

Quality Dosage, Thresholds, and Features in Early Childhood Settings: A Review of the Literature

OPRE 2011-5

August 2010



DISCLAIMER:

The views expressed in this publication do not necessarily reflect the views or policies of the Office of Planning, Research and Evaluation, the Administration for Children and Families, or the U.S. Department of Health and Human Services.

This report and other reports sponsored by the Office of Planning, Research and Evaluation are available at <http://www.acf.hhs.gov/programs/opre/index.html>.

ACKNOWLEDGMENTS:

The authors would like to express their appreciation for the extremely helpful guidance and feedback of Ivelisse Martinez-Beck and Nancy Geyelin Margie of the Office of Planning, Research and Evaluation in the Administration for Children and Families/DHHS throughout the preparation of this literature review. The authors also thank Amy Madigan of the Office of the Assistant Secretary for Planning and Evaluation in DHHS and Tamara Halle of Child Trends for their ongoing input and facilitation of the coordination of the Q-DOT project with the *In the Running for School Readiness* project. We thank our partner on this project, Robert Pianta, for his leadership in this field of research. Members of the Q-DOT Technical Working Group provided invaluable assistance in shaping the coding categories and offering links to newly emerging research, including Stephanie Jones, Lynn Karoly, Samuel Odom, Helen Raikes, Thomas Schultz, Catherine Snow, and Deborah Vandell. Charlotte Cabili of Mathematica has supported the project's organization and management, and Alfreda Holmes, August Pitt, and Jane Retter contributed to the quality of the final document.

Quality Dosage, Thresholds, and Features in Early Childhood Settings: A Review of the Literature

OPRE 2011- 5

August 2010

Submitted to:

Ivelisse Martinez- Beck, *Project Officer*
Nancy Geyelin Margie
Office of Planning, Research, and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

Submitted by:

Martha Zaslow
Rachel Anderson
Zakia Redd
Julia Wessel
Child Trends

Louisa Tarullo
Mathematica Policy Research

Margaret Burchinal
University of California-Irvine and
University of North Carolina-Chapel Hill

Contract Number: HHSP23320095642WC
Mathematica Reference Number: 06671.310
Project Director: Louisa Tarullo
Mathematica Policy Research
Washington, DC

This report is in the public domain. Permission to reproduce is not necessary.
Suggested citation: Zaslow, M., Anderson, R., Redd, Z., Wessel, J., Tarullo, L. and Burchinal, M. (2010). Quality Dosage, Thresholds, and Features in Early Childhood Settings: A Review of the Literature, OPRE 2011-5. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.



CONTENTS

	INTRODUCTION	1
I	BACKGROUND ON THE Q-DOT PROJECT: WHY FOCUS ON QUALITY DOSAGE, THRESHOLDS, AND FEATURES?	4
	A. Changing Context for the Measurement of Quality	4
	B. Evidence on the Association Between Measures of Quality and Child Outcomes.....	5
	C. Differences From Findings of Evaluations of Early Childhood Intervention Studies.....	6
	D. Follow-up Analyses Focusing on Thresholds, Timing, and Quality Features	7
II	METHODOLOGY OF THE LITERATURE REVIEW	9
III	SUMMARY OF RESEARCH ON DOSAGE	11
	A. Overview	11
	B. Operationalizations of Dosage.....	11
	1. Current Participation.....	11
	2. Cumulative Participation	12
	C. Patterns of Findings for Dosage.....	13
	1. Current and Cumulative Overall Participation in ECE	13
	2. Current and Cumulative Participation in Center-Based ECE	15
	3. Current and Cumulative Participation in Care of High Quality.....	17
	D. Summary and Implications.....	18
IV	SUMMARY OF RESEARCH ON THRESHOLDS.....	21
	A. Overview	21
	B. Operationalizations of Thresholds	21
	1. Identification of Thresholds Based on Structural Measures of Quality	21
	2. Identification of Thresholds Based on Rating Labels in Widely Used Observational Measures	22
	3. Identification of Quality Ranges by Examination of Score Distribution	23
	4. Examination of Whether Relationship Between Quality and Child Outcomes Is Linear or Curvilinear	23

Contents (continued)

	5. Testing for Differences in the Relationship Between Quality and Child Outcomes in Different Portions of the Quality Range	24
	C. Patterns of Findings for Thresholds	24
	1. Thresholds Based on Structural Features of Quality.....	24
	2. Thresholds Based on High or Low Ratings in Observational Measures of Quality	25
	3. Thresholds Based on the Distribution of Observed Quality	26
	4. Linear Versus Nonlinear Relationship of Quality and Child Outcomes.....	28
	5. Testing for Differences in the Strength of the Associations Between Quality and Child Outcomes in Differing Portions of the Quality Range	29
	D. Summary and Implications.....	30
V	SUMMARY OF RESEARCH ON QUALITY FEATURES.....	32
	A. Overview	32
	B. Operationalizations	32
	C. Patterns of Findings.....	33
	1. Structural Features of Quality.....	33
	2. Process Features of Quality and the Issue of Alignment	35
	D. Summary and Implications.....	36
VI	CONCLUSION	37
	A. Potential for Joint Consideration of the Core Constructs	37
	B. Issues This Literature Review Suggests Should be Considered in Conducting Secondary Analyses	38
	C. Issues This Literature Review Suggests Should Be Considered in Developing Design Options for New Data Collection.....	39
	REFERENCES.....	42

INTRODUCTION

This literature review is one of multiple components to the Child Care and Early Education Quality Features, Thresholds and Dosage and Child Outcomes (Q-DOT) project.¹ The overarching goal of this project is to examine existing evidence and provide new evidence on the issue of whether it is appropriate to move beyond the widespread assumption that children’s outcomes improve linearly with improvements in overall quality in early childhood settings, to a more complex conceptualization that permits for the possibilities that:

- A certain dosage is needed before quality can be linked with child outcomes;
- Certain thresholds of early care and education quality need to be met before more positive outcomes for children are seen;
- The relationship between quality and child outcomes depends on the features of quality in relation to specific aspects of development (for example, whether young children’s early literacy is better predicted by specific aspects of quality—such as amount and characteristics of adult speech to children—than by global measures of quality).

The project is intended to progress through a series of steps building progressively towards design work for fielding a new study of dosage, thresholds, and features. The steps include:

1. Conducting a review of existing research focusing on these issues
2. Considering how the existing literature points to needed modifications in the prevalent conceptualizations of how quality and child outcomes are linked
3. Carrying out secondary analyses with specific data sets to uncover information on thresholds, dosage, and features of quality, including analyses of data with children during infancy and toddlerhood and later in the preschool years, data from programs for which quality standards and participation goals are clearly articulated, and data from children’s participation in early care and education of more widely varying quality and without prescribed participation goals
4. Revising the conceptual model of the linkages between quality and child outcomes based on the secondary data analyses, aiming for a more fully articulated logic model
5. Identifying the implications of all of these phases of the project for fielding a study of thresholds, dosage, and features of quality, including the use of quality measures that allow a careful examination of specific quality features and a design that permits scrutiny of the role of extent of exposure to care above specific levels of quality with respect to specified quality features.

¹ While the name of the contract for the present projects presents the key constructs in the order of features, thresholds, and dosage, we felt it would be more informative to summarize the literature in a different sequence, with the research on dosage first, followed by the research on thresholds and then features. This sequence also parallels the acronym that has been provided for the project: Q-DOT. Throughout this introduction and the subsequent sections of this literature review, the ordering of the key constructs is dosage, thresholds, and features. This ordering is not intended to carry any implications about the relative importance of each of the constructs. It represents only the logic the authors felt worked best for presenting the research.

The literature review presented here is the first of the planned steps. In order for the secondary analyses to make the most meaningful contributions to the goals of this project, it is essential that these analyses both build on and go beyond work done to date. A review of findings from previous research regarding quality dosage, thresholds, and features can identify the directions that the body of findings suggests will be most promising to pursue further. Secondary analyses will require decisions about how to operationalize quality dosage, thresholds, and features, as well as what analytic approaches to take in examining them. A review of the literature for this project will accordingly need to focus not only on patterns of findings, but also on which alternative operationalizations of each of the key constructs and which analytic approaches have proven most illuminating. Secondary analyses will involve testing of a conceptual model. It will be important for the literature review to guide the development of the conceptual model to be tested, and help ensure that such a model goes beyond prevailing models if the research base raises questions about them.

The literature review and secondary analyses, in turn, are intended to build towards further revision of the conceptual model and planning for new data collection. Secondary data analyses will inevitably be constrained by the nature of the data already collected in terms of sample and data collection approach (for example, whether data were collected in the context of an evaluation study with a relatively small sample, or a large and representative national dataset). Secondary analyses will also be constrained by the measures used in an existing dataset. The literature review may contribute to the planning for new data collection by helping to identify needed sampling approaches and what gaps in current measurement approaches are critical to address. The literature review and secondary analyses may cumulatively suggest further revisions to the conceptual model that can only be examined empirically in new data collection.

Thus, the literature review will be a source of input both into the secondary data analyses and the design work for new data collection, including the further revision of a conceptual model around which the planning for new data collection can be organized. The review has six sections:

- **Section I: Background on the Q-DOT Project.** In the first section, background on the Q-DOT project is provided, summarizing the reasons the project is being undertaken, and clarifying what is meant by each of the three core components of dosage, thresholds, and features of quality.
- **Section II: Methodology of the Literature Review.** The second section will describe the approach used in the literature review, including selection criteria for literature reviewed and how methodology and findings were summarized.
- **Section III: Findings from Previous Research on Dosage.** The third section will provide a summary of how dosage has been operationalized in previous research, and will summarize findings for dosage when considered as extent of current or cumulative participation in early care and education, and when considered in relation to type of care and quality of care.
- **Section IV: Findings from the Research on Thresholds.** The fourth section, in parallel, will provide a summary of how thresholds have been operationalized, and will summarize findings according to whether child outcomes have been analyzed for broad ranges of quality, whether analyses consider if quality and child outcomes follow a nonlinear pattern, and most recently, whether the link between quality and child outcomes is stronger in different parts of the quality range.

- **Section V: Findings from the Research on Features.** The fifth section will similarly provide a summary of how quality features have been operationalized. Findings will be summarized separately for structural and process features, and particular consideration will be given to whether the relationship between quality and child outcomes is stronger when more closely aligned features of quality and child outcomes are related.
- **Section VI: Conclusion.** The conclusion will focus especially on results that consider two of the core constructs (for example, dosage and thresholds) simultaneously. The summary will also identify the implications of the review both for secondary analyses and for the final phase of the project involving design and planning for new data collection.
- **Literature Review Tables.** The information synthesized here is based on a comprehensive set of tables. Please see Zaslow et al., 2010.

I. BACKGROUND ON THE Q- DOT PROJECT: WHY FOCUS ON QUALITY DOSAGE, THRESHOLDS, AND FEATURES?

Measures of quality have long been used in early childhood settings to guide improvements in practice within individual classrooms or home-based groups, in research seeking greater understanding of how quality contributes to children's development, and in research evaluating the impact of a quality improvement approach. In the past decade, a further purpose for the measurement of quality has been added and has become quite widespread: the measurement of quality as part of statewide or community-wide quality rating and improvement systems (QRIS). QRIS are currently in 21 states, with many more piloting or exploring the possibility of developing such systems. These systems are intended to provide parents with readily interpretable and accessible summary ratings of quality as information to use in selecting early childhood settings for their children. They are also intended to track the levels of quality throughout a geographical area in order to determine the prevalence of different levels of quality throughout the area, and to monitor for changes over time in light of quality improvement investments.

The widespread use of quality measures has involved greater scrutiny of their contents and attention to how they should be used in creating summary ratings. This greater scrutiny has sparked new research considering the extent to which quality measures predict children's development overall, whether prediction is stronger when particular quality features are considered in relation to the most closely related child outcomes, and whether there are distinct levels of quality that are important to achieve for stronger prediction to child outcomes.

In providing background for this literature review, we turn first to a summary of the changing context of the measurement of quality in early childhood settings, and then to a review of very recent research conducted to begin to address the new questions emerging in the context of the widespread use of quality measurement.

A. Changing Context for the Measurement of Quality

In policy initiatives such as state QRIS, greater attention is being given to the selection of particular measures of quality and how they are used in creating summary ratings (Tout, Zaslow, Halle, & Forry, 2009; Zaslow, Martinez-Beck, Tour, & Halle, forthcoming). The selection of a particular observational quality measure for use in a statewide or community-wide system—and its weighting relative to other components of quality, which are combined into a summary rating of quality—take on particular importance not only because the measure will be implemented at scale throughout a geographical region, but because it is seen as capturing the vision of the key stakeholders launching the quality initiative.

In addition, QRIS and other policy initiatives to improve quality typically involve consequences and incentives. It may be an important consequence simply to make a quality rating widely available to the public. For example, Witte and Queralt (2004) studied both awareness and responsiveness of providers to having quality information about their child care centers posted online. Specifically, the researchers considered the implications of making licensing inspection reports available on the web in Broward County, Florida, combined with outreach to inform the public of the online reports. While enrollment patterns did not change in centers that passed or failed inspections, rates of accreditation and summary ratings of quality routinely collected in subsidy-receiving centers increased significantly following outreach to inform the public of the availability of the reports online. The authors interpreted the findings as indicating that posting inspection reports exerted pressure on child care quality via provider response to having the information made available.

