where inclusive classrooms were rated as having better overall quality using the ITERS-R scores than non-inclusive classrooms.

In each of the ERSs, there is measurement of care for children with developmental delays/disabilities. Included in the ITERS-R is an item (#32) that addresses programs' Provisions for Children with Disabilities. The item aims to measure a program's collaboration with other providers, adaptation of the physical environment and daily schedules, amount of caregiver interaction with the child(ren), and involvement with parents. The ECERS-R and FCCERS-R also include individual items that address provisions for children with disabilities. In the ECERS-R (Item # 37) and FCCERS-R (Item # 34), the same general categories of inclusion are considered, namely, collaboration with providers and implementation of intervention plans, modification of the environment (physical and non-physical), amount of teacher and peer interaction with the child(ren), and involvement with parents. There is a larger section of the SACERS that includes

supplemental items for children and youth with special needs. There are six items that measure provisions for exceptional children, individualization, multiple opportunities for learning and practicing special skills, engagement, peer interactions, and promoting communication. School-age programs in BB can also choose to be assessed using the Youth PQA (Adams, et al., 2005b), which does not include items specific to inclusion. There is an indicator of promoting psychological and emotional safety that measures evidence of bias and inclusion on the basis of "religion, ethnicity, class, gender, ability, appearance or sexual orientation," but no specific indicators of inclusion were found.

8.2.2 BB FOR CHILDREN WITH DELAYS/DISABILITIES

Arkansas includes children with delays and/or disabilities in that child screening and assessment is a component of each ERS (as described above) and of the PAS. The Child Assessment subscale of the PAS includes two related items: item 10, measuring screening and identification of special needs, and item 11, measuring assessments done in support of learning. Item 10 is specific to the identification of developmental delay and includes making referrals for children identified with needs. The scaling ranges from 1 (inadequate) where children are not screened, to 3 (minimal) screening is completed, to 5 (good) children are screened with a valid tool and referrals are made for further evaluation, to 7 (excellent) where safeguards are built into the identification process, parents are informed of implementation practices, and collaboration with specialists is supported. For school-age programs, the items from the PAS pertaining to developmental screening are excluded because designers assume that screening will take place in the context of the educational system. For children in family child care centers, item 8 on the Business Administration Scale (BAS) measures community resources that includes information about developmental screening. The scaling ranges from 1 (inadequate) where no community resources for parents are known, to 3 (minimal) where the provider has descriptive information about resources that include developmental screening services, to 5 (good) where the provider recommends screening to all parents, to 7 (excellent) where all children have a screening. Item 7 of the BAS measures provider-parent communication in which information about the child (including a developmental history and chronic medical conditions) is gathered, but there are not indices of inclusion, such as making modifications for children or implementing elements of care plans.

For young children in center-based care, Arkansas does not require screening as an element of care independent of the PAS assessment. Given BB ratings are based on a cut score for the average of the PAS items, it is possible for programs to score low on screening and identification (for failure to complete screenings or for not having referral plans for identified children) but still qualify for high ratings of quality. The Level 1 BB standard, "medical and educational care plans involving a child are written and on file, and implementation is documented" (1.E.3) would indicate that programs would need to adhere to an Individualized Family Service Plan (IFSP) or to an Individualized Education Program (IFSP) on record for an individual child. Overall, BB does not require screening outside of the context of the PAS assessment, but once a child is identified as needing specialized services, the system requires that programs implement existing individualized plans. Quality ECE should implement efforts to identify children with special needs and make referrals for early intervention. Without screening in ECE programs, delays and disabilities can stay unaddressed for years. As discussed previously, young children are more responsive to intervention than at any other time (Shonkoff & Phillips, 2000), and the earlier children are identified with delays, the greater the possibility that applied intervention will be effective. The results would be reduced special education costs, and ultimately, reduced hardship for children and their families.

8.3 KEY STATES COMPARISON

BB items related to health and development are not commonly shared with the key comparison states. We see more common ground in the area of sharing information with parents on child health and development, especially in family day care. Other items, though, are unique to Arkansas or only shared with Pennsylvania.

- **Documentation and implementation of medical and education plans**
  Ohio and Pennsylvania address children with special needs as part of their existing QRIS. Each has performance standards specifically for children with disabilities and their families. Ohio's Step Up To Quality Steps 2 and 3 require that all children receive developmental screenings within 60 days of enrollment into a program. Exemptions are made for children in school-age programs. SUTQ programs at Steps 2 and 3 are also required to make needed referrals within 90 days and to formally communicate results to parents. These requirements are necessary for both center-based and family child care settings.

  Pennsylvania's Keystone STARS lists standards for center-based care under the Community Resources/Family Involvement component. Star 2 programs are required to request a child's IEP/IFSP to inform classroom practices, while Star 4 programs must expand upon the Star 2 requirement by implementing "activities appropriate to meet IEP or IFSP goals and/or special needs plans and objectives." Family day care standards for children with special needs are found within Community Resources/Special Needs. Star 2 includes obtaining general information on any of the special needs issues of children in care. Star 3 includes gaining information from special needs assessments, following prescribed special needs treatments, and having a copy of a child's IEP or IFSP and a written plan for meeting needs. Star 4 includes "all staff having at least two hours of training in the last two years on inclusive early education and care practices, including how to access local community health and human services resources for families."

- **ARKids First**
  Pennsylvania requires 2-star family providers to give parents information about publicly funded health insurance programs. The language is less specific for centers, but 1-star providers must give families information about public, social, and community services.

- **Info on child health and development**
  Colorado and Pennsylvania have similar standards. Pennsylvania 1-star centers provide families with information about public, social and community services. For 2 stars, "At least once per year, written information on topics including health and human services, wellness, nutrition and fitness, and/or child development is given and explained to parents and staff." There is more overlap within family day care standards. Colorado, Missouri, Ohio, Oklahoma, and Pennsylvania include standards related to sharing information on health and/or development.

- **Medical homes**
  No key states match found.

- **Sharing info on nutrition and physical activity**
  In Pennsylvania, 2-star centers provide written information on topics including health and human services, wellness, nutrition and fitness, and/or child development is given and explained to

parents and staff at least once per year. Family day care providers share food safety and nutrition info with parents who bring their children's lunches from home.

8.4 CROSSWALKS

8.4.1 HEAD START

Head Start is known for its emphasis on and comprehensive support for child health and development, and standards exceed those in BB. The Head Start standards below address universal screening; subsequent referral and treatment for children who need medical or developmental care; and parental education, communication, and coordination.

- Staff members assist parents with enrolling and participating in ongoing medical care and provide opportunities to learn about preventive and emergency care (45 CFR 1304.40).
- Conferences and home visits address child development and the importance of physical activity in home-based settings (45 CFR 1304.21).
- Parents should be educated in child mental health issues, and programs should create opportunities to identify and discuss mental health issues with parents (45 CFR 1304.24, 45 CFR 1304.40).
- Health screening, treatment and procedures to track health care and the use of IEP/ISFP are required (45 CFR 1304.20).
- Staff and families work together to identify and fulfill each child's nutritional needs. Parent education must address food preparation and nutritional skills (45 CFR 1304.22). Plans for children with disabilities must address prevention of disabilities with a nutrition basis (45 CFR 1308.20).
- Centers provide indoor and outdoor space and equipment and adult guidance to promote active play and the development of gross and fine motor skills. Parents in home-based settings must be encouraged to appreciate the importance of physical development and to provide indoor and outdoor play. (CFR45 1304.21).

8.4.2 NAEYC

NAEYC emphasizes on-site approaches to support child health and development. The organization includes provisions that address health and safety practices that exceed typical minimum licensing, use of health consultants, and education and opportunities to enhance children’s onsite activity and nutrition. NAEYC does offer standards for communicating with parents about their child’s development (7.B.), but not to the degree of either Head Start. For accreditation programs should

- include specific contents in health records (5.A.01).
- coordinate with a health consultant to review program practices (5.A.02).

- have staff “support and encourage families' efforts to negotiate health, mental health, assessment, and educational services for their children (This criterion is an Emerging Practice.)" (7.C.05).
- communicate with families regarding developmental milestones or other issues related to mutual care on a daily basis for children under 2 or weekly for older children (7.B.05).
- allow infants and toddlers to move freely and have multiple opportunities to develop emerging physical skills (2.C.01). Children have opportunities and equipment for large motor experiences (2.C.04).
- provide daily outdoor play (5.A.06).
- implement curriculum that addresses health and safety on topics such as nutrition, medical and dental visits, taking medicine, etc. (2.K).
- assess child progress and implement guidelines for written educational plans (4).
- either participate in a USDA food program or follow evaluations provided by a registered dietitian or public health nutritionist (5.A.02.c).
- provide supports for breastfeeding (5.B.10).

8.4.3 NAFCC

In terms of health and development, NAFCC standards are the least rigorous of those compared. Other than updating medical records and following a received IFSP or IEP, there are no requirements for sharing of information with parents regarding child health and development. NAFCC does require family day care programs to

- compare immunization records to national standards (4.87) and to maintain updated medical information for each child (5.23).
- follow IFSP or IEP for children diagnosed with specific conditions (2.9).
- offer daily opportunities for large motor and small motor activities (3.52, 3.53).
- follow Child and Adult Care Food Program guidelines (4.73).
- share info with parents about common child-rearing issues such as temper tantrums or signs of infectious disease (5.10).
- help families access community and medical services (5.12).

8.4.4 COA

The COA ASP-PS Section 10 asserts that family connections increase a program’s ability to support children and youth, and subsection .03 focuses on well-being: “Personnel and family members share information to support the well-being of children and youth, and families are provided with information

about resources and services that can help meet their needs, when appropriate." Within those broad guidelines, after-school programs are required to show and document how they implement this standard. Nutritional and health needs are addressed by program staff (ASP-PS 8.01 and 8.02), and outdoor play is encouraged (ASP-PS 7). Programs are required to collect information about special needs, including medical needs (ASP-PS 2.03), to keep this information in the child/youth's file (ASP-AM 8.02), and to collaborate with school personnel to address the needs of children and youth (ASP-PA 11.04).

8.4.5 CARF

A majority of the programs that serve children with delays and/or disabilities in the state of Arkansas are accredited by the Commission on Accreditation of Rehabilitation Facilities (CARF). CARF's mission statement includes the promotion of "quality, value, and optimal outcomes of services through a consultative accreditation process that centers on enhancing the lives of the persons served" (p. 1). A CARF-accredited organization goes through a rigorous on-site assessment that includes interviews with management, staff and clients, as well as extensive records review. Child care and early education programs with CARF accreditation in Arkansas typically fall under one of four categories: 1) Behavioral Health Outpatient Treatment, 2) Day Treatment, 3) Child Care, or 4) Home and Community Based Rehabilitation.

The CARF philosophy of child and family-centered care includes: 1) facilitation of parent-professional collaboration at all levels of care, 2) sharing of unbiased and complete information about a child's/youth's care on an ongoing basis in an appropriate and supportive manner, 3) implementation of appropriate policies and programs that are comprehensive and provide necessary support to meet the needs of children/youth/families, 4) recognition of child/youth/family strengths and individuality and respect for different methods of coping, 5) understanding and incorporating the developmental needs of children/youth/families into service systems, and 6) assurance that the design of health and social service delivery systems is flexible, accessible, and responsive to the needs of children/youth/families. Programs with CARF accreditation would exceed the requirements of the Better Beginnings in this area.