Financial incentives and rewards are another key component of quality initiatives such as QRIS (Tout et al., 2009). Incentives may be provided through tiered reimbursement (higher subsidy reimbursement rates for early childhood settings with higher quality ratings), bonuses for attaining higher quality ratings, financial supports for working toward higher levels of quality, and scholarships for low-income families to participate in settings with higher quality. Such incentives increase the priority that the selected quality measure reflect the perception of key stakeholders, can be used reliably, and makes distinctions among levels of quality that provide a basis for incentives and rewards.

An interesting difference is emerging over time among states implementing quality ratings systems: greater emphasis is being placed on choosing measures of quality that reflect both a safe and positive overall environment as well as an environment that supports early learning and provides a strong foundation for academic achievement (Swenson-Klatt, 2008; Tout et al., 2009). There is a particular focus on the contribution that higher-quality early care and education can make to narrowing the gap in measures of school readiness that emerge by kindergarten entry. The most widely used measures of quality in QRIS to date have been the environmental ratings scales. These include the Early Childhood Environment Rating Scale-Revised (ECERS-R) (Harms, Clifford, & Cryer, 2005); the Infant/Toddler Environment Rating Scale-Revised (ITERS-R) (Harms, Cryer, & Clifford, 1990); the Family Day Care Rating Scale (FDCRS) (Harms, Cryer & Clifford, 1989); the Family Child Care Environment Rating Scale-Revised (FCCERS-R) (Harms, Cryer & Clifford, 2007); and the School-Age Care Environment Rating Scale (SACERS) (Harms, Jacobs, & White, 1996). Yet several states in the pilot stage of their QRIS are complementing these measures of quality with further measures that emphasize instructional quality; these include the Classroom Assessment Scoring System (CASS) (Pianta, La Paro, & Hamre, 2007), included in the Minnesota and Virginia pilot quality rating systems, and the Early Childhood Environment Rating Scale-Extended (ECERS-E) (Sylva, Sirai-Blatchford, & Taggart, 2006) in the Missouri pilot.

B. Evidence on the Association Between Measures of Quality and Child Outcomes

This growing emphasis on quality as providing a basis for early learning—rather than simply a safe and positive overall environment—rests on the assumption that our available measures of quality provide a strong basis for predicting child outcomes that indicate school readiness or early academic achievement. Yet recent studies by Burchinal and colleagues (Burchinal, Kainz, & Cai, et al., in press; Burchinal et al., May 2009) on the strength of the association between children's experiences of quality and their developmental outcomes point to consistent but only modest associations between quality and child outcomes when studied using the assumption of a linear relationship (i.e., that as quality increases, child outcomes will improve).

To examine this issue, Burchinal and colleagues conducted both a meta-analysis and coordinated secondary analyses with data from five large studies of early childhood care and education. In the meta-analysis, a literature review identified studies that had undergone peer review, involved examination of the association between quality and child outcomes using common measures of quality, included at least 10 center-based early childhood classrooms, and focused on preschool-age children across all income levels (between the ages of 3 and 5 years). The inclusion criteria were met by 20 projects (some with multiple published papers or reports) reporting on 97 associations between measures of quality and child outcomes. For the meta-analysis, associations between the measures of quality and child outcomes were converted to partial correlations. Effect sizes were estimated to describe the magnitude of the association between the measures of quality and child outcomes (1) overall; (2) for children ages 2 to 3, 3 to 4, and 4 to 6; and (3) by type of child

outcome (categorized as academic and cognitive outcomes, language outcomes, or social-emotional outcomes).

For each age group, type of outcome, and overall, the results indicate that the effects of quality were statistically different from zero, indicating that quality was consistently related to child outcomes. However, the magnitude of the effects was modest, with partial correlations ranging from .05 to .17. Comparisons of the effect sizes by type of outcome indicated stronger associations for language outcomes than for those related to social and emotional development across all ages.

In further work, Burchinal and colleagues conducted coordinated secondary data analyses to examine the strength of the association between measures of quality and child outcomes (Burchinal et al., in press; Burchinal et al., May 2009). These analyses targeted children from low-income families, since the meta-analysis had not been restricted to low-income samples. Five major data sets were identified that had at least 100 children from low-income families and examined quality in at least 50 classrooms: the National Institute of Child Health and Human Development (NICHD) Study of Early Child Care and Youth Development (NICHD Early Child Care Research Network & Duncan, 2003); the Cost, Quality, and Outcomes Study (Peisner-Feinberg & Burchinal, 1997); the National Center for Early Development and Learning (NCEDL) Prekindergarten Evaluation (Howes et al., 2008); and the Head Start Family and Child Experiences Survey (FACES) from 1997 and 2000 (Administration for Children and Families, 2003).

Partial correlations were computed between the measures of quality and fall-to-spring gains on child outcomes, controlling for site, maternal education, ethnicity, and gender. The partial correlations ranged from 0 to .23, with most less than .10. The average partial correlation was .06 between quality and language outcomes; .03 between quality and academic outcomes, and .02 between quality and social outcomes. When the researchers examined a more liberal estimate of the association—zero-order correlations—they found associations that were approximately twice as large but still modest, with an average correlation of .14 between quality and fall-to-spring change in language outcomes, .06 for academic achievement outcomes, and .06 for social-emotional development outcomes.

In sum, the authors concluded that “both the meta-analysis and the secondary data analysis indicated that higher quality [early care and education] is associated with higher language, academic, and social skills and fewer behavior problems, but associations are quite modest” (Burchinal et al., in press).

C. Differences From Findings of Evaluations of Early Childhood Intervention Studies

The meta-analysis and coordinated secondary analyses reported above focused on studies of association between early care and education of varying quality in relation to child outcomes. However, Burchinal and colleagues (Burchinal et al., in press) note that other meta-analyses have focused on the magnitude of effects in experimental evaluations of early childhood interventions in which children are randomly assigned to an early childhood program or a comparison group. In these analyses, the effect size is reported as the difference between the mean for the treatment and control groups divided by the standard deviation for the control group.

As summarized by Burchinal and colleagues (in press), in a recent meta-analysis, Nelson, Westhues, and MacLeod (2003) estimated effect sizes for 34 preschool intervention programs that had at least one follow-up assessment. For cognitive outcomes, moderate effect sizes during

preschool ($d=.52$) were still observed at grade 8 ($d=.30$). While effect sizes for social-emotional outcomes were smaller during the preschool period ($d=.27$), effects were still detected at the end of high school ($d=.33$). Effect sizes varied according to number of years of intervention and when the intervention started, with larger effects for programs that started earlier and involved more years of intervention. Cognitive outcomes were larger for programs that had an intentional instruction component.

In another meta-analysis focusing on the evaluations of 20 early childhood programs, Karoly, Kilburn, and Cannon (2005) found evidence of significant effects in approximately two-thirds of the programs. They also found larger effects on cognitive outcomes for programs that were more intensive and that focused on improving school readiness.

Why are there only modest associations between quality and child outcomes in the new analyses by Burchinal and colleagues and moderate effect sizes in studies of the effects of early childhood programs? The meta-analyses of the early childhood intervention programs raise several important possibilities (see discussion in Burchinal et al., in press). First, effect sizes in the early childhood intervention programs were larger when exposure started earlier and continued longer. Thus, dosage and timing (difficult to disentangle because a program that starts earlier, as long as it continues, also results in greater dosage) appear to be important. Second, effects were larger for cognitive or achievement outcomes in programs that had an intentional instructional component. Thus, particular features of quality may be important, especially in fostering development in specific domains. Finally, in the evaluations of early childhood interventions, program quality may have consistently surpassed a certain threshold of quality needed to produce moderate to strong effects on child outcomes.

The strongest effects may be present in those early childhood programs that combine early, sustained exposure to programs of consistently high quality with specific quality features such as a focus on instruction. The Abecedarian project (Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001; Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002), in which children from birth to age 5 participated in early care and education presumed to be of very high quality (though not directly observed), had strong effects on cognitive outcomes in the preschool period and evidence of sustained effects into young adulthood.

D. Follow-up Analyses Focusing on Thresholds, Timing, and Quality Features

Burchinal and colleagues (in press; May 2009) conducted follow-up analyses to examine some of these possibilities in their meta-analysis and coordinated secondary analyses with the five major early childhood data sets. These further analyses provide some preliminary evidence of:

- **Threshold effects**, with stronger associations between quality and outcomes occurring when observed quality was in the higher range. In particular, secondary analyses were conducted with the five major early childhood data sets considering quality both as a linear and quadratic term. In the FACES 1997 data, quality scores were more strongly related to language outcomes when quality was in the good-to-high range. The same pattern was found for math outcomes in the NICHD SECCYD and FACES 1997 data. In the NCEDL study, quality measures were more strongly associated with reading outcomes when quality was good to high. Finally, in both the NCEDL and FACES data, quality was more strongly negatively predictive of behavior problems when quality was good to high. Thus, across data sets and outcomes, there is some evidence that quality in

the high range shows stronger associations with child outcomes. In the CQO study, however, low-to-average quality was more strongly associated with child outcomes (here, reading scores). The authors caution that these preliminary analyses are not sufficient to identify a particular threshold of quality for which quality-improvement initiatives should aim. However, they do indicate that nonlinear relationships between quality and child outcomes should be considered and that further work is needed on specifying thresholds for particular aspects of quality and for particular ages.

- **Timing**, related to the age of children in care. In the meta-analysis conducted by Burchinal and colleagues, analyses were carried out focusing separately on effect sizes for children ages 2 to 3, 3 to 4, and 4 to 5. Comparison of the effect sizes for these age groups pointed to significantly stronger effects at younger ages, with findings differing specifically between 3- to 4-year-olds and older children. It is important to note that these analyses focused on children in center-based early childhood settings. As will be noted later, participation in center-based settings at early ages is relatively infrequent, with younger children in a regular care arrangement more likely to be in home-based care. While the meta-analyses included controls for background characteristics, there may have been other child or family characteristics (for example, greater emphasis on early learning in the home) that helped to explain early participation in higher quality care as well as stronger child outcomes. Further research more fully isolating the effects of timing is clearly needed. It was not possible with these data sets to carry out coordinated analyses examining dosage. As noted earlier, it is difficult to disentangle dosage and timing of exposure to care of differing quality. Some recent work seeking to do this will be described later. Further work that systematically focuses on both dosage and timing of care of differing quality will also be important for fully understanding where to target quality improvement efforts.
- **Quality features**, related to more alignment between quality and outcomes. Whereas in the first set of secondary data analyses, child outcomes were examined in relation to global or summary measures of quality, in the follow-up analyses, child outcomes were related to subscale scores, and the pattern of associations between more closely aligned subscales and child outcomes were considered. Results point to stronger but still modest associations between aspects of quality and child outcomes that are more closely aligned—for example, when the CLASS Instructional Support subscale is related to language development and reading and math achievement, and when the CLASS Emotional Support subscale is related to measures of social skills and (negatively) to behavior problems.

In sum, preliminary findings underscore the importance of moving beyond an approach to measuring the implications of quality for children's development that (1) involves only global measures, (2) views the relationship only in terms of linear associations, and (3) does not take into account extent of exposure to care of differing quality. However, preliminary evidence also suggests that careful conceptual work and further secondary data analyses will be needed to determine how best to operationalize dosage and thresholds and to identify the specific features of care that are important to particular child outcomes at earlier and later stages during the first five years.

II. METHODOLOGY OF THE LITERATURE REVIEW

The present literature review was undertaken as a foundational step in building towards a conceptual model, secondary analyses, and eventually a revision of the conceptual model and design work for new data collection focusing on quality dosage, thresholds, and features in early childhood settings.

In determining an effective approach in searching for literature to include in the review, as an initial exploratory step, keyword searches were conducted using the three key terms of quality dosage, thresholds, and features. The keyword searches yielded very few studies. We considered it a strong possibility that findings related to these core constructs were embedded within the research examining the relationship between quality and child outcomes. Research may in fact be focusing on the core constructs of the present project without using these labels either in the titles or text.

A broader strategy for including research in the present review was therefore viewed as essential. To “throw a broader net,” we started our search by using the same criteria that Burchinal and colleagues had used in identifying studies for their meta-analysis. These criteria required that studies:

- Had undergone peer review,
- Involved examination of the association between quality and child outcomes utilizing widely used measures of quality,
- Included at least 10 center-based early childhood classrooms, and
- Focused on preschool-age children (between the ages of 3 and 5 years).

We note that the peer review requirement allowed for the inclusion of government reports that had undergone peer review as well as research published in peer reviewed journal articles. We also note that the requirement to include at least 10 center-based early childhood classrooms allowed for the inclusion of major early childhood study samples that included both center-based and home-based care, such as the NICHD Study of Early Child Care and Youth Development.

Using these selection criteria, our starting point was the set of studies included by Burchinal and colleagues in their meta-analysis (a total of 20 publications). A template for summarizing both the methodology and results of each study in table form was developed and reviewed by the project officer and others on the project team. The template was revised and finalized in light of the feedback received. The template includes pages summarizing:

- Research sample,
- Measures of quality included in the study,
- Measures of family context and of child outcomes,
- Study questions, analytic approach and overall findings, and
- Whether and how analyses considered quality dosage, threshold, and features and/or the interrelationships of these, and results of these analyses.

The reference list for the initial set of studies included in the literature review, and preliminary conclusions from the review of these studies, were presented at the first meeting of the Technical

Working Group for the project, held on January 28, 2010 in Washington, DC. With input from the Technical Working Group, the review was extended to include studies of infants and toddlers. The review was also extended to include studies published after the completion of the meta-analysis by Burchinal and colleagues.

In all, 39 studies were reviewed for the present literature review. We note that one study falls beyond the early childhood age range because its explicit focus on attendance and child outcomes during the elementary school years was considered highly relevant to the present review. All other studies focus on samples with children between birth and 5 and not yet in kindergarten at the point of assessment of quality.

The tables completed for the literature review are contained in a separate document (ACF, 2010). We call the attention of readers to the difference between the pages summarizing the findings for the study questions as stated in the article or report being reviewed and the findings pertaining to quality dosage, threshold, and features. There is clear evidence here that studies carried out with other articulated study questions have included results pertinent to the issue of quality dosage, threshold, and features with some frequency. This underscores the importance of going beyond keyword searches for previous research related to the core constructs of this project. We also note that very recent studies--conducted by teams of researchers aware of the May 2009 and in press findings of Burchinal and colleagues and extending that approach to other samples and datasets--have focused explicitly on the issues of quality thresholds and features.

The present review aimed at extracting and summarizing findings in this body of work as they related to quality dosage, thresholds, and features. Our priorities were to summarize how each of the key constructs was operationalized, the analytic approach used in considering the relationship between quality dosage, thresholds, and features and child outcomes, and to identify patterns of results. The approach taken was one of summarizing and synthesizing. We did not attempt to use meta-analytic approaches in summarizing the findings for the core constructs. Such an approach was not deemed feasible given that analyses focusing on quality dosage, thresholds, and features used widely varying operationalizations and did not necessarily provide summary statistics for the analyses: the constructs that are "core" for the present purposes were often secondary in the papers reviewed.

We turn now to a summary of the results of the review as they reflect on each of the core constructs. We conclude with a summary of findings focusing on studies in which more than one of the key constructs was examined at a time (for example, extent of exposure to differing levels of quality). We conclude with a discussion of the implications of the results for secondary analyses and for the design work for new data collection on the issues of quality dosage, thresholds, and features.

III. SUMMARY OF RESEARCH ON DOSAGE

A. Overview

In the literature, dosage has been operationalized in one of two primary ways: as the amount or timing of current participation in early care and education (ECE), and as the amount or timing of cumulative participation in ECE. Current participation measures hours of participation per day or week, or days of attendance in the current year. Cumulative participation involves total hours or days of participation over a period of years. Cumulative care can also refer to the amount of exposure to a certain type of care or a specific program as well as the amount or proportion of available time spent in formal care.

Both current and cumulative participation may focus on extent of participation in particular types of ECE (such as formal, center-based settings), or on overall participation irrespective of type. Analyses also increasingly focus on extent of current or cumulative participation in ECE at certain quality levels. Research focusing on dosage has also considered transitions that children experience across arrangements, types of ECE, or quality levels.

In the sections that follow, we first give examples of studies taking different approaches to operationalizing dosage. While the broad definition of dosage may be current or cumulative extent of participation in early care and education, the operationalization, or way in which a general concept is given enough specification to provide the basis for a variable in analyses, may vary widely within these in different studies. Because we are laying the groundwork for such specifications in the secondary analyses and new data collection for the Q-DOT project, we illustrate differing operationalizations. We then turn to a summary of findings, distinguishing between studies that consider current and cumulative dosage irrespective of type of ECE, studies that consider current and cumulative dosage in light of type of ECE, and then findings that consider current and cumulative dosage in light of quality.

B. Operationalizations of Dosage

1. Current Participation

Studies considering dosage in terms of current participation often measure dosage in hours or days of attendance in the current early care and education setting. However, even for current hours or days of participation, the specific operationalization varies across studies.

Hours per week. For example, while Votruba-Drzal, Coley and Chase-Lansdale (2004) and Vollings and Feagans (1995) included measures of number of hours per week the child is currently in the ECE setting, McCartney et al. (2010) looked both at current hours in ECE, and the cumulative proportion of time spent in ECE in terms of months.

Full and part day. Howes et al. (2008) looked at participation in full-day (defined as 20 or more hours per week) as opposed to half-day programs during the pre-kindergarten year. We note the concern that definitions of a full-day or even full-year program may vary across programs and studies, and thus that it may be preferable for studies to seek to capture actual hours of participation.

Current participation in ECE of specific types and quality. While most studies have looked at current participation in terms of overall exposure, current participation has also been operationalized in some studies in terms of current participation in ECE of a particular type or level

of quality. For example, NICHD, ECCRN, and Duncan (2003), Broberg, Wessels, Lamb, and Hwang . (1997), and Loeb, Fuller, Kagan, and Carrol (2004) all examined the effects of being in a certain type of care at specific ages, while NICHD, ECCRN, and Duncan (2003) also looked at the effects of exposure to certain levels of quality of care at specific ages.

2. Cumulative Participation

Any study that looks over time (either retrospectively or documenting participation prospectively) and considers for how long a child has participated in a particular program, in ECE overall, in a particular type of ECE, or ECE of a particular quality, is considered here to be measuring cumulative participation. As for current participation, cumulative participation has been operationalized in varying ways.

Attendance in a particular program. One approach has involved measuring number of days of attendance in a particular ECE program. Hubbs-Tait et al. (2002) considered the number of days that the child had attended Head Start in the year between the first day of the program year and the date cognitive assessments were performed. Hill, Brooks-Gunn, and Waldfogel (2003) examined effects on children's cognitive outcomes based on the number of days of attendance within a single type of high quality center program. From a program that offered a total of 500 days of care across two years, different cut points were used, with analyses considering whether children had attended between 100 and 300 days, more than 350 days, or more than 400 days of care. In an investigation into the effects of school attendance among elementary school children in Baltimore, Lamdin (1996) summarized attendance at the school level and examined school level standardized tests in light of overall attendance levels.

Cumulative hours of participation in ECE over a period of years. Cumulative dosage experience has frequently been investigated by looking at the cumulative hours or proportion of time spent in ECE overall or ECE of particular types over a period of years. For example, McCartney et al. (2010) looked at both the cumulative number of hours in care as well as the number of hours in care squared. NICHD ECCRN (NICHD Early Child Care Research Network, 1998, 2000, 2006) used cumulative average hours of attendance per week, up to the age at which the child was being assessed as dosage. Studies have also used the cumulative percentage or proportion of time in care as a measure of dosage.

Age of entry into any ECE. In addition to considering cumulative hours of participation in any type of ECE, some studies have also examined age of first entry into ECE. For example, NICHD ECCRN (1998) looked at age of first entry into any type of nonmaternal care.

Cumulative hours of participation in ECE of differing types and hours as a proportion of total. Vandell et al. (2010) considered both the hours and proportion of time spent in nonrelative care and in center care. A publication by the NICHD ECCRN (2006) compared children with no history of center care participation with those who participated in center-based care at least 33 percent of the available time. Using a similar operationalization of dosage, NICHD ECCRN and Duncan (2003) looked at the proportion of three to four month intervals from birth to 54 months spent in center-based care between 3 and 27 months and between 27 and 54 months. Broberg and colleagues (1997) followed infants in Sweden who started on a waiting list for public center-based care as they eventually moved into a public program, another type of care, or continued to be in parental care at 16, 28, and 40 months. Loeb et al. (2004) monitored whether children were in center-based care at one or both of two waves of data collection (done when the children were approximately 2.5 and 4 years old).

Age of entry into specific ECE types. A few studies have elected to go beyond cumulative hours of participation in different type of care to examine the role of age of first entry into a particular type of care. For example, Broberg and colleagues (1997) incorporated this operationalization into their assessment of the effects of center care by comparing children who entered center care at different ages with children who continued to participate in other types of care. Vollings and Feagans (1995) looked at the age of entry into infant day care.

Cumulative exposure operationalized in terms of transitions. Many studies have looked at cumulative dosage by examining changes in ECE arrangements over time, often across the entire early childhood period before kindergarten entry. Tran and Weinraub (2006) considered having 0, 1, 2, or 3 changes in ECE arrangements between 6 and 9 months, 9 and 12 months, and between 12 and 15 months. Arrangement transitions were categorized in terms of whether or not the caregiver was a family member, with categories being: within family changes, out of family changes, and within family to out of family. Arrangement transitions were also categorized in terms of care location as within home to out of home, and out of home to a different out of home setting. The study also summarized the number of months between 6 and 15 (out of a possible four: 6, 9, 12, and 15) during which a child was in two types of nonmaternal ECE simultaneously. Previously, McCartney et al. (1997) had examined the number of care settings that children attended during early childhood; however, this study did not consider how much time was spent in each care setting.

Other studies focusing on transitions over time have concentrated specifically on changes in the quality of the care received. For example, Dearing, McCartney, and Taylor (2009) used cut points on the ORCE to define quality levels and then examined dosage as the number of spells children spent in various levels of quality of care over time (at 6, 15, 24, 36, and 54 months).

C. Patterns of Findings for Dosage

Our review of the research that considers dosage identifies three broad groupings: (1) studies that consider overall extent of exposure (either current or cumulative) irrespective of type of ECE; (2) studies that consider extent of exposure (either current or cumulative) to different types of ECE; and (3) studies that consider extent of exposure (either current or cumulative) to ECE of differing quality. The descriptions of operationalizations above suggest the important caveat that even within broad categories, there are varying operationalizations. Thus, for example, current exposure to a particular type of care may be operationalized in terms of participation in a program that is full day or part of a day, or in terms of current hours per week of participation as reported by the mother. These differences may make it difficult to aggregate across studies, and may underlie differing patterns of findings.

1. Current and Cumulative Overall Participation in ECE

Findings of the studies that examine overall participation are somewhat contradictory. We summarize first the studies looking at current participation, then at studies considering cumulative participation, and finally at studies considering use of multiple arrangements and transitions across arrangements.

Current participation. Votruba-Drzal et al. (2004) looked at behavioral and cognitive outcomes of very low-income children in light of overall current participation in ECE at 3 years of age and found that more current hours in ECE were associated with a diminished likelihood of scoring in the borderline clinical range for behavioral problems as well as with higher scores on a quantitative skills measure.

Looking at school-age children, Lamdin (1996) found that aggregated school level attendance was related to the percent of students performing better than the national median on standardized assessments of reading and mathematics, as well as combined reading and mathematics standardized test scores. However, this study did not examine whether attendance of an individual child was related to child-level outcomes.

Cumulative participation. Multiple analyses of the data collected by the NICHD Early Child Care Research Network have considered cumulative exposure to ECE, net of type and quality. A 1998 study by this group (NICHD Early Child Care Research Network, 1998) reported that more hours in care during the first two years were associated with less social competence and more behavioral problems. Extending this approach over a longer time period, the NICHD ECCRN (2006) found overall extent of participation in ECE to be associated with higher levels of caregiver-reported problem behaviors at 36 and 54 months, and more caregiver-child conflict at 54 months, yet also with stronger social skills at 24 months. In a recent publication using this dataset, Vandell et al. (2010) report that adolescents who spent more hours in early care and education during their first 4.5 years report slightly more risk-taking and impulsive behaviors than adolescents who had spent fewer hours in ECE. McCartney et al. (2010) report that "...there is an effect of child care hours on externalizing behavior at all levels of quality. The association is multiplicative such that the child care hours effect is smallest in high-quality care and largest in low-quality care." Further the "number of hours spent in early child care predicted externalizing scores, controlling for concurrent child care hours as well as selection factors..." (McCartney et al., 2010, p. 10). It is important to note, however, that these studies operationalized exposure differently and used different samples: Vandell and colleagues looked at overall care exposure and adolescent behavioral outcomes while McCartney and colleagues considered the role of quality in their analyses. Two previous NICHD ECCRN studies (NICHD Early Child Care Research Network, 2000; NICHD Early Child Care Research Network & Duncan, 2003) found no corresponding relationship between cumulative participation in ECE and outcomes when the focus was child cognitive rather than social outcomes.

Numerous studies have considered cumulative exposure to ECE in terms of duration of use of multiple arrangements and transitions across arrangements over time. While our focus here is primarily on exposure to multiple arrangements and number of transitions overall (rather than on consideration of these factors in light of type of care or quality of care), as will be seen, results in this set of studies quickly point towards the need to consider both multiplicity and transitions in light of type and quality.

For example, Tran and Weinraub (2006) considered use of multiple arrangements and transitions between 6 and 15 months (measured at 6, 9, 12, and 15 months). They found that while using multiple arrangements for 3-4 months (as opposed to 0 or 1-2 months) was associated with better child language production and comprehension when the quality of care was good, child language scores were lower when multiple arrangements were used and care quality was moderate or low. Transitions, although only specific types, also played a role in child outcomes. The overall number of changes in care did not predict language comprehension, language production, or cognitive performance. However, infants who transitioned from family member care to nonfamily member care or from care inside the home to care provided outside of the home showed poorer language comprehension. It should be noted, however, that these results only approached significance. McCartney et al. (1997) looked at a similar issue and did find a significant role for the number of care changes. They found that the number of care changes during early childhood was associated with dependency and behavior problems in preschoolers.