8.5 CONCLUSIONS

The majority of the standards in Better Beginnings Health and Development involve sharing information with parents, which is a structural measure. Topics chosen are appropriate to the needs of children, and there is some empirical precedence from pediatric literature to suggest that anticipatory guidance, such as print information about child development or medical conditions, increases use of medical and preventive care as well as parental willingness to communicate with providers. While some studies found that written information produced results and patient satisfaction, others find that written guidance coupled with conversational guidance was more effective.

For young children in center-based care, Arkansas does not require screening as an element of care independent of the assessment of administrative practices (using the PAS and the BAS). There is a requirement at the lowest Level of Better Beginnings that "medical and educational care plans involving a child are written and on file, and implementation is documented" (1.E.3), which would indicate that

programs would need to adhere to an individualized plan (IFSP/IEP) for children with identified delays/disabilities.

We would recommend that if Better Beginnings adds higher Levels in the future, standards in the health and development category should observe processes that occur with the child during hours of care. We also believe that quality ECE should implement efforts to identify children with special needs and make referrals for early intervention. Without screening in ECE programs, delays and disabilities can stay unaddressed for years. Young children are more responsive to intervention than at any other time (Shonkoff et al., 2000). The earlier children are identified with delays, the greater the possibility that applied intervention will reduce special education costs and hardship for children and families.

9 LIMITATIONS OF RESEARCH

Our ability to draw firm conclusions about the effect of child care on child outcomes is compromised by a number of limitations within early childhood research. Within the studies we see:

- **Use of altered versions of the ERS or composite indices combining ECERS-R with other measures of global quality.** The fine print in measurement descriptions reveals that although the ERS instruments may be used, they are sometimes not applied as originally written and tested. We would like to say that as we compare ERS studies we are comparing apples to apples, but in fact, we may be comparing apples to crabapples.

- **Concurrent associations.** Some studies test children's development at the same time they measure quality. Positive associations between the two may not indicate that better quality caused better development, rather that children with better development are placed in better quality care (Mashburn, 2008). Our analyses using Head Start and Early Head Start data to glean the outcomes for children using BB guidelines avoid this weakness. The Early Head Start evaluation study randomized children into the program, and the Head Start FACES study included a representative sample of Head Starts, which would be assumed to be of equivalent quality.

- **Attrition and non-representative samples.** Non-respondents or dropouts on the child/family level are more likely to be from higher risk groups. (NICHD 1999, 2000, 2001a). Likewise, on the program level, lower quality providers are more likely to decline participation or to drop out of longitudinal studies than high-quality providers (G Zellman, et al., 2008). Less representative samples likely blunt detection of the effects of quality on child development. Highly regulated programs that have built in mandatory quality checks, such as Head Start, also provide a limited range of quality, typically in the good to excellent range. It is possible that if research was able to observe lower quality childcare, we would be able to pinpoint how elements of care affect children in various ways.

- **Probable omitted variables.** It is likely that most research does not fully capture and control for all of the parent, child, or even neighborhood variables that may affect the family's selection of a particular type of care and child outcomes. "Omitted variables can arise from either parent- or child-based characteristics and can impart either upward or downward bias to the estimated child care quality coefficients. Simple correlations between child care quality and child outcomes are not informative about the nature of 'true' effects, given the uncertain direction of bias, and attempts to control for some but not all of them will fail to provide policy makers with valid bottom-line impact estimates" (Duncan & Gibson-Davis, 2006, p. 615).

- **Correlation rather than cause.** Researchers can identify associations between one factor and another but are only rarely able to prove that factor X caused factor Y. It is possible that variables not included in the study and/or sample selection methods are responsible for differences in outcomes. Designers of recent studies include increasingly complex covariates and often collect baseline child performance against which to measure progress over the course of the study to uncover influences on child outcomes with more precision (D. L. Vandell, 2004). Nevertheless, it is still possible that they may not be nuanced enough to accurately capture the features that

have the most direct effects on quality.

- **Application of the ERS measures in settings or for purposes in which they have not been adequately tested.** While there is consensus that ERSs are helpful for self study and improvement within centers, scales have barely been tested in high-stakes settings, where results are used to publicly fund incentives for programs with higher scores or are used to sway parental choice based on assessment results, to merit their widespread use. Dickinson (2002) argues that ECERS-R has not kept up with recent research on child development illuminated by National Research Council (National Research Council, 2001). Despite literacy having become a national imperative, the only ECERS-R item addressing the practice of print is "Staff link spoken communication with written language," and it does not come into play until determining whether a program will rate the highest score of 7. If a state has established a goal of improving children's writing skills, unless ECERS-R is revised again, additional instruments that match this goal should be chosen.

10 STRENGTHS

BB includes some characteristics of an overall level of quality that have been found to contribute the most heavily to child outcomes. BB draws on the elements common to successful model preschool programs when deemed feasible for local providers. It includes standards for the education and training of staff, parent involvement, and use of a curriculum. It also employs global measures of program quality as rated by independent observers. Further, the component choices that were made in Better BB were similar to those of the other states we reviewed.

10.1 WILL FULFILL GOALS FOR GREATER PARTICIPATION

Whereas states with combined systems reap about 60% participation among state child care providers, states with voluntary systems, like BB, usually achieve 30% or less. Short of integrating licensing into the system, the components of BB were developed to promote quality as well as to encourage provider participation. In addition to BB, the state of Arkansas is investing training and technical assistance to help programs prepare and make improvements in quality and is providing grants to programs to help divert the costs of quality improvements and professional development needs.

10.2 INCORPORATES VALID AND RELIABLE MEASURES BY INDEPENDENT OBSERVERS

BB incorporates widely accepted, valid, and reliable instruments to assess quality. The use of the ERS suite of instruments, despite the potential issues raised above, provides BB the opportunity to compare quality across multiple studies and types of childhood programs. Of the 26 QRS/QRIS systems included in the *Compendium of Quality Rating Systems and Evaluations* (Kathryn Tout, et al., 2010), the majority use the ERS to assess the environment in child care centers (N=20) and family child care programs (N=17).

10.3 ADDRESSES MULTIPLE DOMAINS OF INFLUENCE

One of the most important decisions in QRIS design is component selection. A review of the five pioneering QRIS systems (also chosen as key comparison states in this study) found consensus concerning the quality components (G. Zellman & Perlman, 2008). Each system included measures of (1) staff training and education and (2) classroom or learning environment. In general, the component choices that were made in BB were similar to those of the other states we reviewed. BB also includes the use of valid and reliable measures of program quality that have been and are being used in other studies of quality in child care settings.

11 CONCERNS

"In most statewide or pilot QRSs, stakeholders describe their program as having an implicit logic model, rather than an explicit model. This means that stakeholders have discussed a common set of goals for the program and developed a set of activities to reach the goals. However, stakeholders have not articulated these activities and expected outcomes in a formal, written document that details the specific assumptions or pathways of change for their program" (K Tout, Zaslow, Halle, & Forry, 2009, p. 7).

Although the expressed goal of this particular investigation is better child outcomes, the implicit goal of BB is provider participation in program improvement. BB items do not include some characteristics or an overall level of quality that have been found to contribute the most heavily to child outcomes in recent years. We believe the goals for BB and the rationale for item selection is two-fold:

1. Designers selected items that local observers, technical assistants, and policy-makers have found to be lax in child care in Arkansas.

2. Designers wanted a system in which providers would be amenable to participation. Arkansas providers have expressed deep resistance to the idea of the rating system. Therefore, the designers have compromised with stakeholders, in particular with child care administrators concerned that BB standards will be far more demanding than current practice and that participation in the QRIS will require prohibitively high financial commitments at a time when many programs are already underfunded.

11.1 HEAVY ON STRUCTURAL MEASURES, LIGHT ON PROCESS MEASURES

As previously stated, research typically identifies an indirect relationship between structural features and child outcomes, with structural features impacting child outcomes by setting the stage for good processes to occur (Child Trends, 2009). Structural features of childhood programs are cheaper to implement and easier to observe and mandate, but there are drawbacks to their use. By not measuring the components of care that directly impact children's outcomes (such as classroom interactions), QRIS systems can ultimately increase costs in training and technical assistance without having influence on the real concern: children's development.

11.2 ALTERED USE OF MEASURES NOT TESTED FOR HIGH STAKES USAGE

BB excludes PAS items 22-25 that address administrator and teacher qualifications and PAS items 5 and 6 that rate staff benefits, staffing patterns, and scheduling. Evidence from analyses conducted by the evaluation team suggests that these exclusions may have negative impacts on the instrument's validity. It is unknown how the altering of the scoring of this measure will ultimately work in the state.

In addition to changes to the actual scoring of instruments, there will be alterations to the use of the ERS measures. Program assessment will be completed by trained and reliable data collectors; however, BB will not assess quality in every classroom, but in a random sampling of one-third of classrooms for participating providers. This is not an uncommon practice, but existing studies raise concerns about the accuracy of the prediction. For example, a study of QRS pilot studies in Missouri reported matching 75% of programs within their Tier system based on all classrooms versus one-third sampling distribution ERS scores (Thornburg, et al., in press). Further, a study of the Illinois QRS system reported a 67% match for

the same sampling distribution (McCormick Center for Early Childhood Leadership, 2010). Again, this is a common practice in QRS/QRIS systems across the nation, but there are also questions that remain about whether the financial savings of sampling fewer classrooms outweigh the benefits of better prediction of quality (K Tout, et al., 2009).

11.3 LEVELS DO NOT ADDRESS HIGHER LEVELS OF QUALITY

BB is similar to other comparison state rating systems in permitting self-assessment for entry-level quality ratings but diverges from other states by permitting scores in the adequate range for upper tiers of quality. Of the QRS/QRIS systems included in the *Compendium of Quality Rating Systems and Evaluations* (Kathryn Tout, et al., 2010) that included the ERS suite, the lowest average ERS score for center-based care typically recognized was either 3.0-3.75 (N=13) or 4.0-4.5 (N=6). The highest average ERS score recognized was typically in the 5.0 – 5.5 range (13) or higher (4). This was similar for family child care programs, with the lowest average ERS score recognized usually in the 3.0-3.5 range (8), or in the 4.0-4.5 range (7) and the highest average ERS score in the 5.0 – 5.5 range for most (12) programs. The highest range for BB is 4, which represents a significant divergence from what is seen in the majority of quality rating systems across the nation.

12 RECOMMENDATIONS

Within any system, there is the opportunity for improvement. The evaluation team recommends refinements and revisions that will reduce redundancy in the components and strengthen the influence of BB while being mindful that changes to a system that has been implemented state-wide should be minimized to prevent possible resentment (G. Zellman & Perlman, 2008).

12.1 REDUCE REDUNDANCY

There are many elements of family involvement documented in BB. There are items of the Strengthening Families (SF) training materials that are being assessed as part of the administrative and environmental assessments. We recommend that the SF component be modified to exclude content areas already gauged with the PAS and ERS assessments.

12.2 USE MEASURES AS WRITTEN AND TESTED

Better Beginnings should assess and score PAS items that are currently excluded. Teacher education measured by PAS is related to more optimal classroom practices. We recognize that providers may have difficulty achieving high scores on the items, but the original scaling of the instrument outperforms the scale with the excluded items. The state has already invested in the PAS and should take advantage of its validity and reliability testing.

12.3 DESIGNATE CAREGIVER-CHILD RATIOS

A key component present in other state rating systems but absent from Better Beginnings is required teacher-child ratios. States and organizations seeking to improve child outcomes via research-based practices have adopted guidelines for limiting the number of children in a teacher's care. Arkansas minimum licensing allows less optimal teacher-child ratios, especially for birth to two years, than most key comparison states, the National Association for the Education of Young Children (NAEYC), and Head Start. Improving ratios is expensive. Within Arkansas, stakeholders have rejected attempts to adjust minimum licensing regulations. Nevertheless, in light of consistent evidence that ratios affect child outcomes, we recommend that Better Beginnings include requirements for teacher-child ratios that exceed those found in minimum licensing.

12.4 INCORPORATE PROCESS MEASURES

The evaluation team recommends incorporating process measures because they are stronger predictors of child outcomes. There are structural measures in Better Beginnings that could be strengthened by altering how information is collected. One example that is strongly supported in the literature is use of a curriculum. There is evidence that relying on program-reported use of curricula will be less accurate than using independent observations of use. Developing methods to observe adherence to a curriculum during an assessment visit versus permitting self-report of curriculum is advised.

There is also strong evidence to support inclusion of teacher-child interactions assessments. The incorporation of a new instrument would be costly. As an alternative, Better Beginnings could more

closely track teacher-child interactions already being observed with the ERS instruments. The evaluation team found evidence that children in programs with ERS Interaction subscale scores that do not meet the minimum criterion score for the overall ERS have less optimal cognitive and social development. These findings could be used to support revised Better Beginnings standards requiring programs to meet an overall minimum ERS score and also the same minimum for the Interactions subscale. Increased technical support for programs in the area of teacher-child interactions is warranted when scores on the Interaction subscale of the ECERS-R are substantially lower than the overall score.

12.5 ADDRESS LOWER LEVELS OF QUALITY

Analyses examining current cut scores on the ERS would lead us to caution accrediting programs with a quality rating when at least a minimum score of 3 has not been met. An implicit goal of Better Beginnings is to communicate to parents the importance of quality child care for their children's development. We recommend that Better Beginnings Level 1 be considered a "getting ready" level that invites participation but also communicates to parents that programs have not yet been assessed and may not reflect a minimal level of environmental quality.

12.6 ADDRESS HIGHER LEVELS OF QUALITY

We recommend the development of levels beyond the current highest level of Better Beginnings to encourage programs to make improvements that promote optimal child development. The range of low scores typically recognized in other states' QRIS is either 3.0-3.75 (N=13) or 4.0-4.5 (N=6). The range of high scores is typically in the 5.0 – 5.5 range (N=13) or higher (N=4). [45] Better Beginnings Level 3, the highest rating in the system, requires average ERS scores of 4, a substantial divergence from other quality systems. Analyses showed that children in higher quality programs (meeting cut scores of 5 and 5.5 that the UAMS evaluation team proposes for future Better Beginnings Levels 4 and 5) had higher cognitive and academic skill scores than children in lower levels. In awareness of print materials and phonemic knowledge, children in our proposed Level 5 fared even better than children in our proposed Level 4 programs.

12.7 INCLUDE CHILD SCREENING

At the lowest level of Better Beginnings programs must adhere to an individualized plan (IFSP/IEP) for children with identified delays/disabilities. However, Arkansas does not require screening as an element of care independent of the assessment of administrative practices. (In PAS and BAS, programs with good to excellent scores facilitate screenings for children in their care). We recommend that quality ECE programs implement efforts to identify children with special needs and make referrals for early intervention. Without screening, delays and disabilities can stay unaddressed for years. Furthermore, young children are more responsive to intervention than at any other time. Early identification increases the possibility that applied intervention will be effective, reduces education costs, and alleviates hardship for children and their families.

★★★★★★★

[45] Tout, K., R. Starr, et al. (2010). Compendium of Quality Rating Systems and Evaluations. Child Trends, Mathematica Policy Research.

13 REFERENCES

Abbott-Shim, M. (2000). Sure Start effectiveness study: Final report. Atlanta, GA: Quality Assist, Inc.

Achenbach, T. M. (1992). *Teacher/caregiver report form for ages 2-5*. Burlington, VT: Center for Children, Youth, and Families; University of Vermont.

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms and profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families.

Action for Healthy Kids. (2008). Annual report Retrieved 6/3/2010, from http://www.actionforhealthykids.org/resources/files/annualreport.pdf

Adams, K., Brickman, N., & McMahon, T. (Eds.). (2005a). *Younger youth program quality assessment, Form A*. Ypsilanti, MI: High/Scope Press.

Adams, K., Brickman, N., & McMahon, T. (Eds.). (2005b). *Youth program quality assessment, Form A*. Ypsilanti, MI: High/Scope Press.

Adams, K., Brickman, N., & McMahon, T. (Eds.). (2005c). *Youth program quality assessment, Form B*. Ypsilanti, MI: High/Scope Press.

Afterschool Investments Project. (2007). Using a state child care quality rating system to promote quality in afterschool programs Retrieved 6/17/2010, from http://nccic.acf.hhs.gov/afterschool/qrs_afterschool.pdf

Aizpuru, B. (1993). Barriers to verbal communication and consumer satisfaction with consultations in general medicine. *Graceta Sanitaria, 7*(34), 27-31.

Alexander, K. L., & Entwisle, D. R. (1988). Achievement in the first two years of school: Patterns and processes. *Monographs of the Society for Research in Child Development, 53*.

Alkon, A., Ramley, M., & MacLennan, K. (2003). Evaluation of mental health consultation in child care centers. *Early Childhood Education Journal, 31*(2), 91-99.

Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology, 10*(4), 541-552.

Bailey, D. B., Harms, T., & Clifford, R. M. (1982). Comparison of preschool environments for handicapped and non-handicapped children. *Topics for Early Childhood Special Education, 2*(1), 9-20.

Barnard, W., Smith, W. E., Fiene, R., & Swanson, K. (2006). Evaluation of Pennsylvania's Keystone STARS Quality Rating System in child care settings. Pittsburgh, PA: University of Pittsburgh Office of Child Development and the Pennsylvania State University Prevention Research Center.

Barnett, W. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *The Future of Children, 5*(3), 25-50.

Barnett, W. (2003). Better teachers, better preschools: Student achievement linked to teacher qualifications. *Preschool Policy Matters, 2*, 1-12.

Bayley, N. (1993). *Bayley scales of infant development, second edition, manual*. New York, NY: The Psychological Corporation, Harcourt Brace & Company.

Beckett, M., Hawken, A., & Jacknowitz, A. (2001). Accountability for after-school care: Devising standards and measuring adherence to them. Los Angeles, CA: RAND Corporation.

Beller, E., Stahnke, M., Butz, P., Stahl, W., & Wessels, H. (1996). Two measures of the quality of group care for infants and toddlers. *European Journal of Psychology of Education, 11*(2), 151-167.

Belsky, J., Steinberg, L., & Walker, A. (1982). The ecology of day care. In M. Lamb (Ed.), *Childrearing in nontraditional families* (pp. 71-116). Hillsdale, NJ: Lawrence Erlbaum.

Bigras, N., Bouchard, C., Cantin, G., Brunson, L., Coutu, S., Lemay, L., et al. (2010). A comparative study of structural and process quality in center-based and family-based child care services. *Child & Youth Care Forum, 39*(3), 129-150.

Birmingham, J., Pechman, E. M., Russell, C. A., & Mielke, M. (2005). Shared features of high-performing after-school programs: A follow-up to the TASC evaluation Retrieved 06/02/2010, from http://www.sedl.org/pubs/fam107/fam107.pdf

Black, J., Jones, T., Nelson, C., & Greenough, W. (1998). Neuronal plasticity and the developing brain. *Handbook of child and adolescent psychiatry, 6*, 31–53.

Blau, D. M. (2000). The production of quality in child-care centers: Another look. *Applied Developmental Science, 4*(3), 136.

Bloom, P. (1996). *Improving the quality of work life in the early childhood setting: Resource guide and technical manual for the Early Childhood Work Environment Survey*. Wheeling, IL: McCormick Tribune Center for Early Childhood Leadership, National Louis University.

Bodilly, S., & Beckett, M. K. (2005). Making out-of-school time matter: Evidence for an action agenda, 06/24/2010, from http://www.rand.org/pubs/monographs/2005/RAND_MG242.pdf

Bokony, P., McKelvey, L., & Patrick, T. (2009). OFA promoting responsible fatherhood demonstration program semi-annual progress report. Little Rock, AR: University of Arkansas for Medical Sciences.

Bokony, P. A., McKelvey, L., Swindle, T., Patrick, T., & Shaw-Bailey, D. (in preparation). Feasibility of brief parenting interventions in early childhood education.

Boreham, C., & Riddoch, C. (2001). The physical activity, fitness and health of children. *Journal of sports sciences, 19*(12), 915-929.

Brandt, K. (2005). Strengthening Families: The impact of New Hampshire family support programs. Concord, NH: New Hampshire Children's Trust Fund.

Brennan, E. M., Bradley, J. R., Allen, M. D., & Perry, D. F. (2008). The evidence base for mental health consultation in early childhood settings: Research synthesis addressing staff and program outcomes. *Early Education & Development, 19*(6), 982 - 1022.

Bronfenbrenner, U. (1974). Is early intervention effective? A report on the longitudinal evaluations of preschool programs (Vol. 2). Washington, DC: Department of Health, Education, and Welfare, Office of Child Development.

Bronson, M., Tivnan, T., & Seppanen, P. (1996). Relations between teacher and classroom activity variables and the classroom behaviors of prekindergarten children in Chapter 1 funded programs. *Journal of Applied Developmental Psychology, 16*, 253-282.

Bryant, D. M., Burchinal, M., Lau, B., & Sparling, J. (1994). Family and classroom correlates of Head Start children's developmental outcomes. *Early Childhood Research Quarterly, 9*, 289-309.

Bryant, D. M., Maxwell, K., Taylor, K., Poe, M., Peisner-Feinberg, E. S., & Bernier, K. (2003). Smart Start and preschool child care quality in North Carolina: Change over time and relation to children's readiness. Chapel Hill, NC: Frank Porter Graham Child Development Center; FPG Child Development Institute

Bryant, D. M., Maxwell, K. L., & Burchinal, M. (1999). Effects of a community initiative on the quality of child care. *Early Childhood Research Quarterly, 14*(4), 449-464.

Burchinal, M., Howes, C., & Kontos, S. (2002). Structural predictors of child care quality in child care homes. *Early Childhood Research Quarterly, 17*(1), 87-105.