Along with attendance at specific ages, some studies also examined the role of starting ages in determining child outcomes. A NICHD ECCRN study (1998) that employed both concurrent and lagged analyses done at 24 months found an association between later entry into care during the first two years of life and more negative behavior. However, a study by Volling and Feagans (1995) seemed to suggest the opposite: children entering into infant care later on in their first year of life were more likely to have positive peer interactions and negative caregiver interactions. Another NICHD ECCRN study (2000) found that while overall cumulative hours in child care did not predict children's cognitive or language development, age of entry did: the longer that children were in center care beginning at 6 months, the better they performed on the measures of cognitive and language development. It is important to note that definitions of "earlier" and "later" entry into care varied across these studies.

2. Current and Cumulative Participation in Center-Based ECE

Studies that have defined dosage as the extent of participation in ECE of particular types have found that children with more exposure to center-based care show more positive outcomes, especially in the cognitive area. Results are concentrated especially among studies looking at cumulative exposure, with one major study of extent of current participation in pre-kindergarten not finding the pattern.

Current participation. Howes et al. (2008) investigated this issue by contrasting participation in half- versus full-day (20 or more hours per week) pre-kindergarten programs in a major national study of pre-kindergarten. They found no evidence that being in a full-day program as opposed to a half-day program was associated with stronger scores on measures of cognitive or social development. However, it is important to note that the study compared outcomes based on the program's given operating hours, not the children's documented daily participation. It may be that some children enrolled in pre-kindergarten programs have inconsistent attendance, and this may act to diminish the difference between actual attendance in full- versus part-day programs. Additionally, if there is an hourly threshold of attendance above which effects become apparent, it may be greater than the 20 hours per week used in this study.

Vollings and Feagans (1995) examined the relationship between hours in care per week and social behaviors for infants and toddlers in center-based care and found an association between being in "day care" for more hours per week with less solitary play and slightly more positive peer interaction. Owen, Klausli, Mata-Otero, and Caughy (2008) found that as children spent more than one year in the center-based programs considered, cognitive and social outcomes improved. There was no control group, so these improvements may reflect developmental changes; however, caregivers in relationship-focused programs (that provided continuity of relationships over time with particular caregivers) did report improved parent-caregiver relations over time.

Cumulative participation. The NICHD Early Child Care Research Network and Duncan (2003) found that center-based care, especially when experienced at specific ages, was related to more positive cognitive outcomes. When exposure occurred between 27 and 54 months, the proportion of time spent in center care was consistently correlated with more positive cognitive outcomes than when the exposure occurred between 3 and 24 months. Broberg et al. (1997) found that children who were attending center-based care at both 16 and 28 months had better verbal and mathematical skills at age 8 than children who were in parental care at those ages. In addition, children who were participating in center-based care at 40 months had better mathematical skills than children in parental care or home-based care at this age.

Loeb et al. (2004) also looked at the effects of being in center care at different time points. Loeb and colleagues examined whether children were in center care at both waves of data collection (at roughly 2 ½ and 4 years of age), were in family, friend, and neighbor care at both waves, or whether they transitioned into center care between the two waves. They found that being in center-based care at both waves was associated with the best outcomes on the Bracken measure of cognitive school readiness and FACES measures of book mechanics, comprehension, and familiarity. Children who transitioned from family, friend, and neighbor care into center care between the two data collection waves also had better outcomes than children who were in family, friend, and neighbor care at both waves, suggesting that even some center care or center care at a later age was associated with benefits to children's cognitive achievement.

The associations between center settings and improved child cognitive outcomes have been reported in samples varying as to income level. Loeb et al. (2004) found that greater participation in center-based care was associated with better cognitive outcomes among low-income children, while NICHD ECCRN (2000) and NICHD ECCRN and Duncan (2003) found similar results with a sample including varying socio-economic status and relatively few low-income children.

Behavioral outcomes associated with increased center-based care exposure were more varied than the almost universally positive cognitive outcomes. One NICHD ECCRN study (2006) found stronger cognitive outcomes but more problematic social outcomes including more behavioral problems at 36 and 54 months and lower social skills at 24 months associated with more hours per week of center-based care at those ages. However, these negative findings are not entirely consistent across the literature. NICHD ECCRN (2006) also found that more time in center-based care was associated with more positive peer interactions at 54 months. And, center-based care has been found to have positive effects on child behavior and social-emotional development when considered jointly with grouping characteristics and age of exposure (see NICHD ECCRN 1998 below).

Type of care was sometimes analyzed with type features, such as the caregiver's status as a relative or nonrelative, that could potentially overlap with care setting (i.e. some nonrelative care could be center-based). Tran and Weinraub (2006) found that while greater use of multiple child care arrangements consisting of family members was associated with stronger language comprehension, a greater use of multiple care arrangements consisting of both family and nonfamily caregivers was associated with weaker language comprehension. In a recently published study, Vandell and colleagues (2010) looked at the outcomes of children in all types of nonrelative care and found that children who were in more hours of nonrelative care across the first 4 ½ years of life had increased risk-taking and impulsivity at age 15. The same study also found that proportion of time spent in center care did not predict to adolescent outcomes.

Center care may affect child outcomes in a number of ways related to factors that are perhaps more often found in such programs, such as defined curricula, caregiver trainings or certifications, presence of books and other educational materials, and access to mental health and curriculum specialists. The findings of a NICHD ECCRN study (1998) suggest that the timing of exposure to groups of at least three children in any type of child care setting may be important to social and behavioral outcomes. For example, at 36 months, children cared for in groups of this size in any setting were less negative, had fewer problems in care, and resisted playing with a forbidden toy better when observed in a lab setting. However, lagged results at 36 months found that children who had experienced group care in their first year had more mother-reported problems at age 3, while children who experienced group care in the second or third year had fewer observed and caregiver-reported problems in the care setting. In addition, McCartney and colleagues (2010) report that "children who spent a greater proportion of time with a large group of peers in any type of child

care setting had higher externalizing scores than other children, and this difference was greater for children who spent more hours in child care." (p. 12).

3. Current and Cumulative Participation in Care of High Quality

A growing number of studies are choosing to define dosage as the extent of exposure to ECE at particular quality levels, particularly high quality early care and education. Some studies such as NICHD ECCRN (2003), which looked at exposure to high quality care at 36 and 54 months, did not find a dosage effect related to quality of care. However, the majority of studies have found associations with extent of exposure to ECE of high quality and cognitive and/or socioemotional child outcomes.

Current participation in care of varying quality. In a study of low-income children, Votruba-Drzal and colleagues (2004) found that children in high quality care, but not those in low quality care, show a steep decline in both internalizing and externalizing behavior problems as hours in care increase. Quality of care was not found to impact cognitive outcomes. In addition, as hours in low quality care increased, children's externalizing behavior problems increased. The authors interpret the findings as suggesting "that extensive hours of care in high-quality arrangements may be protective for children's socioemotional functioning, whereas long hours of care in low-quality settings may be particularly detrimental for children's rates of externalizing behavior problems" (p. 307). It is not clear whether this pattern (and the authors' interpretation) extend to all children or is specific to children from low-income families.

McCartney et al. (2010) found an interaction between hours in care and quality in care with respect to children's externalizing behavior problems. More hours in ECE predicted more externalizing behavior problems at all levels of quality, but the size of the effect was greatest when the care was of poor quality and smallest when the care was of high quality.

Cumulative participation in care of varying quality. Dearing et al. (2009) looked at both predictors and outcomes associated with the number of spells in different quality care during early childhood. Dearing and colleagues measured care quality at 6, 15, 24, 36, and 54 months to determine during how many of those "spells" children were in care of a certain quality. Dearing and colleagues found that spells in high quality care could reduce the gap on achievement measures associated with the income-to-needs ratio. Children who were in high quality care at three or more of the time points during early childhood demonstrated no association between their income-to-needs ratio and their outcomes on broad math, broad reading, and letter-word identification measures. Even one spell of high quality care in early childhood had statistically significant impacts on the math scores of low-income children.

Tran and Weinraub (2006) looked at spells of quality at 6 and 15 months and found that both average quality level and the slope of quality between these two points were predictive of cognitive outcomes in infancy. Although effect sizes were small, average quality was predictive of language comprehension and increasing quality over time predicted cognitive performance.

Although Hill et al. (2003) operationalized dosage in terms of days of attendance over time, because their sample included only high quality center programs, their analyses pertain to dosage of high quality care as well. Hill and colleagues found that days of exposure to a high quality center-based program for low-birth-weight babies was related to outcomes on assessments of intelligence at age 8. Children who attended over 400 of the possible 500 days of care demonstrated a 7 to 10 point increase on the Wechsler Intelligence Scale for children's full and verbal scores at age 8. Low birth-

weight-babies who attended over 350 days of the high quality program also showed significant improvement at age 8, although the results were not as large as those experienced by the group that attended over 400 days.

Hubbs-Tait et al. (2002) also looked at the effects of days of attendance at Head Start, which was classified as a higher quality program required to meet program performance standards. They found that for Head Start children who were at highest socio-demographic risk, as days of Head Start during a year increased, so did sociability and receptive vocabulary scores. Owen et al. (2008) took a somewhat similar approach by using accreditation as a marker of high quality. With longitudinal analyses, Owen and colleagues found that sustained enrollment in an accredited center for at least one year was associated with improvements over time in numerous cognitive and social outcomes including mean school readiness, receptive language, and caregiver reports of internalizing behavior problems.

Dosage may also affect different subgroups differently; high quality care and attendance may most strongly benefit the most socio-economically and socio-demographically disadvantaged. In their investigation into the effects of the number of spells in high quality care, Dearing and colleagues (2009) found pronounced results for children from economically disadvantaged families. Each additional episode of high quality care was associated with a 5 percent of a standard deviation increase in math achievement for children who were at 200 percent of the poverty level. Each additional episode of high quality care was associated with a 7 percent of a standard deviation increase in applied problem scores for children at 194 percent of the poverty level (cut point defined by family income-to-need calculations) and a 6 percent of a standard deviation increase in letter-word identification for children at 185 percent of the poverty level. In looking at days of attendance, Hubbs-Tait and colleagues (2002) also found differential benefits of attendance based on subgroup characteristics. They also found that while attendance itself did not differ by socio-demographic risk, it related to receptive vocabulary scores for children at greater socio-demographic risk but not those at lower risk.

D. Summary and Implications

Perhaps the most striking pattern of findings that we have identified in this review of the research on dosage of young children's exposure to early care and education is the increase in positive outcomes (and in some studies, decrease in negative outcomes) when children attend high quality early care and education program for more time. The pattern of findings is identified in studies focusing on concurrent participation as well as cumulative participation, in both large national studies and in studies with smaller local samples, and is noted for both cognitive and social-emotional outcomes. In recent research, more sustained exposure to high quality care has been found to narrow the gap on measures of achievement between low income and higher income children.

This review of research concurs with other reviews in finding greater exposure to center-based care to be associated with stronger cognitive outcomes in young children. Results are inconsistent for social outcomes. The pattern of results for cognitive outcomes, found in both multi-site national studies and smaller local studies, is found especially in studies considering cumulative dosage of participation.

Finally, there is a cluster of results pointing to greater overall participation in early care and education (irrespective of type and quality) as associated with less positive social outcomes in young children. Multiple reports of this pattern are based on rigorous analyses conducted with the NICHD

ECCRN, a multi-site study that is national in scope. Yet the pattern of results is not consistent across studies, with some smaller-scale studies reporting different results. There are emerging indications that the negative patterns of social behavior may be linked to more exposure to larger size groups of children.

These findings point to the need for further analyses involving the joint examination of dosage with ECE quality and ECE type. We need better understanding of the extent of exposure to high quality ECE that may be associated with positive outcomes in young children. To date, the studies considering more sustained dosages of high quality have used global measures of quality. Future work looking at the specific quality features that young children most benefit from in sustained dosages would be extremely valuable.

In terms of operationalization, this review suggests that wherever possible, future analyses of dosage should aim to move beyond measures of program operation (for example, whether a particular program is offered part day or full day; part year or full year), to measures of individual children's actual participation. Both current and total cumulative actual participation have emerged as potentially of importance in the literature reviewed thus far, especially when considered in combination with quality.

In secondary analyses of existing data, we will be constrained by the limited number of datasets in which actual attendance has been documented prospectively. Thus, for current dosage, in seeking to carry out coordinated analyses involving multiple datasets, it will likely be necessary to rely on operationalizations of dosage in terms of parent report of individual children's hours of current participation in a particular early care and education type or specific setting, based on questions about recent weeks or months. For cumulative dosage, we anticipate that existing datasets will afford operationalizations involving summaries of data from recurrent interviews with parents or teachers regarding current participation, retrospective parental or teacher report of children's participation over a period of a year, or global summaries of whether a child has participated in a particular program for a year as opposed to multiple years. It will be important to take into account issues of data quality for retrospective data on participation for a period of a year or years. It will also be important to give careful consideration to selection effects: whether any association between dosage and child outcomes may be attributable at least in part, to the characteristics of children or families who participate in an early childhood setting or settings in a more or less sustained manner.

Looking towards future data collection, it will be important to consider the potential added value of prospective data collection on participation at the individual child level. Attendance data will involve placing a burden for data collection on caregivers and teachers. It will also involve the possibility of data quality issues, especially missing data. Ideally, before determining that new data collection should involve such prospective collection of attendance data at the individual child level, it would be valuable to conduct specific analyses with a limited number of datasets in which such data have already been collected. Comparisons of findings for dosage that emerge in such datasets from daily attendance data versus parental summary reports, and more specifically findings on prediction to child outcomes from these two sources, would be extremely informative. Irrespective of the findings of such comparative analyses, we note that individual level attendance data may be important simply as descriptive data. That is, it may be valuable for programs to be aware of variation in attendance, and the extent to which programs are indeed achieving an intended dose. However, where the goal is prediction to child outcome, before recommending that new research involve the collection of dosage data through prospective collection of attendance data, it will be important to have preliminary findings indicating that such fine-grained prospective data on dosage

provide a stronger basis for understanding contributors to developmental outcomes than parental or teacher summary report of days or years of participation.

IV. SUMMARY OF RESEARCH ON THRESHOLDS

A. Overview

In turning to the issue of thresholds, the central question is whether young children benefit especially (or only) from participation in early childhood settings that are at or above a certain level of quality. The issue of how to identify thresholds has become increasingly important in the context of efforts by federal, state, and local governments to provide summary ratings of quality. In such systems, specific levels of quality may come with financial incentives for higher quality programs, and with the potential that consumers may be more likely to choose early childhood settings that receive ratings indicating high quality.

For thresholds (in particular among the three core constructs focused upon in this review), operationalization is heavily dependent upon the analytic approach taken. This review finds five different analytic approaches that have been taken to operationalizing thresholds: (1) setting thresholds based on professional recommendations that dichotomize structural measures of quality to identify a positive level, for example, whether an early educator has a bachelor's degree, and whether group size meets recommendations articulated by such professional organizations as the National Association for the Education of Young Children (NAEYC);² (2) identification of ranges of quality based on labels used in ratings in widely used observational measures of quality (for example, distinguishing between care that receives a summary rating of "good" or "excellent" on an environmental rating scale and care that receives a rating below good); (3) identification of ranges of quality based on an empirical method, such as the confidence interval or standard deviation from a particular score on a widely used observational measure; (4) analyses examining whether the relationship between quality and child outcomes is linear or curvilinear, and use of the inflection point, or point at which the relationship between quality and outcomes becomes stronger, to identify a threshold; and (5) analyses examining whether the strength of the relationship between quality and child outcomes differs above and below an identified level of quality.

While we see a progression towards more rigorous analytic approaches over time, key issues about analytic approaches remain unresolved. This review identifies as a key next step the articulation of recommendations for analytic approaches in identifying thresholds.

B. Operationalizations of Thresholds

This review finds an evolution in the research on quality as it pertains to the identification of quality thresholds, with five "waves" of work, each defined by a distinct analytic approach. After detailing the analytic approaches, we turn to an examination of the findings yielded by each.

1. Identification of Thresholds Based on Structural Measures of Quality

In a first set of studies, a high (or low) threshold of quality is set in light of whether or not structural measures of quality, such as caregiver/educator educational attainment, group size, and ratio, meet recommended levels as articulated by professional organizations or government documents. In an early study by Howes, Phillips, and Whitebook (1992), in one of multiple

² As will be noted in the section of this literature review on quality features, structural measures of quality, such as group size, have also been analyzed as continuous variables.

approaches to measuring quality, the researchers set thresholds for poor and high quality in relation to the ratio recommendations articulated in the Federal Interagency Day Care Requirements (FIDCR). Based on the FIDCR recommendations for ratio, the threshold for poor quality was set at five or more children being cared for in an infant classroom by one adult, and nine or more children in toddler and preschool classrooms. We note, however, that this study considered these thresholds in relation to observed quality rather than child outcomes.

As further examples, using data from the Cost Quality and Child Outcomes study as well as the Florida Quality Improvement Study, Howes (1997) examined children's language development and engagement with objects and with peers when their teachers had a bachelor's degree, an associate's degree, or a CDA. In more recent work, Mashburn and colleagues (2008) examined children's development in light of whether classrooms met nine individual "benchmarks of quality" for pre-kindergarten classrooms articulated by the National Institute on Early Education Research (NIEER), as well as a summary score of number of benchmarks met. The benchmarks each set a threshold for high quality based on recommendations of professional organizations (for example, 20 or fewer children in the class; lead teacher has a bachelor's degree).

2. Identification of Thresholds Based on Rating Labels in Widely Used Observational Measures

In a second set of studies, thresholds are defined using the labels on observational measures of quality to identify ranges of quality, including high and low ranges. Labels on the Early Childhood Environment Rating Scale and its variants (ECERS, ITERS, FCCERS), as well as the Classroom Assessment Scoring System (CLASS), all assign evaluative labels to ratings indicating the strength of quality. On the ECERS-R, for example, a rating of 1 corresponds to "inadequate" quality, a rating of 3 to minimal quality, 5 to good quality, and 7 to excellent quality. The CLASS also involves 7 point ratings scales, with ratings of 1 and 2 labeled low quality, 3-5 involving medium levels of quality, and 6-7 labeled as high quality, (see overview of each measure in Halle, Vick, Whittaker, & Anderson, 2010). For each measure, these labels have been assigned based on a review of the literature and understanding of best practice. In research using this approach, analyses test for whether differences in child outcomes are statistically significant above and below the demarcations created based on rating labels. The number of different demarcations and where they are placed differ somewhat across studies, and there is sometimes also slight variation between the way in which evaluative labels are assigned in the original measure and in the ranges used in the study.

Howes and colleagues (1992), in an article referring to thresholds explicitly, "Thresholds of quality: Implications for the social development of children in center-based care", considered quality ranges using data from a longitudinal study of center-based care in California and from the National Child Care Staffing Study. In a first step in the analyses, factor analyses were used to create summary scores of Appropriate Caregiving and Developmentally Appropriate Activities using items from the ECERS and ITERS.³ Four ranges were then delineated on each of the following: (1) scores below 3 on these summary measures were considered inadequate; (2) scores between 3 and 3.9 were considered barely adequate; (3) scores between 4 and 4.9 were labeled as good; and (4) scores at 5

³ The factors that have emerged in more recent work with the ECERS-R have been labeled Teaching and Interactions and Provisions for Learning (see for example, Howes, Burchinal, Pianta, Bryan, Early, Clifford & Barbarin, 2008). Howes et al. (2008) note that these factors are generally consistent with the factors that emerged in earlier work using the ECERS, such as the in the study being described by Howes, Phillips & Whitebook (1992).

and above were labeled very good. These labels correspond roughly to the labels assigned on the ECERS ratings, with the exception that the label “good” is assigned to scores of 5 or above on the ECERS and ITERS (rather than starting at 4). Analyses considered whether children’s social development differed when their center-based settings had summary scores for Appropriate Caregiving or Developmentally Appropriate Activity in these four ranges.

Using a similar approach, Burchinal, Peisner-Feinberg, Bryant, and Clifford (2000) divided the ECERS into three ranges. Analyses were carried out in three datasets that used similar measures of quality and child outcomes: the Cost, Quality and Child Outcomes study, the North Carolina Head Start Partnership study, and the Public Preschool Evaluation Project. This study distinguished among summary rating scores of below 3, 3-4.9, and 5 or higher, referring to these as low, medium and high quality. Scores below a 3 on the ECERS fall below the label of “minimal quality,” while scores below a 5 do not attain a label of “good,” and scores of 5 or higher are labeled “good to excellent.” Analyses tested for whether children’s scores on measures of cognitive and social development differed significantly for those in care quality in these three ranges.

3. Identification of Quality Ranges by Examination of Score Distribution

In a variant on the approach of using rating labels to form quality ranges, in further studies, quality ranges have been delineated based on an examination of the distribution of scores within a sample. Dearing and colleagues (2009) defined high quality on the ORCE within the NICHD SECCYD as care observed to be above the median on the quality score composite generated with all of the quality measures done at each age. The authors note that a score of 3 or higher on the ORCE is generally seen as indicating high quality. Settings with average scores above the median using this approach also had scores on nearly all individual ORCE items of 3 or higher.

Other analyses with the data from the NICHD study have used quartile splits to identify cut points on ratings of quality based on the ORCE (see, for example, NICHD ECCRN 2000 and 2006; Tran & Weinraub, 2006). Analyses have sometimes focused on the extreme quartiles (NICHD ECCRN 2006; Tran & Weinraub, 2006). Hynes and Habasevich-Brooks (2008) conducted analyses using both ranges based on the rating labels used in the ORCE and quartile splits. These approaches resulted in slightly different demarcations in terms of ORCE scores.

Finally, some researchers have created quality ranges by identifying scores falling above or below a confidence interval (Gallagher & Lambert, 2006), or have examined child outcomes in relation to quality scores varying by a standard deviation (Burchinal & Cryer, 2003).

4. Examination of Whether Relationship Between Quality and Child Outcomes Is Linear or Curvilinear

In the introduction to this literature review, we summarized work by Burchinal and colleagues (Burchinal et al., in press; Burchinal et al., May 2009) considering whether the relationship between quality and child outcomes was linear or curvilinear. To briefly recapitulate, the analyses by Burchinal and colleagues found some evidence of a curvilinear relationship, with the shape of the relationship suggesting that child outcomes started to improve only when quality reached a certain fairly high level on widely used quality measures. This approach, in which linear and quadratic terms in regression analyses are examined, has also been used in two publications by the NICHD Early Child Care Research Network (NICHD ECCRN & Duncan, 2003; NICHD ECCRN, 2006).

5. Testing for Differences in the Relationship Between Quality and Child Outcomes in Different Portions of the Quality Range

In a fifth analytic approach, separate “splines” or linear regressions are estimated for the classrooms falling in a higher and lower portion of the quality range. One slope is determined for the classrooms falling in the lower quality range and another in the higher quality range, and the difference between these is tested.

Burchinal, Vandergrift, Pianta, and Mashburn (2010) used this approach in a recent article that also refers to thresholds in its title, “Threshold analysis of the association between child care quality and child outcomes for low income children in pre-kindergarten programs.” The explicit reference to thresholds in this article and the much earlier one by Howes, Phillips, and Whitebook (1992), illustrates how the issue of thresholds has long been a concern in the research on quality and child outcomes, and at the same time how analytic approaches have changed over time. This group of researchers looked at the distribution of scores to define high quality separately for the domains of Emotional Support and Instructional Support on the CLASS while also taking into account scale definitions (similar to the approach noted above involving use of quality labels to form ranges). In this study, the scores of 5 or higher were seen as in the high range on Emotional Support, but scores of 3.25 or higher were seen as in the high range for Instructional Support. Vandell and colleagues use this analytic approach in their very recent study examining the effects of early child care quality on development in adolescence (Vandell et al., 2010).

C. Patterns of Findings for Thresholds

1. Thresholds Based on Structural Features of Quality

It is important to note at the outset of this subsection that not all analyses considering the relationship between structural features of quality and child outcomes take an approach involving thresholds. For example, while analyses taking a thresholds approach may examine children’s outcomes in relation to whether or not a caregiver has a bachelor’s degree, other analyses looking at the relationship between caregiver education and child outcome consider all levels of educational attainment without creating a demarcation. Our focus here is on the research that does take a thresholds approach. However, the next section of this review, focusing on quality features, discusses the research on structural features of quality, irrespective of whether the approach taken in analyses involves thresholds.

The most recent research focusing on structural features of quality and taking a thresholds approach finds very limited evidence for quality thresholds. Early and colleagues (2007) carried out coordinated secondary analyses with the data from seven major studies of early care and education to examine whether observed classroom quality in early childhood classrooms and gain scores in children’s achievement during the year prior to kindergarten were related to the lead teacher’s educational attainment. Of particular importance to the issue of thresholds, these coordinated analyses found little evidence of higher scores on observed classroom quality or of larger gain scores on children’s academic achievement during the pre-kindergarten year when their classroom teachers had completed a bachelor’s degree. Further findings indicated little association with quality or child outcomes according to the highest education level of teachers among those with an early childhood major, or according to whether those teachers with a bachelor’s degree had an early childhood major.

Mashburn and colleagues (2008) examined the pattern of prediction to child outcomes when pre-kindergarten classrooms met each of nine benchmarks of quality proposed by the National Institute of Early Education Research, as well as a summary score of number of benchmarks met. The nine benchmarks included whether: the lead teacher has a bachelor's degree, the lead teacher has training in early childhood and child development, the assistant teacher has a CDA, class size is less than or equal to 20, a comprehensive curriculum is used, there was a child-teacher ratio of 10 to 1 or better, at least one meal a day is offered, the program provides vision, hearing, or other health screening or referral services, and at least one family support service is provided. With only a single exception (having a lead teacher with a bachelor's degree predicted higher teacher ratings of child social competence), findings indicated that the quality benchmarks were not positively associated with children's cognitive, language, or social outcomes.