Burchinal, M., Kainz, K., & Cai, K. (in press). How well are our measures of quality predicting to child outcomes: A meta-analysis and coordinated analysis of data from large-scale studies of early childhood settings. In M. Zaslow, K. Tout, T. Halle & I. Martinez-Beck (Eds.), *Next steps in the measurement of quality in early childhood settings*. Baltimore: Brookes Publishing.

Burchinal, M., Roberts, J., Nabors, L., & Bryant, D. (1996). Quality of center child care and infant cognitive and language development. *Child Development*, 606-620.

Burchinal, M., Vandergrift, N., Pianta, R. C., & Mashburn, A. J. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly, 25*(2), 166-176.

Burchinal, M. R., Cryer, D., Clifford, R. M., & Howes, C. (2002). Caregiver training and classroom quality in child care centers. *Applied Developmental Science, 6*(1), 2-11.

Burchinal, M. R., Peisner-Feinberg, E., Bryant, D. M., & Clifford, R. (2000). Children's social and cognitive development and child-care quality: Testing for differential associations related to poverty, gender, or ethnicity. *Applied Developmental Science, 4*(3), 149.

Burchinal, M. R., & Roberts, J. E. (2000). Relating quality of center-based child care to early cognitive and language development longitudinally. [Article]. *Child Development, 71*(2), 339.

Burdette, H. L., & Whitaker, R. C. (2005). Resurrecting free play in young children: Looking beyond fitness and fatness to attention, affiliation, and affect. *Archives of Pediatrics & Adolescent Medicine, 159*(1), 46-50.

Buysse, V., & Bailey, D. B. (1993). Behavioral and developmental outcomes in young children with disabilities in integrated and segregated settings: A review of comparative studies. *Journal of Special Education, 26*(4), 434-461.

Buysse, V., Wesley, P., Bryant, D., & Gardner, D. (1999). Quality of early childhood programs in inclusive and noninclusive settings. *Exceptional Children, 65*(3), 301-314.

Caldwell, B. M., & Bradley, R. H. (1984). *Home observation for measurement of the environment*. Little Rock, AR: University of Arkansas at Little Rock.

Campbell, F. A., Ramey, C. T., Pungello, E., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian project. *Applied Developmental Science, 6*(1), 42-57.

Casey, P. H., & Whitt, J. K. (1980). Effect of the pediatrician on the mother-infant relationship. *Pediatrics, 65*(4), 815.

Casey, R., McMahon, F., McCormick, M., Pasquariello Jr, P., Zavod, W., & King Jr, F. (1984). Fever therapy: An educational intervention for parents. *Pediatrics, 73*(5), 600.

Cassidy, D., Hestenes, L., Mims, S., & Hestenes, S. (2003). North Carolina rated license assessment project: A three-year summary of assessed facilities: An executive summary 1999-2002. Greensboro, NC: NCRLAP; UNCG.

Cassidy, D. I., Buell, M. I., Pugh-Hoese, S., & Russell, S. (1995). The effect of education on child care teachers' beliefs and classroom quality: Year one evaluation of the teach early childhood associate degree scholarship program. *Early Childhood Research Quarterly, 10*(2), 171-183.

Center for the Study of Social Policy. (2008a). State evaluation strategies Retrieved 05/12, 2010, from http://www.strengtheningfamilies.net/images/uploads/state_tools/State_Evaluation_Strategies.pdf

Center for the Study of Social Policy. (2008b). Strengthening Families online resources: Guidebook and self-assessment Retrieved 05/11/2010, from http://www.strengtheningfamilies.net/index.php/online_resources/guide_assess/category/self_assessment/

Centers for Disease Control and Prevention. (1997). Guidelines for school and community programs to promote lifelong physical activity among young people. *Morbidity and Mortality Weekly Report, 46*(RR-6).

Centers for Disease Control and Prevention. (2003). First reports evaluating the effectiveness of strategies for preventing violence: Early childhood home visitation *Morbidity and Mortality Weekly Report* (pp. 1-9).

Chaffin, M. (2004). Is it time to rethink Healthy Start/Healthy Families? *Child Abuse & Neglect, 28*(6), 589-595.

Chamberlin, R., & Szumowski, E. (1980). A follow-up study of parent education in pediatric office practices: Impact at age two and a half. *American Journal of Public Health, 70*(11), 1180.

Chamberlin, R., Szumowski, E., & Zastowny, T. (1979). An evaluation of efforts to educate mothers about child development in pediatric office practices. *American Journal of Public Health, 69*(9), 875.

Chambers, B., Cheung, A., & Slavin, R. (2006). Effective preschool programs for children at risk of school failure: A best-evidence synthesis. In B. Spodek & O. Saracho (Eds.), *Handbook of research on the education of young children* (2nd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Child Trends. (2009). What we know and don't know about measuring quality in early childhood and school-age care and education settings (Vol. OPRE Issue Brief #1 Publication #2009-12). Washington, DC: Child Trends.

Chin-Quee, D. S., & Scarr, S. (1994). Lack of early child care effects on school-age children's social competence and academic achievement. *Early Development and Parenting, 3*(2), 103-112.

Clarke-Stewart, K. A., Vandell, D. L., Burchinal, M., O'Brien, M., & McCartney, K. (2002). Do regulable features of child-care homes affect children's development? *Early Childhood Research Quarterly* (Vol. 17, pp. 52-86).

Clifford, R., Bryant, D., & Early, D. M. (2005). What we know about pre-kindergarten programs. *Principal, 85*(1), 20-24. Retrieved from http://www.naesp.org/resources/2/Principal/2005/S-Op20.pdf

Clifford, R. M., Harms, T., & Harms, T. (1989). *Family day care rating scale*. New York: Teachers College Press.

Cochran, M. (1992). Parent empowerment: Developing a conceptual framework. *Family Science Review, 5*(1), 3–21.

Cochran, M. (2007). *Finding our way: The future of American early care and education*. Washington, DC: ZERO TO THREE National Center for Infants, Toddlers and Families.

Comer, J. P. (1984). Home-school relationships as they affect the academic success of children. *Education and urban society, 16*(3), 323-337.

Committee on After-School Research and Practice. (2005). Moving towards success: Framework for after-school programs. Washington, DC: C. S. Mott Foundation.

Conners-Burrow, N. A., McKelvey, L. M., Amini-Vermani, E., & Sockwell, L. (In review). Evaluation results from the Arkansas Early Childhood Mental Health Consultation Project. *Infant Mental Health*.

Copple, C., & Bredekamp, S. (Eds.). (2009). *Developmentally appropriate practice* (Third ed.). Washington, DC: National Association for the Education of Young Children.

Daro, D., & Donnelly, A. C. (2002). Charting the waves of prevention: Two steps forward, one step back. *Child Abuse & Neglect, 26*(6-7), 731-742.

de Kruif, R. E. L., McWilliam, R. A., Ridley, S. M., & Wakely, M. B. (2000). Classification of teachers' interaction behaviors in early childhood classrooms. *Early Childhood Research Quarterly, 15*(2), 247-268.

Dearing, E., McCartney, K., & Taylor, B. A. (2009). Does higher quality early child care promote low-income children's math and reading achievement in middle childhood? *Child Development, 80*(5), 1329-1349.

Deater-Deckard, K., Pinkerton, R., & Scarr, S. (1996). Child care quality and children's behavioral adjustment: A four-year longitudinal study. *Journal of Child Psychology and Psychiatry, 37*(8), 937-948.

DEC/NAEYC. (2009). Early childhood inclusion: A summary. Chapel Hill, NC: The University of North Carolina, FPG Child Development Institute.

Dettling, A. C., Parker, S. W., Lane, S., Sebanc, A., & Gunnar, M. R. (2000). Quality of care and temperament determine changes in cortisol concentrations over the day for young children in childcare. *Psychoneuroendocrinology, 25*(8), 819-836.

Dickinson, D. (2002). Shifting images of developmentally appropriate practice as seen through different lenses. *Educational Researcher, 31*(1), 26.

Doherty, G., Forer, B., Lero, D. S., Goelman, H., & LaGrange, A. (2006). Predictors of quality in family child care. *Early Childhood Research Quarterly, 21*(3), 296-312.

Dowda, M., Brown, W. H., McIver, K. L., Pfeiffer, K. A., O'Neill, J. R., Addy, C. L., et al. (2009). Policies and characteristics of the preschool environment and physical activity of young children. *Pediatrics, 123*(2), e261-266.

Dowda, M., Pate, R. R., Trost, S. G., Almeida, M. J., & Sirard, J. R. (2004). Influences of preschool policies and practices on children's physical activity. *Journal of Community Health, 29*(3), 183-196.

Dowsett, C. J., Huston, A. C., Imes, A. E., & Gennetian, L. (2008). Structural and process features in three types of child care for children from high and low income families. *Early Childhood Research Quarterly, 23*(1), 69-93.

Drummond, R. L., Staten, L. K., Sanford, M. R., Davidson, C. L., Magda Ciocazan, M., Khor, K.-N., et al. (2009). Steps to a healthier Arizona. *Health Promotion Practice, 10*(2 suppl), 156S-167S.

Duggan, A., McFarlane, E., Fuddy, L., Burrell, L., Higman, S. M., Windham, A., et al. (2004). Randomized trial of a statewide home visiting program: Impact in preventing child abuse and neglect. *Child Abuse & Neglect, 28*(6), 597-622.

Duncan, G. J., & Gibson-Davis, C. M. (2006). Connecting child care quality to child outcomes: Drawing policy lessons from nonexperimental data. *Evaluation Review, 30*(5), 611-630.

Dunn, C., Thomas, C., Ward, D., Pegram, L., Webber, K., & Cullitan, C. (2006). Design and implementation of a nutrition and physical activity curriculum for child care settings. *Preventing chronic disease, 3*(2), A58.

Dunn, L., Beach, S. A., & Kontos, S. (1994). Quality of the literacy environment in day care and children's development. *Journal of Research in Childhood Education, 9*(1), 24 - 34.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary test-third edition*. Circle Pines, MN: American Guidance Service.

Dunst, C., & Trivette, C. (2005). *Measuring and evaluating family support program quality*. Asheville, NC: Winterberry Press.

Dunst, C. J., Trivette, C. M., & Hamby, D. W. (2006). *Family support program quality and parent, family and child benefits*. Asheville, NC: Winterberry Press.

Dunst, C. J., Trivette, C. M., & Hamby, D. W. (2007). Meta-analysis of family-centered helpgiving practices research. *Mental Retardation & Developmental Disabilities Research Reviews, 13*(4), 370-378.

Duran, F., Hepburn, K., Irvine, M., Kaufmann, R., Horen, N., & Perry, D. (2009). What works?: A study of effective early childhood mental health consultation programs. Washington, DC: Georgetown University Center for Child and Human Development.

Durlak, J. A., & Weissberg, R. P. (2007). The impact of after-school programs that promote personal and social skills. Chicago, IL: Collaborative for Academic, Social and Emotional Learning.

Dworkin, P. (2000). Preventive health care and anticipatory guidance. In J. P. Shonkoff & S. J. Meisels (Eds.), *Handbook of early childhood intervention* (2nd ed., pp. 327-338). Cambridge, UK: Cambridge University Press.