In contrast, global quality as measured by the ECERS-R, and especially measures of relationships and interactions in the classroom as measured by the Emotional Support and Instructional Support domain scores on the CLASS, were associated with child outcomes. Mashburn and colleagues interpret the findings as indicating that the measures of quality that most consistently and strongly predict child outcomes are those that capture children's direct experience of interactions with teachers.

The findings by Mashburn and colleagues (2008) might be seen as suggesting that if we are to seek thresholds of quality, we should focus on measures that more directly capture children's immediate experiences of interactions in early care and education. We turn to this possibility in subsequent subsections of the review of evidence on thresholds. However before proceeding to these results, we must note a further possibility raised by Early and colleagues (2007). The authors ask whether there may be differences in findings by cohort. They note that studies from the 1990s focusing on caregiver educational attainment (and taking a thresholds approach) did find evidence of associations with child outcomes. Thus, for example, in the set of studies encompassed by the present review, Howes (1997) found that in the Florida Quality Improvement Study sample, when caregivers had either a bachelor's degree with a focus on early childhood or a CDA (focusing explicitly on early childhood), children engaged in more complex play with objects and with peers.

It is not clear if a difference in findings by cohort regarding teachers who do or do not have a bachelor's degree might reflect greater methodological rigor in more recent studies (for example, research more fully controlling for children's initial achievement levels), changes over time in which teachers with a bachelor's degree remain in early childhood classrooms versus move to elementary education, or issues of the quality of higher education programs.

We turn now to studies focusing on observational measures of quality and using different analytic approaches to consideration of possible thresholds of quality.

2. Thresholds Based on High or Low Ratings in Observational Measures of Quality

Two studies included in this review established quality ranges using the levels of quality labeled as poor or good on widely used quality measures. Analyses involve contrasting the development of children in either high quality or low quality care with children in the other settings. While both studies point to differences in the development of children according to whether they were in settings labeled high or low quality (than other), the range identified to be high quality differs somewhat across the studies.

Howes and colleagues (1992) used as their threshold ratings of good (4 or higher) or very good (5 or higher) on two summary scores based on the ECERS and ITERS developed by factor analysis: Appropriate Caregiving and Developmentally Appropriate Activities. Data from three samples were used: two from longitudinal studies in California of children who had entered care as infants and data from the Atlanta site of the National Child Care Staffing Study (Whitebook, Howes & Phillips, 1990). Findings indicate that children from settings rated good or very good on appropriate caregiving were more likely than children from other settings to show secure attachment (as opposed to avoidant or ambivalent). In addition, children from settings rated good in general or very good on developmentally appropriate activities were more likely to be classified as both peer- and adult-oriented rather than solitary.

Work by Burchinal and colleagues (2000) taking this approach found evidence of differences in development, especially between children in low quality versus other ECE settings. This group analyzed data from three studies of center-based ECE that used the same observational measure of quality (the ECERS) and had similar child outcome measures: the Cost, Quality and Child Outcomes study, North Carolina Head Start Partnership study, and the Public Preschool Evaluation Project. This study distinguished between high quality (in which classrooms received a rating of 5 or higher on the ECERS), medium quality (in which classrooms received a rating of 3-4.9) and low quality (in which classrooms received a rating of 1-2.9).

Findings indicate that “children experiencing poor-quality child care on average displayed more behavior problems, fewer language skills, and lower levels of academic skills than did children in medium- or high-quality care” (p. 160). The difference between children in poor and high quality care on a standardized measure of language development approached a standard deviation, while the difference on the measures of reading and math was nearly a third of a standard deviation.

3. Thresholds Based on the Distribution of Observed Quality

Among the studies reviewed, the approach used most frequently to examine thresholds of quality was to create ranges of quality through consideration of the distribution of quality scores. Ranges of quality were created in several ways by: (1) looking at quality scores differing by one standard deviation, (2) creating a confidence interval and defining high and low quality as scores falling above and below the confidence interval, (3) using the median to distinguish high and low quality, and (4) separating quality scores into quartiles and focusing on the highest and lowest.

As with studies creating ranges of quality by relying on rating labels (e.g., identifying settings receiving ratings corresponding with summary scores of good or better), these studies consistently report differences in child outcomes for children participating in ECE settings in different ranges of quality. However, as for the studies relying on quality labels to create ranges, it is difficult to aggregate the findings because of the differences in how the quality ranges were distinguished on empirical grounds and which portions of the range were focused on in analyses (in addition to differences in quality measures used and sample characteristics). We turn now to an overview of findings from studies taking this set of approaches to defining thresholds.

Using data from the NICHD Study of Early Child Care, Burchinal and Cryer (2003) examined child outcomes at 36 months in settings differing in quality by one standard deviation on a summary rating of positive caregiving based on the ORCE. A one standard deviation increase in positive caregiving was associated with an increase of 4.85 points on the Bracken School Readiness Scale, of 2.03 points on the Reynell Developmental Language Comprehension Scale for receptive vocabulary,

a decrease of 1.95 on the total behavior problem score of the Child Behavior Checklist, and an increase of .87 on the Adaptive Social Behavior Inventory prosocial scale.

Gallagher and Lambert (2006) considered the development of children in Head Start programs in light of both quality and the percentage of children with special needs in the classroom. The measure of quality used was the Assessment Profile for Early Childhood Programs (Research Version). High and low quality classrooms were identified as those falling outside of a confidence interval. In these analyses, classrooms with scores in the mid range were dropped. Typically developing children in high quality classrooms had higher scores on measures of early literacy development than those in low quality classrooms, yet their teachers rated them as more disruptive. However, interactions of classroom quality and percentage of children in the classroom with special needs suggested that the behavioral outcomes were moderated by percentage of children with special needs. For example, typically developing children in high quality classrooms with no children with special needs scored higher than children in classrooms with 20 percent or more children with special needs on teacher-report measures of compliance and prosocial behavior. In addition, parent-report of problem behaviors were higher for typically developing children in high quality classrooms with 20 percent or more of the children having special needs than in classrooms with no children with special needs.

Schlieker, White, and Jacobs (1991) looked at the distribution of total raw scores on the ECERS in a sample of Montreal child care centers selected to include children with a range of socioeconomic backgrounds and classrooms with a range of ECERS scores. Summary scores of 190 to 239 were labeled as high quality and summary scores of 93 to 131 as low quality. (The researchers dropped a single score that fell between these.) Analyses contrasted PPVT-R (3) scores for children in the high and low quality centers. Quality was found to explain 7 percent of the variance in PPVT-R scores. Results also indicated that quality explained a higher percentage of the variance in PPVT scores for children from one-parent families (19 percent) than children from two-parent families (4 percent).

Dearing and colleagues (2009) also dichotomized quality scores, using the median scores on measures of quality based on the ORCE to distinguish between high and low quality in the NICHD Study of Early Child Care, at each of five age points or spells between 6 and 54 months. As noted in the section on dosage, findings indicate that differences in child outcomes according to family income-to-needs ratio diminished with the number of spells in high quality care. When children's primary care arrangement was observed to be of high quality for three spells or more, there was no longer an association between income-to-needs ratio and children's scores on assessments of math, reading, and letter-word identification.

Two studies authored by the NICHD ECCRN group used quartile splits to identify quality ranges on the ORCE. In NICHD ECCRN (2000), analyses focused on quality as a predictor of cognitive and language outcomes. Preliminary analyses identified observed language stimulation to be the strongest predictor of child outcomes at 15 and 24 months, and positive caregiving to be the strongest predictor of child outcomes at 36 months. Quartile splits on these measures resulted in groups labeled low quality, low/medium quality, average/high quality, and high quality. Differences in child outcomes between children in the high and low quality groups ranged in terms of effect sizes from .18 to .48, with the largest effect sizes occurring on measures of expressive language at 24 months and comprehension at 36 months.

A 2006 publication by the same research group also used quartile splits on quality measures. Analyses contrasted outcomes for children in settings in the highest and lowest quartiles and also

looked at the quartiles of quality as a continuous variable. Findings contrasting the high and low quartiles indicated that participation in high quality care was modestly related to more positive language development, cognitive outcomes, school readiness and social skills and fewer behavior problems.

4. Linear Versus Nonlinear Relationship of Quality and Child Outcomes

Pianta, in comments at the Technical Working Group meeting for the Q-DOT project in January 2010, questioned whether differences in child outcomes in different ranges of quality, even if these ranges are identified as high or low quality, should be considered tests of the existence of quality thresholds. It is only in research in the past decade that we are seeing analyses explicitly posing the question of whether the relationship between quality and child outcomes is nonlinear, suggesting the presence of thresholds, and whether the relationship between quality and child outcomes is stronger in specific portions of the quality range.

A nonlinear relationship between quality and child outcomes would identify points in the range of quality where it is particularly valuable to invest resources for quality improvement. For example, findings indicating that there is no or very little improvement in child outcomes until a certain level of quality is reached (a relatively flat slope up until a certain quality level) would suggest that efforts should be made to help classrooms get above that level of quality. As another example, findings indicating that there is no longer an improvement in child outcomes above a certain level of quality (a relatively flat slope after a certain quality level) would suggest allocating resources to get to but not necessarily above that quality level.

Research to date has used differing analytic approaches to test for linear versus nonlinear relationships between quality and child outcomes. In the NICHD ECCRN study just described (NICHD ECCRN, 2006), the researchers juxtaposed analyses in which the quartile splits were treated as a continuous variable and when high and low quality levels were contrasted. Findings were similar using these two approaches. The authors interpret the findings as providing evidence of a linear relationship between quality and child outcomes.

In NICHD ECCRN and Duncan (2003), “to test for nonlinear relationships, level, change and residualized change models were run in which child care quality was categorized and means were compared across groups” (p. 1466). The researchers conclude that “although all analyses indicated that children in the highest quality care scored higher than children in the lowest quality care, there was no consistent pattern of evidence regarding thresholds at either low- or high-quality levels.” However, the authors caution that this pattern may be attributable, in part, to a lack of settings in the low quality range in the sample.

Peisner-Feinberg and colleagues (2001), in the follow-up study into elementary school of the Cost, Quality and Child Outcomes study, conclude that the relationship between quality and child outcomes is linear. While these researchers indicate that they carried out preliminary analyses to examine whether there was a nonlinear relationship between quality of early care and education and later child outcomes, they do not describe the analytic approach taken:

[O]ur initial examinations of the data showed no evidence of nonlinear relations between child care quality and children's outcomes, indicating that there is not a specific threshold at which quality begins to have a positive effect. Rather, the linear relations that were found indicate that better quality child care is related to better outcomes for children across the spectrum of quality, so that the more quality is increased the better off children are (p. 1551).

Burchinal and colleagues (Burchinal et al., in press; Burchinal et al., May 2009) in work summarized in the introduction to this literature review, carried out coordinated analyses in five major datasets that include measures of quality and child outcomes: the NICHD Study of Early Child Care, Cost Quality and Child Outcomes, the National Center for Early Learning and Development (NCEDL) study of pre-kindergarten, the Family and Child Experiences Survey 1997 and 2000. To test for the possibility of a nonlinear relationship between quality and child outcomes, regression analyses were carried out including quality as both linear and quadratic terms. Positive quadratic associations were interpreted as indicating that the relationship between quality and child outcomes was stronger at higher levels of quality, whereas negative quadratic associations indicated that quality and child outcomes were more strongly associated at lower quality levels. Findings for selected child outcomes in the FACES 1997, NCEDL, and NICHD studies indicated that quality was more strongly associated with child outcomes when quality was in the high range (good or stronger). The most consistent findings showing this pattern pertained to math outcomes (identified in the NCEDL, NICHD, and FACES 1997 studies). In two of these studies (NCEDL and FACES 1997), a similar pattern was found for behavior problems. In the FACES 1997 data, language scores were more strongly associated with quality in the higher range of quality and in the NCEDL measures of quality were more strongly related to reading outcomes when quality was in the good to high range. In the Cost, Quality, and Child Outcomes Study, however, quality was more strongly related to reading outcomes when in the low to average range.

5. Testing for Differences in the Strength of the Associations Between Quality and Child Outcomes in Differing Portions of the Quality Range

The results reported by Burchinal and colleagues (in press) pointing to nonlinear relationships between quality and child outcomes indicate that for certain outcomes, the relationship between quality and child outcomes is stronger in certain portions of the quality range. In two further very recent publications (just published or in press), analyses include but go beyond reporting on whether the relationship between quality and child outcomes is linear or nonlinear. This most recent work reports on spline analyses, testing for differences in the strength of the association between quality and child outcomes in differing portions of the quality range. We see this work as building and extending the work summarized at the end of the previous section, rather than representing a different or contrasting approach.

In an important study still in press, Burchinal and colleagues examined child outcomes in relation to two summary scores on the CLASS in the NCEDL and SWEEP studies of pre-kindergarten. Examination of the distribution of quality suggested that the higher range of quality on the Emotional Support summary score should be set at scores of 5 or above on a 7-point scale, whereas for Instructional Support, the high range started only at 3.25. The authors predicted that scores in the higher range on Emotional Support would be needed to affect children's social and emotional development, whereas scores in the higher range for Instructional Support would be needed for children to show improvements on measures of academic development.