Dworkin, P., Allen, D., Geertsma, M., Solkoske, L., & Cullina, J. (1987). Does developmental content influence the effectiveness of anticipatory guidance? *Pediatrics, 80*(2), 196.

Early, D. M., Bryant, D. M., Pianta, R. C., Clifford, R. M., Burchinal, M. R., Ritchie, S., et al. (2006). Are teachers' education, major, and credentials related to classroom quality and children's academic gains in pre-kindergarten? *Early Childhood Research Quarterly, 21*(2), 174-195.

Early, D. M., Maxwell, K. L., Burchinal, M., Alva, S., Bender, R. H., Bryant, D., et al. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development, 78*(2), 558-580.

Elicker, J., Fortner-Wood, C., & Noppe, I. C. (1999). The context of infant attachment in family child care. *Journal of Applied Developmental Psychology, 20*(2), 319-336.

Elliot, S. N., Gresham, F. M., Freeman, R., & McCloskey, G. (1988). Teacher and observer ratings of children's social skills: Validation of the social skills rating scales. *Journal of Psychoeducational Assessment, 6*, 152-161.

Epstein, A. S. (1993). *Training for quality: Improving early childhood programs through systematic inservice training. Number nine, monograph series*. Ypsilanti, MI High/Scope Educational Foundation.

Epstein, J. L. (1996). Perspectives and previews on research and policy for school, family and community partnerships. In A. Booth & J. F. Dunn (Eds.), *Family-school links: How do they affect educational outcomes?* (pp. 209-246). Mahwah, NJ: Lawrence Earlbaum Associates, Inc.

Field, T. M. (1980). Preschool play: Effects of teacher/child ratios and organization of classroom space. *Child Study Journal, 10*(3), 191-205.

Fiene, R. (2002). Improving child care quality through an infant caregiver mentoring project. *Child & Youth Care Forum, 31*(2), 79-87.

Fiene, R., Greenberg, M. T., Bergsten, M., Fegley, C., Carl, B., & Gibbons, E. (2002). The Pennsylvania early childhood quality settings study: Executive summary. Harrisburg, PA: Governor's Task Force on Early Childhood Education; Pennsylvania State University. Prevention Research Center

Fiester, L., White, R. N., Reisner, E. R., & Castle, A. M. (2001). Evaluation results from the TASC after-school program's second year: Summary of findings. Washington, DC: Policy Studies Associates.

Finney, J. W., Friman, P. C., Rapoff, M. A., & Christophersen, E. R. (1985). Improving compliance with antibiotic regimens for otitis media: Randomized clinical trial in a pediatric clinic. *American Journal of Diseases of Children, 139*(1), 89-95.

Fischer, J. L., & Krause Eheart, B. (1991). Family day care: A theoretical basis for improving quality. *Early Childhood Research Quarterly, 6*(4), 549-563.

Fordham, I., Boyd, P., & Apicella, T. (2004). Building quality in out-of-school time in the United Kingdom. *The Evaluation Exchange, 9 X*(1 Spring).

Frankenburg, W. K., & Thornton, S. M. (1989). Denver developmental activities: A preliminary report. *Journal of Developmental & Behavioral Pediatrics, 10*(1), 22-26.

Fraser, J. A., Armstrong, K. L., Morris, J. P., & Dadds, M. R. (2000). Home visiting intervention for vulnerable families with newborns: Follow-up results of a randomized controlled trial. *Child Abuse & Neglect, 24*(11), 1399-1429.

Frede, E. (1995). The role of program quality in producing early childhood program benefits. *The Future of Children, 5*(3), 115-132.

Fries, A. B. W., & Pollak, S. D. (2004). Emotion understanding in postinstitutionalized Eastern European children. *Development and Psychopathology, 16*(02), 355-369.

Fukkink, R. G., & Lont, A. (2007). Does training matter? A meta-analysis and review of caregiver training studies. *Early Childhood Research Quarterly, 22*(3), 294-311.

Fuller, B., Kagan, S. L., Loeb, S., & Chang, Y. (2004). Child care quality: Centers and home settings that serve poor families. *Early Childhood Research Quarterly, 19*(4), 505-527.

Fuller, B., Livas, A., & Bridges, M. (2006). How to expand and improve preschool in california: Ideals, evidence, and policy options. Working paper 05-1. Berkeley, CA: Policy Analysis for California Education.

Galinsky, E., Howes, Kontos, S., & Shinn, M. (1994). *The study of children in family child care and relative care. Highlights of findings*. New York, NY: Families and Work Institute.

Galinsky, E., Howes, C., & Kontos, S. (1995). The family child care training study. Highlights of findings. New York, NY: Families and Work Institute.

Garces, E., Thomas, D., & Currie, J. (2002). Longer term effects of Head Start. *The American Economic Review, 92*(4), 999-1012.

Garet, M., Porter, A., Desimone, L., Birman, B., & Yoon, K. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*(4), 915.

Gerber, E. B., Whitebook, M., & Weinstein, R. S. (2007). At the heart of child care: Predictors of teacher sensitivity in center-based child care. *Early Childhood Research Quarterly, 22*(3), 327-346.

Glewwe, P., Jacoby, H. G., & King, E. M. (2001). Early childhood nutrition and academic achievement: A longitudinal analysis. *Journal of Public Economics, 81*(3), 345-368.

Goelman, H., & Pence, A. R. (1988). Children in three types of day care: Daily experiences, quality of care and developmental outcomes. *Early Child Development and Care, 33*(1), 67 - 76.

Goffin, S. (2000). The role of curriculum models in early childhood education. *ERIC Digest, 443*, 597.

Goldsmith, J., Arbreton, A. J. A., & Bradshaw, M. (2004). Promoting emotional and behavioral health in preteens: Benchmarks of success and challenges among programs in Santa Clara and San Mateo Counties. Oakland, CA: Public/Private Ventures.

Gomby, D., Culross, P., & Behrman, R. (1999). Home visiting: Recent program evaluations: Analysis and recommendations. *The Future of Children, 9*(1), 4-26.

Green, B. L., Everhart, M., Gordon, L., & Garcia Gettman, M. (2006). Characteristics of effective mental health consultation in early childhood settings. *Topics in Early Childhood Special Education, 26*(3), 142-152.

Grisham-Brown, J., Cox, M., Gravil, M., & Missall, K. (2010). Differences in child care quality for children with and without disabilities. *Early Education & Development, 21*(1), 21-37.

Gullo, D. F. (2006). Alternative means of assessing children's learning in early childhood classroom. In B. Spodek & O. Saracho (Eds.), *Handbook of research on the education of young children* (2nd ed., pp. 443-455). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Hagekull, B., & Bohlin, G. (1995). Day care quality, family and child characteristics and socioemotional development. *Early Childhood Research Quarterly, 10*(4), 505-526.

Hamre, B. K., & Pianta, R. C. (2004). Self-reported depression in nonfamilial caregivers: Prevalence and associations with caregiver behavior in child-care settings. *Early Childhood Research Quarterly, 19*(2), 297-318.

Harms, T., Clifford, R., & Cryer, D. (1998). *Early childhood environment rating scale, revised edition*. New York: Teachers College Press.

Harms, T., Cryer, D., & Clifford, R. (2003). *Infant/toddler environment rating scale, revised edition*. New York: Teachers College, Columbia University.

Harms, T., Cryer, D., & Clifford, R. (2007). *Family child care environment rating scale, revised edition*. New York: Teachers College, Columbia University.

Harms, T., Vineberg Jacobs, E., & Romano White, D. (1996). *School-age care environment rating scale*. New York, NY: Teachers College Press.

Hausfather, A., Toharia, A., LaRoche, C., & Engelsmann, F. (1997). Effects of age of entry, day-care quality, and family characteristics on preschool behavior. *Journal of Child Psychology and Psychiatry, 38*(4), 441-448.

Helburn, S. (1995). Cost, quality and child outcomes in child care centers. Technical report, public report, and executive summary. Denver, CO: Economics Department, University of Colorado at Denver.

Henderson, A. T., & Berla, N. (1994). A new generation of evidence: The family is critical to student achievement. Washington, DC: Center for Law and Education.

Henderson, A. T., & Mapp, K. L. (2002). A new wave of evidence: The impact of school, family and community connections on student achievement. Austin, TX: Southwest Educational Development Laboratory.

Herman, J. L., & Winters, L. (1994). Portfolio research: A slim collection. *Educational Leadership, 52*(2), 48-55.

Hestenes, L. L., Kontos, S., & Bryan, Y. (1993). Children's emotional expression in child care centers varying in quality. *Early Childhood Research Quarterly, 8*(3), 295-307.

Holochwost, S., DeMott, K., Buell, M., Yannetta, K., & Amsden, D. (2009). Retention of staff in the early childhood education workforce. *Child and Youth Care Forum, 38*(5), 227-237.

Honeycutt, D. (2008). *The relationship of teacher qualifications, quality of care and student achievement outcomes in the Arkansas Better Chance program.* Ed.D., University of Arkansas, Fayetteville, AR.

Horton, C. (2003). Protective factors literature review: Early care and education programs and the prevention of child abuse and neglect. Retrieved from http://strengtheningfamilies.net/images/uploads/pdf_uploads/LiteratureReview.pdf

Howes, C. (1987). Quality indicators in infant and toddler childcare: The Los Angeles study. In D. Phillips (Ed.), *Quality in childcare: What does research tell us?* (pp. 81-88). Washington, DC: National Association for the Education of Young Children.

Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., et al. (2008). Ready to learn? Children's pre-academic achievement in pre-kindergarten programs. *Early Childhood Research Quarterly, 23*(1), 27-50.

Howes, C., James, J., & Ritchie, S. (2003). Pathways to effective teaching. *Early Childhood Research Quarterly, 18*(1), 104-120.

Howes, C., Phillips, D., & Whitebook, M. (1992). Thresholds of quality: Implications for the social development of children in center-based child care. *Child Development, 63*(2), 449-460.

Howes, C., & Smith, E. (1995). Relations among child care quality, teacher behavior, children's play activities, emotional security, and cognitive activity in child care. *Early Childhood Research, 10*(4), 381-404.

Howes, C., & Stewart, P. (1987). Child's play with adults, toys, and peers: An examination of family and child-care influences. *Developmental Psychology, 23*(3), 423-430.

Howes, C., Whitebook, M., & Phillips, D. (1992). Teacher characteristics and effective teaching in child care: Findings from the national child care staffing study. *Child and Youth Care Forum, 21*(6), 399-414.

Huang, D., Cho, J., Mostafavi, S., & Nam, H. (2008). What works? Common practices in high functioning afterschool programs: The national partnership for quality afterschool learning final report. Austin, TX: SEDL.

Hubbs-Tait, L., Culp, A. M., Huey, E., Culp, R., Starost, H.-J., & Hare, C. (2002). Relation of Head Start attendance to children's cognitive and social outcomes: Moderation by family risk. *Early Childhood Research Quarterly, 17*(4), 539-558.