Results confirm that Emotional Support predicts child outcomes only in the high range of quality. Further, for Emotional Support, the relationship between quality and child outcomes was significantly stronger in the high quality range than in the low to medium range. Findings were more complex for Instructional Support, with results differing for expressive language as an outcome than for reading and math outcomes. In particular, reading and math scores were related to quality only in the high quality range, as for social-emotional outcomes. However expressive language was associated with Instructional Support throughout the full range of quality. At the same time, the strength of the association between instructional quality and the child outcomes was significantly

greater in the higher than the lower range of quality for all three outcomes. The authors interpret the findings as indicating that there are thresholds of quality at which improvements in child outcomes start to be detected.

Vandell and colleagues (2010), in research considering the long-term effects of quality during the early childhood years, also found evidence of a quadratic relationship between quality and cognitive outcomes in adolescence. Results of a spline model indicate that quality in the early years was related to achievement at age 15 in the higher but not lower quality range, with the distinction in quality set at the mean. Differences in the magnitude of the association of quality and achievement in adolescence for the higher and lower portions of the range were marginally significant. Setting the demarcation between higher and lower quality using different values on the ORCE did not result in substantially different patterns of results, “suggesting that our data may not be able to identify a single cut-point for defining thresholds” (p. 21).

D. Summary and Implications

This review of the evidence on thresholds identifies an evolution in the literature in how the possibility of thresholds of quality has been addressed analytically. Research over a period of years has asked (and continues to ask) whether children’s outcomes are stronger when quality is above rather than below a priori cut points for structural or observational measures of quality. It is only in recent work, however, that researchers have entered quadratic as well as linear terms into regression analyses, examining whether the relationship between quality and child outcomes is nonlinear, and contrasting the strength of the relationship between quality and child outcomes in different segments of the quality range.

We have noted that some researchers have questioned whether using a priori cut points for structural and observational measures of quality and contrasting child outcomes above and below the cut points suffices as an approach to identifying thresholds. We need to work towards agreement on what analytic approach and what types of evidence point to the presence of thresholds. Agreement on analytic approach will be critical to the next step of this project, involving secondary analyses.

Across the studies reviewed here, if we prioritize the approaches looking at linear and quadratic terms in regressions and contrasting the strength of association between quality and child outcomes in differing portions of the quality range, we see emerging evidence that quality and child outcomes are associated more strongly in the higher portions of the quality range. These analyses have been carried entirely in large multi-site studies, including parallel secondary data analyses in five major early childhood datasets (Burchinal et al., in press). We note that while there is some consistency in the pattern of findings across these major early childhood datasets, it is also clear that results have not indicated a nonlinear pattern for all outcomes or all measures of quality. (See the detailed results in the replicated analyses in five datasets by Burchinal and colleagues (in press) for variation in the pattern across outcomes.) It will be critical to extend this set of initial results in the secondary analyses for the current project to further examine the consistency of the patterning of results.

In the emerging body of work contrasting the strength of the relationship between quality and child outcomes in different portions of the quality range (analyses all carried out in large multi-site datasets), questions remain about how best to set the point in the quality range above and below which the strength of association is compared. Indeed, Vandell and colleagues (2010) report on alternative approaches to set this “knot.” As part of future work, it will be important to think through the best approach to setting this point: through examination of the distribution of quality

scores, through identification of the inflection point when a curvilinear relationship between quality and child outcomes is identified, or on a conceptual basis.

The research reported here suggests that it will be important in both secondary analyses and in developing design options and planning for new data collection to ensure that the full range of quality is sampled. NICHD and Duncan (2003) cautioned that their conclusions are tempered by the limited number of early childhood settings in the low quality range in the NICHD dataset. In secondary analyses, it will be important to describe the quality range included in specific samples. In developing design options and planning for future data collection, efforts will be needed to ensure that settings span the full range of quality and that there is adequate representation of quality at all levels. While we have been able to include some studies of infant and toddler care, the number of such studies was limited. In planning for new data collection, it will be important to include sufficient numbers of center-based classrooms of infants and toddlers to permit systematic examination of whether thresholds of quality differ for infants and toddlers as opposed to older preschool-age children.

Finally, the research reported here also raises the possibility that findings regarding thresholds may be moderated by such variable as family characteristics (for example, the findings by Schlieker et al. (1991) regarding one versus two parent families) and classroom composition characteristics (for example, the findings reported by Gallagher and Lambert [2006] regarding proportion of children in the classroom with special needs). It will be important in planning for secondary analyses to decide on the priority that will be placed on consideration of such moderators.

V. SUMMARY OF RESEARCH ON QUALITY FEATURES

A. Overview

We turn now to consideration of quality features. The underlying goal in considering quality features is to discern which specific aspects of quality would be of greatest importance, especially at sufficiently high levels and for sustained periods, in order to support children's positive development. In very recent work on this issue, a corollary has been added: which features are most important to which child outcomes?

As summarized in the introduction to this literature review, an emerging hypothesis is that quality features may be particularly important for aspects of development that are most closely aligned with the nature of stimulation and support considered. Thus, for example, the extent and quality of language stimulation in an early childhood setting may be most important to the directly aligned outcome of language development.

Recent work by Burchinal and colleagues (Burchinal et al., in press) provides the basis for this emerging hypothesis. Burchinal and colleagues conducted a set of analyses in which child outcomes were examined in relation to global or summary measures of quality. In the follow-up analyses, child outcomes were related to subscale scores or individual items, and the patterns of associations between more closely aligned items or subscales and child outcomes were considered. Results point to stronger but still modest associations between aspects of quality and child outcomes that are more closely aligned, for example, when the CLASS Instructional Support domain is related to language development and reading and math achievement, and when the CLASS Emotional Support domain is related to measures of social skills and (negatively) to behavior problems. In this section of the literature review, we consider the evidence for alignment by looking at the patterning of results across studies.

B. Operationalizations

Previous research has frequently categorized quality features into two broad groupings: *structural features* of quality and *process features* of quality. Structural features are often described as “regulable” in that standards (thresholds) can be articulated for acceptable or exemplary practice. Structural features are more enduring characteristics of the classroom or group (such as number of children present, ratio of teachers to children, use of a curriculum) and of caregivers or teachers (such as educational attainment and training related to early childhood development). In contrast, process features of quality focus on the young child's immediate experiences of interactions with people and with materials in the early childhood setting and are dynamic.

The NICHD ECCRN (2002) articulated and tested a conceptual model in which structural features of quality affected child outcomes not directly, but indirectly, through process features of quality. More positive structural features of quality, in this model, are seen as increasing the likelihood of, although not assuring, more positive caregiver-child interactions, a key direct predictor of child outcomes. Structural equation modeling found evidence in support of this model.

In a recent reexamination of the relationship between structural and process measures of quality, Mashburn and Pianta (in press) argue that in addition to mediation, in which structural features of quality affect child outcomes indirectly through process quality, structural features of quality can also act as moderators of process quality. For example, the effects of positive caregiver-child interaction can be magnified if there are fewer children overall in the group or a more optimal

ratio of staff to children, because such interactions will occur more often. Poor structural quality can dampen the direct effects of positive interactions through decreasing their frequency.

Mashburn and Pianta (in press) hold that inconsistent findings in the literature on the effects of structural quality on children occur because of inappropriate placement within the model: structural features are included in analyses as if they had direct effects on child outcomes. Few analyses explicitly test the indirect effects on child outcomes of structural features of quality, and virtually none have examined structural features as moderators.

In the present review, we first summarize the evidence on structural features of quality. We consider whether analyses implicitly assume direct effects or explicitly test for indirect effects on child outcomes. We look for consistency vs. inconsistency in the evidence on structural features in light of the conceptualization by Mashburn and Pianta. In turning to process features of quality, the key issue examined is whether results looking across studies follow a pattern in which specific features of quality are noted to be predictors of more closely aligned child outcomes.

C. Patterns of Findings

1. Structural Features of Quality

Structural features of quality, specifically group size, ratio, caregiver education and caregiver training, were examined in a number of the studies encompassed in this review. For each of these structural quality features, the evidence regarding associations with child outcomes is inconsistent.

Group size was found to be a significant predictor of child outcomes in an expected direction in studies by Howes and colleagues (1992), the NICHD ECCRN and Duncan (2003), and Volling and Feagans (1995), but not in those by Blau (1999) and Gallagher and Lambert (2006).

More specifically, path analyses by Howes and colleagues (1992) suggest that group size is indirectly related to social competence with peers, through a process quality measure (increased exposure to developmentally appropriate activities), which leads to improved social orientation with teachers and peers. It is noteworthy that this study, with findings in the predicted direction, considered the indirect relationship of a structural feature of quality through process quality.

In addition, the NICHD ECCRN and Duncan (2003) found smaller group size to be consistently, though modestly, associated with higher cognitive development in some models, but not others. Volling and Feagans (1995) found larger group size to be associated with more nonsocial play. Proportion of time with a large group of peers (a variation of a group size variable) moderated the relationship between number of hours in child care and externalizing behaviors in one recent study (McCartney et al., 2010).

Blau (1999), on the other hand, reported an unexpected finding with larger group sizes associated with higher reading scores, while Gallagher and Lambert (2006) found no association between class size and social behavior and pre-literacy child outcomes.

Ratio was found to be a significant predictor of child outcomes in an expected direction in studies by Blau (1999), Deater-Deckard, Pinkerton, and Scarr (1996), Howes and colleagues (1992), Howes (1997), NICHD ECCRN (2002), and Volling and Feagans (1995), but not in a study by the NICHD ECCRN and Duncan (2003).

More specifically, Blau (1999) found that better child-adult ratios were associated with lower PIAT-reading scores, but not with PIAT-math scores, vocabulary scores or behavior problems. Deater-Deckard and colleagues (1996) found better child-adult ratios to be associated with improved maternal ratings of social withdrawal. Howes (1997), in an analysis of data from the CQO study, found that children in classrooms with child-staff ratios that meet the recommendations of professional organizations had higher pre-literacy scores than children in classrooms that do not meet with compliance standards. Volling and Feagans (1995) found higher child-adult ratios to be associated with more nonsocial play.

Two studies that considered ratio in models considering indirect effects on child outcomes through process quality both report evidence of such a relationship. The path analyses by Howes and her colleagues (1992) indicate that child-adult ratio is indirectly related to social competence with peers, through a process quality measure (increased likelihood of receiving good caregiving), which leads to an increased sense of security with teachers and better relationships with peers. The NICHD ECCRN (2002), using a path analysis model, found child-adult ratio to be indirectly related to social competence through process measures of quality.

However, the NICHD ECCRN and Duncan (2003) examined the relationship between child-adult ratio and child outcomes in several models across multiple time points and found only a marginally significant relationship between ratio and cognitive development outcomes in one set of models at a single time point.

Caregiver Education and Training. The NICHD ECCRN (2002) found caregiver training to be related to child cognitive and socioemotional outcomes indirectly through observed positive caregiving. Further, NICHD ECCRN and Duncan (2003) found caregiver education to be consistently related to child achievement at 54 months, as indicated by a composite score made up of three scales designed to assess math skills, reading skills and phonological knowledge. Howes and colleagues (1997), in their analysis of data from the Florida Quality Improvement Study, found evidence that teacher's specialized education and training and early childhood education is associated with increased complex play with peers, complex play with objects. Similarly, in their analysis of data from the CQO study, Howes and colleagues (1997) found that children in classrooms with teachers who have at least an AA in early childhood education have higher vocabulary and pre-literacy scores than children who do not.

However, Blau (1999) found specialized caregiver training to have inconsistent effects. Specifically, Blau found caregiver training to be associated with higher PIAT-math scores and vocabulary outcomes (but not with PIAT-reading scores or behavior problems) in one very simple model out of several regression models in which relationships between quality and outcomes were examined. In addition, in coordinated secondary analyses of multiple major early childhood datasets, Early et al. (2007) found very little evidence of a relationship between lead teacher educational attainment and children's gain scores on measures of achievement in the year prior to kindergarten. As noted earlier in the discussion of findings for thresholds operationalized in terms of structural features of quality, Mashburn and colleagues (in press) examined the structural quality features of pre-kindergarten classrooms based on nine individual benchmarks for structural quality (e.g., whether or not a classroom met the benchmark of having 20 or fewer children in a class), and a summary index score across all nine benchmarks, and found that structural features did not predict child outcomes, whereas process quality measures did predict child outcomes.

In summary, findings from the present review support the summary articulated by Mashburn and Pianta that structural features of quality are inconsistently related to child outcomes. The

approaches that Mashburn and Pianta call for in order to clarify why findings may be inconsistent were rarely used in this body of work. In particular, there was limited consideration of the relationship between structural features and child outcomes as either mediated by process quality or of structural features as moderating the effects of process quality. However, it is important to note that studies that did look at structural variables as indirectly related to child outcomes through linkages with measures of process quality did provide supporting evidence. There is a clear need for further work on taking this approach.