Huffman, L. R., & Speer, P. W. (2000). Academic performance among at-risk children: The role of developmentally appropriate practices. *Early Childhood Research Quarterly, 15*(2), 167-184.

Hunt, K. (1998). *Abandoned to the state: Cruelty and neglect in Russian orphanages*: Human Rights Watch.

James-Burdumy, S., DSynarski, M., Moore, M., Deke, J., Mansfield, W., & Pistorino, C. (2005). When schools stay open late: The national evaluation of the 21st century community learning centers program: Final report: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Jordan, C., Parker, J., Donnelly, D., & Rudo, Z. (2009). A practitioner's guide: Building and managing quality afterschool programs Retrieved 7/8/2010, from http://www.sedl.org/afterschool/practitioners_guide_to_afterschool_programs.pdf

Kadlec, M. B., & Cermak, S. A. (2002). Activity level, organization, and social-emotional behaviors in post-institutionalized children. *Adoption Quarterly, 6*(2), 43-57.

Kantrowitz, E. J., & Evans, G. W. (2004). The relation between the ratio of children per activity area and off-task behavior and type of play in day care centers. *Environment and Behavior, 36*(4), 541-557.

Keith, T., Keith, P., Troutman, G., Bickley, P., Trivette, P., & Singh, K. (1993). Does parental involvement affect eighth-grade student achievement? Structural analysis of national data. *School Psychology Review, 22*(3), 474-496.

Kelley, P., & Camilli, G. (2007). The impact of teacher education on outcomes in center-based early childhood education programs: A meta-analysis (Vol. 15, pp. 2007). New Brunswick, NJ: National Institute for Early Education Research, Rutgers, The State University of New Jersey.

Keltner, B. (1990). Family characteristics of preschool social competence among black children in a Head Start program. *Child Psychiatry and Human Development, 21*(2), 95-108.

Kempe, A., Beaty, B., Englund, B. P., Roark, R. J., Hester, N., & Steiner, J. F. (2000). Quality of care and use of the medical home in a state-funded capitated primary care plan for low-income children. *Pediatrics, 105*(5), 1020-1028.

Kisker, E. E., Hofferth, S., Phillips, D., & Farquhar, E. (1991). A profile of child care settings: Early education and care in 1990. Princeton, NJ: Mathematica Policy Research, Inc.

Klebanov, P. K., & Brooks-Gunn, J. (1998). The contribution of neighborhood and family income to developmental test scores over the first three years of life. *Child Development, 69*(5), 1420-1436.

Knopf, H. T., & Swick, K. J. (2007). How parents feel about their child's teacher/school: Implications for early childhood professionals. *Early Childhood Education Journal, 34*, 291-296.

Koh, S., & Neuman, S. B. (2009). The impact of professional development in family child care: A practice-based approach. *Early Education & Development, 20*(3), 537 - 562.

Kontos, S. (1994). The ecology of family day care. *Early Childhood Research Quarterly, 9*(1), 87-110.

Kontos, S., Howes, C., & Galinsky, E. (1996). Does training make a difference to quality in family child care? *Early Childhood Research Quarterly, 11*(4), 427-445.

Kontos, S., Howes, C., Shinn, M., & Galinsky, E. (1995). *Quality in family child care and relative care*. New York, NY: Teachers College Press.

Kontos, S., Hsu, H.-C., & Dunn, L. (1994). Children's cognitive and social competence in child-care centers and family day-care homes. *Journal of Applied Developmental Psychology, 15*(3), 387-411.

Kontos, S., & Wilcox-Herzog, A. (1997). Influences on children's competence in early childhood classrooms. *Early Childhood Research Quarterly, 12*(3), 247-262.

Kontos, S. J. (1991). Child care quality, family background, and children's development. *Early Childhood Research Quarterly, 6*(2), 249-262.

Krauss, M., Upshur, C., Shonkoff, J., & Hauser-Cram, P. (1993). The impact of parent groups on mothers of infants with disabilities. *Journal of Early Intervention, 17*(1), 8-20.

Kumpfer, K., Alvarado, R., Smith, P., & Bellamy, N. (2002). Cultural sensitivity and adaptation in family-based prevention interventions. *Prevention Science, 3*(3), 241-246.

La Paro, K. M., Sexton, D., & Snyder, P. (1998). Program quality characteristics in segregated and inclusive early childhood settings. *Early Childhood Research Quarterly, 13*(1), 151-167.

Lamb, M. E., Hwang, C., Broberg, A., & Bookstein, F. L. (1988). The effects of out-of-home care on the development of social competence in Sweden: A longitudinal study. *Early Childhood Research Quarterly, 3*(4), 379-402.

Landry, S. H., Anthony, J. L., Swank, P. R., & Monseque-Bailey, P. (2009). Effectiveness of comprehensive professional development for teachers of at-risk preschoolers. *Journal of Educational Psychology, 101*(2), 448-465.

Layzer, J., & Goodson, B. (2006). National study of child care for low-income families care in the home: A description of family child care and the families and children who use it: Wave 1 report Cambridge, MA: Abt Associates.

Li, V., Coates, T., Spielberg, L., Ewart, C., Dorfman, S., & Huster, W. (1984). Smoking cessation with young women in public family planning clinics: The impact of physician messages and waiting room media. *Preventive medicine, 13*(5), 477-489.

Lightfoot, S. (2003). *The essential conversation: What parents and teachers can learn from each other*. New York, NY: Random House.

Linver, M. R., Brooks-Gunn, J., & Kohen, D. E. (2002). Family processes as pathways from income to young children's development. *Developmental Psychology, 38*(5), 719-734.

Little, P., Wimer, C., & Weiss, H. B. (2007). After school programs in the 21st century: Their potential and what it takes to achieve it. *Harvard Family Research Project Brief*. Retrieved from http://www.hfrp.org/

Long, N., Rickert, V., & Ashcraft, E. (1993). Bibliotherapy as an adjunct to stimulant medication in the treatment of attention-deficit hyperactivity disorder. *Journal of pediatric health care: official publication of National Association of Pediatric Nurse Associates & Practitioners, 7*(2), 82.

Lonigan, C., Wagner, R., Torgeson, J., & Rashotte, C. (2002). *Preschool comprehensive test of phonological and print processing (pre-ctopp)*: Department of Psychology, Florida State University.

Love, J. M., Constantine, J., Paulsell, D., Boller, K., Ross, C., Raikes, H., et al. (2004). The role of Early Head Start programs in addressing the child care needs of low-income families with infants and toddlers: Influences on child care use and quality: U.S. Department of Health and Human Services; Administration for Children and Families; Office of Planning, Research and Evaluation; Administration on Children, Youth and Families; Head Start Bureau.

Love, J. M., Harrison, L., Sagi-Schwartz, A., van Ijzendoorn, M. H., Ross, C., Ungerer, J. A., et al. (2003). Child care quality matters: How conclusions may vary with context. *Child Development, 74*(4), 1021-1033.

Love, J. M., Kisker, E., Ross, C., Schochet, P., Brooks-Gunn, J., Paulsell, D., et al. (2002). Making a difference in the lives of infants and toddlers and their families: The impacts of Early Head Start volume 1: Final technical report: Child Outcomes Research and Evaluation; Office of Planning, Research, and Evaluation; Administration for Children and Families; Head Start Bureau Administration on Children, Youth and Families; Department of Health and Human Services.

Lower, J. K., & Cassidy, D. J. (2007). Child care work environments: The relationship with learning environments. *Journal of Research in Childhood Education, 22*(2), 189 - 204.

Makino, S., Gold, P. W., & Schulkin, J. (1994). Corticosterone effects on corticotropin-releasing hormone mrna in the central nucleus of the amygdala and the parvocellular region of the paraventricular nucleus of the hypothalamus. *Brain Research, 640*(1-2), 105-112.

Marcon, R. A. (1999). Differential impact of preschool models on development and early learning of inner-city children: A three-cohort study. *Developmental Psychology, 35*(2), 358-375.

Mashburn, A. J. (2008). Quality of social and physical environments in preschools and children's development of academic, language, and literacy skills. *Applied Developmental Science, 12*(3), 113-127.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., et al. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development, 79*(3), 732-749.

Mason, J., & Stewart, J. (1989). *The CAP early childhood diagnostic instrument* (prepublication edition ed.): American Testronics.

McCartney, K. (1984). Effect of quality of day care environment on children's language development. *Developmental Psychology, 20*(2), 244-260.

McCartney, K., Scarr, S., Rocheleau, A., Phillips, D., Abbott-Shim, M., Eisenberg, M., et al. (1997). Teacher-child interaction and child-care auspices as predictors of social outcomes in infants, toddlers, and preschoolers. *Merrill Palmer Quarterly, 43*, 426-450.

McCartney, K., Scarr, S., Phillips, D., & Grajek, S. (1985). Day care as intervention: Comparisons of varying quality programs. *Journal of Applied Developmental Psychology, 6*, 247-260.

McCormick Center for Early Childhood Leadership. (2010). Head Start administrative practices, director qualifications, and links to classroom quality *Research Notes*. Chicago, IL: National-Louis University.

McIntosh, N., Clark, N., & Howatt, W. (1994). Reducing tobacco smoke in the environment of the child with asthma: A cotinine-assisted, minimal-contact intervention. *Journal of Asthma, 31*(6), 453-462.

McKelvey, L. M., Bokony, P., & Patrick, T. (2010). *Tips for Great Kids! Findings from the year 4 parent surveys.* University of Arkansas for Medical Sciences.

Medical Home Initiatives for Children with Special Needs Project Advisory Committee. (2002). The medical home. *Pediatrics, 110*, 184-186.

Meisels, S. J., Xue, Y., & Shamblott, M. (2008). Assessing language, literacy, and mathematics skills with "Work Sampling for Head Start". *Early Education and Development, 19*(6), 963-981.

Mill, D., & Romano-White, D. (1999). Correlates of affectionate and angry behavior in child care educators of preschool-aged children. *Early Childhood Research Quarterly, 14*(2), 155-178.

Miller, B. M. (2003). *After-school time and children's behavioral adjustment*. Boston, MA: Nellie Mae Foundation.

Miller, B. M. (2005). Pathways to success for youth: What counts in after-school. Massachusetts after-school research study (MARS) report. Wellesley, MA: National Institute for Out-of-School Time.

Modigliani, K. (1993). *Child care as an occupation in a culture of indifference*. Boston, MA: Wheelock College.

Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *Future Child, 5*(2), 113-127.

Nader, P. R., Bradley, R. H., Houts, R. M., McRitchie, S. L., & O'Brien, M. (2008). Moderate-to-vigorous physical activity from ages 9 to 15 years. *Jama, 300*(3), 295-305.

National Association for the Education of Young Children. (1993). A conceptual framework for early childhood professional development. Washington, DC: National Association for the Education of Young Children.

National Institute of Child Health and Human Development Early Child Care Research Network. (2000). Characteristics and quality of child care for toddlers and preschoolers. *Applied Developmental Science, 4*(3), 116.

National Institute on Out-of-School Time. (2000). Making an impact on out-of-school time. Wellesley, MA: National Institute on Out-of-School Time.

National Institute on Out-of-School Time. (2009). Making the case: A 2009 fact sheet on children and youth in out-of-school time Retrieved 05/27/2010, from http://www.niost.org/pdf/factsheet2009.pdf

National Research Council. (2001). Eager to learn: Educating our preschoolers. In B. Bowman, M. Donovan & S. Burns (Eds.). Washington, DC: Committee on Early Childhood Pedagogy, Commission on Behavioral and Social Sciences Education.

New, R. (1998). Theory and praxis in Reggio Emilia: They know what they are doing, and why. *The hundred languages of children: The Reggio Emilia approach–Advanced reflections*, 261-284.

NICHD Early Child Care Research Network. (1996). Characteristics of infant child care: Factors contributing to positive caregiving. *Early Childhood Research Quarterly, 1*, 269-306.

NICHD Early Child Care Research Network. (1999). Child outcomes when child care center classes meet recommended standards for quality. *American Journal of Public Health, 89*(7), 1072-1077.

NICHD Early Child Care Research Network. (2000). Characteristics and quality of child care for toddlers and preschoolers. *Applied Developmental Science, 4*(3), 116-135.

NICHD Early Child Care Research Network. (2001). Nonmaternal care and family factors in early development: An overview of the NICHD study of early child care. *Journal of Applied Developmental Psychology, 22*(5), 457-492.

NICHD Early Child Care Research Network. (2002). Child-care structure → process → outcome: Direct and indirect effects of child-care quality on young children's development. *Psychological Science, 13*(3), 199-206.

NICHD Early Child Care Research Network. (2003). Does quality of child care affect child outcomes at age 4 1/2? *Developmental Psychology, 39*(3), 451-469.

Nicklas, T. A., Baranowski, T., Baranowski, J. C., Cullen, K., Rittenberry, L., & Olvera, N. (2001). Family and child-care provider influences on preschool children's fruit, juice, and vegetable consumption. *Nutrition Reviews, 59*(7), 224-235.

Noam, G. G., Biancarosa, G., & Dechausay, N. (2001). Learning beyond school: Developing the field of afterschool education. Cambridge, MA: Harvard Graduate School of Education: Program in Afterschool Education and Research.

Nores, M., Belfield, C., Barnett, W., & Schweinhart, L. (2005). Updating the economic impacts of the High/Scope Perry preschool program. *Educational Evaluation and Policy Analysis, 27*(3), 245.

Norris, D. J. (2001). Quality of care offered by providers with differential patterns of workshop participation. *Child & Youth Care Forum, 30*(2), 111-121.

Norris, D. J., & Dunn, L. (2004). "Reaching for the Stars" family child care home validation study: Early Childhood Collaborative of Oklahoma.

Norris, D. J., Dunn, L., & Eckert, L. (2003). "Reaching for the Stars" center validation study final report (November 2003 ed.): Early Childhood Collaborative of Oklahoma.

O'Connor, T. G., Rutter, M., Beckett, C., Keaveney, L., & Kreppner, J. M. (2000). The effects of global severe privation on cognitive competence: Extension and longitudinal follow-up. *Child Development, 71*(2), 376-390.

Ohio Collaborative. (2009). Step up to quality research initiative: The Ohio State University.

Olds, D. L., Eckenrode, J., Henderson, C. R., Jr, Kitzman, H., Powers, J., Cole, R., et al. (1997). Long-term effects of home visitation on maternal life course and child abuse and neglect: Fifteen-year follow-up of a randomized trial. *Jama, 278*(8), 637-643.

Olds, D. L., Robinson, J., Pettitt, L., Luckey, D. W., Holmberg, J., Ng, R. K., et al. (2004). Effects of home visits by paraprofessionals and by nurses: Age 4 follow-up results of a randomized trial. *Pediatrics, 114*(6), 1560-1568.

Ouellet-Morin, I., Tremblay, R. E., Boivin, M., Meaney, M., Kramer, M., & Cote, S. M. (2010). Diurnal cortisol secretion at home and in child care: A prospective study of 2-year-old toddlers. *Journal of Child Psychology and Psychiatry, 51*(3), 295-303.

Palfrey, J. S., Sofis, L. A., Davidson, E. J., Liu, J., Freeman, L., & Ganz, M. L. (2004). The pediatric alliance for coordinated care: Evaluation of a medical home model. *Pediatrics, 113*(5), 1507-1516.

PAS; Talan, T., & Bloom, P. (2004). *Program administration scale*. New York: Teachers College Press.

Peisner-Feinberg, E., & Burchinal, M. (1997). Relations between preschool children's child-care experiences and concurrent development: The cost, quality, and outcomes study. *Merrill Palmer Quarterly, 43*, 451-477.

Peisner-Feinberg, E., Burchinal, M., Clifford, R., Culkin, M., Howes, C., Kagan, S., et al. (2000). The children of the cost, quality, and outcomes study go to school: Technical report. Chapel Hill, NC: FPG Publications Office, University of North Carolina at Chapel Hill, Frank Porter Graham Child Development Center.

Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., et al. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development, 72*(5), 1534-1553.

Pence, A., & Goelman, H. (1991). The relationship of regulation, training, and motivation to quality of care in family day care. *Child & Youth Care Forum, 20*(2), 83-101.

Perlman, M., Zellman, G. L., & Le, V.-N. (2004). Examining the psychometric properties of the early childhood environment rating scale-revised (ECERS-r). *Early Childhood Research Quarterly, 19*(3), 398-412.

Perry, C. L., Luepker, R. V., Murray, D. M., Kurth, C., Mullis, R., Crockett, S., et al. (1988). Parent involvement with children's health promotion: The Minnesota home team. *American Journal of Public Health, 78*(9), 1156-1160.

Perry, D., Dunne, M., McFadden, L., & Campbell, D. (2008). Reducing the risk for preschool expulsion: Mental health consultation for young children with challenging behaviors. *Journal of Child and Family Studies, 17*(1), 44-54.

Phillips, D., Howes, C., & Whitebook, M. (1991). Child care as an adult work environment. *Journal of Social Issues, 47*(2), 49-70.

Phillips, D., McCartney, K., & Scarr, S. (1987). Child-care quality and children's social development. *Developmental Psychology, 23*(4), 537-543.

Phillips, D., Mekos, D., Scarr, S., McCartney, K., & Abbott-Shim, M. (2000). Within and beyond the classroom door: Assessing quality in child care centers. *Early Childhood Research Quarterly, 15*(4), 475-496.

Phillipsen, L., Burchinal, M., Howes, C., & Cryer, D. (1997). The prediction of process quality from structural features of child care. *Early Childhood Research Quarterly, 12*(3), 281-303.

Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., et al. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science, 9*(3), 144 - 159.

Pianta, R., La Paro, K., & Hamre, B. (2008). *Classroom assessment scoring system (CLASS) manual, pre-k*. Baltimore, MD: Paul H. Brookes Publishing Company.

Pierce, K., Bolt, D., & Vandell, D. (2010). Specific features of after-school program quality: Associations with children's functioning in middle childhood. *American Journal of Community Psychology, 45*(3), 381-393.

Powell, D. R. (1978). The interpersonal relationship between parents and caregivers in day care settings. *American Journal of Orthopsychiatry, 48*(4), 680-689.

Puma, M., Bell, S., Cook, R., & Heid, C. (2010). Head Start impact study. Final report. Washington, DC: U.S. Department of Health and Human Services; Administration for Children and Families; U.S. Department of Health and Human Services.

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. New York: Simon and Schuster.

Raikes, H. A., Raikes, H. H., & Wilcox, B. (2005). Regulation, subsidy receipt and provider characteristics: What predicts quality in child care homes? *Early Childhood Research Quarterly, 20*(2), 164-184.

Raikes, H. H., Torquati, J. C., Hegland, S., Raikes, H. A., Scott, J., Messner, L., et al. (2006). Studying the culture of quality early education and care: A cumulative approach to measuring characteristics of the workforce and relations to quality in four midwestern states. In M. Zaslow & I. Martinez-Beck (Eds.), *Critical issues in early childhood professional development* (pp. 111-136). Baltimore, MD: Paul H. Brookes Publishing Co.

Raikes, H. H., Wilcox, B., Peterson, C., Hegland, S., Atwater, J., Summers, J., et al. (2003). Child care quality and workforce characteristics in four midwestern states. Lincoln, NE: The Gallup Organization and the University of Nebraska.

Ramey, C. T., Campbell, F. A., Burchinal, M., Skinner, M. L., Gardner, D. M., & Ramey, S. L. (2000). Persistent effects of early childhood education on high-risk children and their mothers. *Applied Developmental Science, 4*(1), 2-2.

Raver, C., Jones, S., Li-Grining, C., Metzger, M., Champion, K., & Sardin, L. (2008). Improving preschool classroom processes: Preliminary findings from a randomized trial implemented in Head Start settings. *Early Childhood Research Quarterly, 23*(1), 10-26.

Reisner, E., White, R., Russell, C., & Birmingham, J. (2004). Building quality, scale and effectiveness in after-school programs: Summary report of the TASC evaluation. Washington, DC: Policy Studies Associates, Inc.

Reynolds, A. J., & Robertson, D. L. (2003). School–based early intervention and later child maltreatment in the Chicago longitudinal study. *Child Development, 74*(1), 3-26.

Reynolds, A. J., Temple, J. A., Robertson, D. L., & Mann, E. A. (2001). Long-term effects of an early childhood intervention on educational achievement and juvenile arrest: A 15-year follow-up of low-income children in public schools. *Jama, 285*(18), 2339-2346.

Roach, M., Kim, Y. B., O'Connor, C., & Laurion, J. (2009). Impact of a statewide collaborative on Strengthening Families through early care and education. *Early Education & Development, 20*(5), 865 - 879.

Roberts, C., Imrey, P., Turner, J., Hosokawa, M., & Alster, J. (1983). Reducing physician visits for colds through consumer education. *Jama, 250*(15), 1986.

Roid, G. H., & Miller, L. J. (1997). *Leiter-R international performance scale, revised*. Wood Dale, IL: Stoelting Co.

Rosenthal, M. (1994). *An ecological approach to the study of child care: Family day care in Israel*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Rosenthal, R., & Vandell, D. L. (1996). Quality of care at school-aged child-care programs: Regulatable features, observed experiences, child perspectives, and parent perspectives. *Child Development, 67*(5), 2434-2445.

Ryan, S., & Ackerman, D. (2004). Getting qualified: A report on the efforts of preschool teachers in New Jersey's Abbott Districts to improve their qualifications *Creating a Qualified Preschool Teaching Workforce Part I*. New Brunswick, NJ: National Institute for Early Education Research.

Sakai, L. M., Whitebook, M., Wishard, A., & Howes, C. (2003). Evaluating the early childhood environment rating scale (ECERS): Assessing differences between the first and revised edition. *Early Childhood Research Quarterly, 18*(4), 427-445.

Scales, P. C., Benson, P. L., Leffert, N., & Blyth, D. A. (2000). Contribution of developmental assets to the prediction of thriving among adolescents. *Applied Developmental Science, 4*(1), 27 - 46.

Scarr, S., Eisenberg, M., & Deater-Deckard, K. (1994). Measurement of quality in child care centers. *Early Childhood Research Quarterly, 9*(2), 131-151.

Scarr, S., Phillips, D., McCartney, K., & Abbott-Shim, M. (1993). Quality of child care as an aspect of family and child care policy in the United States. *Pediatrics, 91*(1), 182.

Schuster, M. A., Duan, N., Regalado, M., & Klein, D. J. (2000). Anticipatory guidance: What information do parents receive? What information do they want? *Archives of Pediatrics and Adolescent Medicine, 154*(12), 1191-1198.

Schweinhart, L. J., & Weikart, D. P. (1997). The high/scope preschool curriculum comparison study through age 23. *Early Childhood Research Quarterly, 12*(2), 117-143.

SEDL Research Consortium. (2008). The national partnership for quality afterschool learning randomized controlled trial of studies of promising afterschool programs: Summary of findings.

Sheldon, J., Arbreton, A., Hopkins, L., & Grossman, J. (2010). Investing in success: Key strategies for building quality in after-school programs. *American Journal of Community Psychology, 45*(3), 394-404.

Shonkoff, J. P., & Phillips, D. (2000). *From neurons to neighborhoods: The science of early childhood development*. Washington, DC: National Academy Press.

Smith, C., Devaney, T. J., Akiva, T., & Sugar, S. A. (2009). Quality and accountability in the out-of-school-time sector. *New Directions for Youth Development, 2009*(121), 109-127.

Smith, C., & Hohmann, C. (2005). Full findings from the youth program quality assessment validation study. Ypsilanti, MI: High/Scope Educational Research Foundation.

Snider, M. H., & Fu, V. R. (1990). The effects of specialized education and job experience on early childhood teachers' knowledge of developmentally appropriate practice. *Early Childhood Research Quarterly, 5*(1), 69-78.

Spiker, D., Hebbeler, K., & Barton, L. (in press). Measuring quality of early care and education programs for children with disabilities. In M. Zaslow (Ed.), *Measuring quality in early childhood settings*.

Spodek, B. (1996). The professional development of early childhood teachers. *Early Child Development and Care, 115*(1), 115 - 124.

St. Pierre, R., & Layzer, J. (1999). Using home visits for multiple purposes: The comprehensive child development program. *The Future of Children, 9*(1), 134-151.

Starfield, B., & Shi, L. (2004). The medical home, access to care, and insurance: A review of evidence. *Pediatrics, 113*(5 Part 2), 1493-1498.

Steinberg, L., Brown, B. B., & Dornbusch, S. M. (1996). *Beyond the classroom: Why school reform has failed and what parents need to do*. New York, NY: Simon and Schuster.

Stipek, D. J., Daniels, D., Galluzzo, D., & Milburn, S. (1992). UCLA early childhood observation form. *Early Childhood Research Quarterly, 7*(1), 1-19.

Stone, E. J., McKenzie, T. L., Welk, G. J., & Booth, M. L. (1998). Effects of physical activity interventions in youth - an epidemiological perspective. *American Journal of Preventive Medicine, 15*, 298-315.

Supovitz, J. A., MacGowan, A., & Slattery, J. (1997). Assessing agreement: An examination of the interrater reliability of portfolio assessment in Rochester, New York. *Educational Assessment, 4*(3), 237-259.

Sylva, K., Siraj-Blatchford, I., & Taggart, B. (2003). *Assessing quality in the early years: Early childhood environment rating scale: Extension (ECERS-e), four curricular subscales*: Trentham Books Ltd.

Sylva, K., Siraj-Blatchford, I., Taggart, B., Sammons, P., Melhuish, E., Elliot, K., et al. (2006). Capturing quality in early childhood through environmental rating scales. *Early Childhood Research Quarterly, 21*(1), 76-92. doi: DOI: 10.1016/j.ecresq.2006.01.003

Talan, T., & Bloom, P. (2004). ***Program administration scale*****. New York: Teachers College Press.**

Talan, T., & Bloom, P. (2009). ***Business administration scale for family child care*****. New York: Teachers College Press.**

Thornburg, K., Mauzy, D., Mayfield, W., Scott, J., Sparks, A., Mumford, J., et al. (in press). Data driven decision making in preparation for large-scale QRS implementation. In M. Zaslow, K. Tout, T. Halle & Martinez-Beck (Eds.), ***Next steps in the measurement of quality in early childhood settings*****. Baltimore: Brookes Publishing.**

Thornburg, K., Mayfield, W., Hawks, J., & Fuger, K. (2009). The Missouri quality rating system school readiness study.

Timmons, B., Naylor, P., & Pfeiffer, K. (2007). Physical activity for preschool children--how much and how? ***Applied Physiology, Nutrition, and Metabolism, 32*****(S2E), S122-S134.**

Tout, K., Starr, R., Soli, M., Moodie, S., Kirby, G., & Boller, K. (2010). Compendium of quality rating systems and evaluations. Compendium of Quality Rating Systems and Evaluations: Child Trends, Mathematica Policy Research.

Tout, K., Zaslow, M., Halle, T., & Forry, N. (2009). Issues for the next decade of quality rating and improvement systems: Office of Planning, Research, and Evaluation; Administration for Children and Families; U.S. Department of Health and Human Services.

Trost, S., Fees, B., & Dzewaltowski, D. (2008). Feasibility and efficacy of a "move and learn" physical activity curriculum in preschool children. ***Journal of physical activity & health, 5*****(1), 88.**

U.S. Department of Health and Human Services. (1998). Head Start program standards and other regulations ***Part 1305.6*****. Washington, DC: U.S. Government Printing Office.**

U.S. Department of Health and Human Services. (2001a). Building their futures: How Early Head Start programs are enhancing the lives of infants and toddlers in low-income families. Washington, DC: The Commissioner's Office of Research and Evaluation and the Head Start Bureau, Administration on Children, Youth and Families.

U.S. Department of Health and Human Services. (2001b). The surgeon general's call to action to prevent and decrease overweight and obesity. Rockville, MD: US Department of Health and Human Services, Public Health Service, Office of the Surgeon General.

Vandell, D., Burchinal, M., Steinberg, L., & Vandergrift, N. (2010). Do effects of early child care extend to age 15 years? Results from the NICHD study of early child care and youth development. ***Child Development, 81*****(3), 737-756.**

Vandell, D., Reisner, E., & Pierce, K. (2007). Outcomes linked to high-quality afterschool programs: Longitudinal findings from the study of promising afteschool programs. Irvine, CA and Washington, DC: Policy Studies Associates, Inc.

Vandell, D., & Wolfe, B. (2000). Child care quality: Does it matter and does it need to be improved? : Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services.

Vandell, D. L. (2004). Early child care: The known and the unknown. ***Merrill-Palmer Quarterly, 50*****(3), 387-414.**

Vu, J. A., Howes, C., & Jeon, H.-J. (2008). Formal education, credential, or both: Early childhood program classroom practices. *Early Education and Development, 19*(3), 479-504.

Wall, M. A., Severson, H. H., Andrews, J. A., Lichtenstein, E., & Zoref, L. (1995). Pediatric office-based smoking intervention: Impact on maternal smoking and relapse. *Pediatrics, 96*(4), 622-628.

Wallhead, T. L., & Buckworth, J. (2004). The role of physical education in the promotion of youth physical activity. *Quest, 56*(3), 285-301.

Ward, D., Benjamin, S., Ammerman, A., Ball, S., Neelon, B., & Bangdiwala, S. (2008). Nutrition and physical activity in child care: Results from an environmental intervention. *American Journal of Preventive Medicine, 35*(4), 352-356.

Watamura, S. E., Donzella, B., Alwin, J., & Gunnar, M. R. (2003). Morning-to-afternoon increases in cortisol concentrations for infants and toddlers at child care: Age differences and behavioral correlates. *Child Development, 74*(4), 1006-1020.

Westmoreland, H., & Little, P. (2006). Exploring quality in after school programs for middle school-age youth *Harvard Family Research Project Brief*. Cambridge, MA: Harvard Family Research Project; Harvard Graduate School of Education.

Whitebook, M. (2003). Early education quality: Higher teacher qualifications for better living environments. A review of the literature. Berkeley, CA: Center for the Study of Child Care Employment, Institute of Industrial Relations, University of California at Berkeley

Whitebook, M., & et al. (1989). Who cares? Child care teachers and the quality of care in america. Final report, national child care staffing study. Oakland, CA: Child Care Employee Project.

Woodcock, R., McGrew, K., & Mather, N. (2001). *Woodcock-Johnson III tests of achievement*. Itasca, IL: Riverside Publishing.

Yeung, W. J., Linver, M. R., & Brooks–Gunn, J. (2002). How money matters for young children's development: Parental investment and family processes. *Child Development, 73*(6), 1861-1879.

Yohalem, N., Wilson-Ahlstrom, A., Fischer, S., & Shinn, M. (2009). *Measuring youth program quality: A guide to assessment tools, second edition*. Washington, DC: The Forum for Youth Investment.

Yohalem, N., Wilson-Ahlstrom, A., & Yu, D. (2005). Youth program quality assessment and improvement: Celebrating progress and surfacing challenges. Washington, DC: The Forum for Youth Investment, Impact Strategies, Inc.

Yoshikawa, H. (1995). Long-term effects of early childhood programs on social outcomes and delinquency. *The Future of Children, 5*(3), 51-75.

Zaslow, M. E., & Martinez-Beck, I. E. (2005). *Critical issues in early childhood professional development*. Baltimore, MD: Brookes Publishing Company.

Zellman, G., & Perlman, M. (2006). Parent involvement in child care settings: Conceptual and measurement issues. *Early Child Development and Care, 178*(no. 5), 521-538.

Zellman, G., & Perlman, M. (2008). Child care quality rating and improvement systems in five pioneer states: Implementation issues and lessons learned. Santa Monica, CA.

Zellman, G., Perlman, M., Le, V., & Setodji, C. (2008). Assessing the validity of the qualistar early learning quality rating and improvement system as a tool for improving child-care quality. Santa Monica, CA: RAND Corporation.

Zigler, E., & Muenchow, S. (1992). *Head Start: The inside story of America's most successful educational experiment*. New York: Basic Books.

Zigler, E., Taussig, C., & Black, K. (1992). Early childhood intervention: A promising preventative for juvenile delinquency. *American Psychologist, 47*(8), 997-1006.

Zill, N., Resnick, G., Kim, K., McKey, R. H., Clark, C., Pai-Samant, S., et al. (2001). Head Start FACES: Longitudinal findings on program performance. Third progress report. Washington, DC: Research, Demonstration and Evaluation Branch & Head Start Bureau, Administration on Children, Youth and Families, U.S. Department of Health and Human Services.

Zill, N., Resnick, G., Kim, K., O'Donnell, K., Sorongon, A., McKey, R. H., et al. (2003). Head Start FACES 2000: A whole-child perspective on program performance. Fourth progress report. Alexandria, VA Head Start Information and Publication Center.

Zill, N., Sorongon, A., Kim, K., Clark, C., & Woolverton, M. (2006). FACES 2003 research brief: Children's outcomes and program quality in Head Start. Alexandria, VA: Head Start Information and Publication Center.