A number of limitations in the existing research temper our conclusion that findings for structural features of quality are inconsistent. First, most of the studies encompassed in this review focused on preschool-age children. Yet some previous theory suggests that such structural quality features as caregiver education may be relatively more important for the development of infants and toddlers. Another limitation is that our review focuses primarily on studies of center-based care. Yet the range on structural features of quality in center-based care may be truncated because of licensing regulations or because there are quality standards in publicly funded programs (such as Head Start).

2. Process Features of Quality and the Issue of Alignment

There is some indication in the studies reviewed of alignment between quality features studied and particular child outcomes. For example, measures of instructional support tend to predict cognitive and language outcomes. While studies of social and emotional support in early childhood classrooms also show alignment in terms of prediction to social competence and fewer behavior problems, there is also evidence of prediction to cognitive outcomes.

Language stimulation. A number of studies reviewed here identify language stimulation as predictive of language and cognitive outcomes (Howes et al., 2008; McCartney, 1984; NICHD Early Child Care Research Network, 2000; NICHD Early Child Care Research Network & Duncan, 2003). McCartney found that type of language stimulation was also important, with representational language a better predictor than language used in interactions to control behavior.

Measures of instruction. Howes and colleagues (2008) found the ECERS factor score on Teaching and Interactions and the CLASS Instructional Support domain score to be related to language and cognitive outcomes. Mashburn and his colleagues (2008) found the CLASS Instructional Support domain score to be related to language and cognitive outcomes.

Supportive relationship with caregiver/educator. Three of the studies included in this review found supportive relationships with caregiver or teacher to be related to behavioral outcomes, such as fewer behavior problems (Howes et al., 2008; Mashburn et al., 2008; Peisner-Feinberg et al., 2001). In addition, these studies also reported associations of supportive relationships with cognitive outcomes. For example, Howes et al., (2008) found CLASS Emotional Support to be related to behavior problems as well as to math skills.

Finally, while this is only a single study, the work of Kontos and Wilcox–Herzog (1997) has interesting findings regarding activity settings. In particular, direct engagement with the caregiver appeared to be more important than the simple presence of the caregiver. In addition, children’s participation in “high yield” (highly engaging) activities was associated with more positive outcomes.

D. Summary and Implications

This review finds inconsistent relationships between measures of structural features of quality and child outcomes. It is noteworthy, however, that the small subset of studies that consider structural features of quality as related indirectly to child outcomes, with the relationship mediated through process quality measures, provide consistent evidence of such indirect linkages. Among the features of process quality considered, while measures of instructional support were found to be related to cognitive outcomes, measures of emotional support were related both to cognitive and social-emotional outcomes. This set of conclusions is based on large, multi-site studies as well as small, locally based studies.

To further consider these patterns, the secondary analyses conducted for the Q-DOT project could specify a conceptual model in which structural features of quality are seen as affecting child outcomes indirectly through measures of process quality. Inclusion of the direct as well as indirect paths between structural features of quality and child outcomes would help to test the hypothesis articulated by Mashburn and Pianta that structural features of quality are often “misplaced” in models of the effects of quality on child outcomes. With virtually no previous examinations of whether structural features of quality function as moderators, examinations of this possibility would also make a contribution to the understanding of the role of quality in contributing to child outcomes.

There are a number of important limitations in the capacity of this review to provide a thorough examination of the issue of alignment. These limitations underscore the importance of addressing this issue in a systematic way in secondary analyses conducted for the Q-DOT project. First, in order to test for alignment, measures of quality must have sufficient specificity. Yet in the studies covered here, when the measures of quality go beyond global or summary scores, they still tend to focus on broad dimensions (such as instructional support) rather than describe very specific quality features (such as frequency of discussions of mathematical concepts). Second, consideration of alignment requires the inclusion of both conceptually aligned and nonaligned measures of provider quality and child outcomes, so that it is possible to examine differential patterns of association. Yet studies may include only closely aligned quality and child outcome measures because of an implicit hypothesis that these are the particular outcomes that will be related to quality.

Secondary analyses including measures of quality capturing more specific features of process quality would make a valuable contribution. Such analyses should include consideration of both more and less closely aligned quality and child outcome measures, if available.

VI. CONCLUSION

Each section of the literature review has had its own separate summary and discussion of implications. Rather than reiterating the key findings regarding quality dosage, thresholds, and features, we focus here especially on the implications of the review in building towards the next steps of the Q-DOT project. We begin by noting the potential of the next phases of the work to focus especially on the simultaneous or joint consideration of quality dosage, thresholds, and features. We then focus on specific recommendations that the literature review suggests should be considered high priorities for the planned secondary analyses and for the final phase of the project involving developing design options for a new data collection effort.

A. Potential for Joint Consideration of the Core Constructs

Among the most interesting and important findings covered in this review are the results of a small set of studies that give joint consideration to quality dosage and thresholds. For example, we have summarized findings of a study indicating that when low-income children are in high quality care (but not when they are in low quality care), there is a steep decline in both internalizing and externalizing behavior problems as current hours in care increase (Votruba-Drzal et al., 2004). We have also noted research that found that across a period of years, an increase in the number of spells that young children spent in care observed to be of high quality was found to be associated with a decrease in the gap on achievement measures associated with the income-to-needs ratio of the children's family. Three or more spells of high quality care virtually eliminated the gap (Dearing et al., 2009).

These findings need to be placed in the context of research on high quality early intervention studies. As noted in the introduction to this literature review, studies that show sustained effects, even into adulthood, of early intervention programs, such as the Abecedarian project, have involved more than one year of participation. For example, a new implementation study of the Educare project, an extended and particularly comprehensive version of Early Head Start and Head Start, yields compelling findings. As years of participation increase, children's scores on standardized measures of receptive vocabulary increase in a stepwise manner and reach national norms when the full intended dosage of five years is reached. Parallel findings are reported for measures of social behavior in the children (Yazejian & Bryant, 2009).

This review, taken together with the findings of early intervention studies, underscores the need for a new conceptualization of how quality affects child outcomes: to more systematically take into account not only the level of quality but also the duration of exposure. A high priority for the future steps of this project will be to work towards greater understanding of how dosage and quality interact to influence the care experiences of young children. In addition, we need to consider the availability of data in existing datasets that will make it possible to examine further and extend our understanding of outcomes for children exposed to high quality over a period of years.

While the work encompassed in this review is beginning to give indications of the potential importance of joint consideration of quality dosage and thresholds, we found no work that provides similar joint consideration of quality features with either thresholds or with dosage. Our review summarized evidence pointing to patterns of findings in which child outcomes in specific domains appeared to be more systematically related to the features of quality most closely related to that domain (especially for instructional quality and children's language and cognitive outcomes). Yet we found no studies considering children's sustained exposure to early childhood settings involving high quality in terms of very specific quality features.

We also found that studies that could reflect on the issue of alignment have, as yet, gone only one step towards greater specification of quality features: they move beyond consideration of global quality but only to the point of separating out instructional and emotional quality, and in some studies, language stimulation. That is, they specify an *area of interaction* in which quality may or may not be high (they are interaction-specific), but with the exception of language stimulation, they do not yet go so far as to describe high quality in terms of supports for development in specific domains (we are lacking measures of domain-specific quality).

There is a clear need for studies that drill down deeper into more specific features of quality, including a wider range of quality facets. It would be extremely valuable to consider, for example, features of quality related to supporting early mathematics skills in relation to gains in children's math scores, or features of quality related to supporting children's behavioral and attentional self-regulation in relation to children's growth in executive functioning.

This review suggests that as we move forward, we should work towards consideration of more specific quality features. In planning for secondary analyses, it will be very important to seek datasets that permit consideration of both interaction-specific and domain-specific measures of quality in relation to child outcomes. An important underlying question will be whether going beyond interaction-specific measures of quality to consider also domain specific measures yields even stronger prediction to child outcomes.

While consideration of more detailed quality features in relation to child outcomes concurrently will address an important gap in the literature, we need to consider the potential also to seek opportunities for joint examination of quality features and dosage over time. If it is not possible to identify existing datasets that would permit simultaneous consideration of interaction-specific and domain-specific quality features and dosage, this should be considered as a priority for the final phase of the project involving developing design options for new research.

B. Issues This Literature Review Suggests Should Be Considered in Conducting Secondary Analyses

Beyond the high priority that should be placed on joint consideration of quality dosage, thresholds, and features, to the extent possible, in secondary analyses, this review of the literature identified several further issues that should be considered in conducting these analyses:

- This review identifies the need to seek datasets that will permit more fine-grained analyses of quality features. While work to date is promising in suggesting that stronger or more consistent effects obtain when more closely aligned features of quality and child outcomes are considered, the degree of specification of quality features is as yet limited. The identification of datasets in which more fine-grained, domain-specific, quality features are incorporated in the measurement of quality will make it possible to consider thresholds of specific quality features in relation to child outcomes.
- This review suggests that we move away from measuring dosage in terms of program hours or days offered (such as full or part time) to extent of actual participation (for example, in terms of current or cumulative hours of participation). The review of the evidence points to the importance of consideration especially of exposure over multiple years. Secondary analyses will need to take steps to control for possible selection effects when examining the relationship of dosage to child outcomes, as previous research (e.g.,

NICHD ECCRN, 2006) suggests that family characteristics are related to extent of utilization of early care and education.

- Secondary analyses provide the potential for detailed consideration of how thresholds of quality should be operationalized. Emerging work is considering both whether there is evidence of a nonlinear relationship of quality and child outcomes, and of differences in the strength of the relationship in different portions of the quality range. However, there is an important discussion, and lack of resolution, regarding how to identify the point in the range to place the demarcation (for example, above and below which the strength of association between quality and child outcomes will be assessed). Possibilities include setting this point conceptually, through examination of the distribution of scores, or through consideration of the shape of the curve (using the inflection point in the curve) if the relationship of quality and outcomes is found to be nonlinear. It would be extremely helpful if secondary analyses built in methodological work contrasting findings when different approaches to setting the threshold are taken and working towards recommendations for analytic approach.
- We have noted the concern, articulated by Mashburn and Pianta (in press), that structural features of quality are rarely analyzed as part of more comprehensive models of how quality is related to child outcomes. Secondary analyses provide an important opportunity to replicate and extend previous analyses in which structural features of quality have been found to be related to child outcomes indirectly through process features of quality. Mashburn and Pianta (in press) also raise the possibility that structural features of quality may moderate the relationship between process quality and child outcomes. For example, the effects of supportive caregiver-child interactions maybe magnified when group size is smaller. This hypothesis has not, as yet, been tested empirically. The highest priority should be placed on examining the relationship between structural features and child outcomes as mediated through more proximal measures of quality, and extending previous consideration of this issue to include both interaction-specific and domain-specific measures of quality in models. A further goal might be to examine the hypothesis that structural measures of quality moderate the relationship between interaction- or domain-specific quality measures and child outcomes.
- We have noted some indications that children from families with greater demographic risk may benefit more from ECE that attains quality thresholds. We have also noted some evidence that classroom composition, such as number or proportion of children with special needs in the classroom, may also moderate the effects on children of participation in high quality care. It would be helpful to consider the potential to examine family demographic characteristics and classroom composition as moderators of the effects of quality that reaches specified levels and/or is sustained over time. It should be noted that examination of moderators as a priority will need to be balanced with other priorities in data analyses, as consideration of moderators will affect power.

C. Issues This Literature Review Suggests Should Be Considered in Developing Design Options for New Data Collection

- The design phase of the Q-DOT project will provide an important opportunity to address gaps in the extent to which existing datasets permit consideration of specific features of quality. It would be a particularly important contribution if the planned new data collection could incorporate and permit contrasts of quality measures as predictors of the most closely aligned child outcomes at all three levels of specificity with respect to

quality features: (1) global measures of quality, (2) interaction-specific measures of quality (including such measures as Instructional Support and Emotional Support in the CLASS), and (3) domain-specific measures of quality (focusing on environmental supports for development in such specific domains as health, social and emotional development, and early mathematics skills).

- It would be a valuable contribution if the design phase of the project could lay out a plan for assessing the reliability of both interaction-specific and domain-specific measures of quality. A member of the Technical Working Group for this project, Thomas Cook (personal communication, May, 2010) has suggested that domain-specific measures of quality may require observations on more occasions in order to reach reliable estimates of quality than interaction-specific or global measures. He notes that in elementary school classrooms, there is greater day-to-day variation in whether domain-specific content is introduced. It will be very important to assess what is required to reach stable estimates of domain-specific quality measures and to balance this against what is gained in terms of prediction to child outcomes with such measures. If a substantial increment in prediction to aligned child outcomes is found with domain-specific quality measures, this will need to be weighed against the cost of carrying out more observations if this is found to be necessary to attain reliable quality estimates with such measures. This cost-benefit analysis could be quite important if, for example, whole states were to consider incorporating domain-specific measures of quality into Quality Rating and Improvement Systems.
- The design phase of the Q-DOT project will also provide an opportunity to consider the highest priority for examination of the core constructs in the context of a rigorous experimental design. In the United States, families determine whether to use early care and education, and if so, the type, amount and quality. It is important to acknowledge that family decision making is constrained by economic circumstances, available options for early care and education, and information about options. It is also important to acknowledge that family decision making can be affected by public support for early care and education, which can increase financial resources and extend the range of available settings to include publicly supported programs. Overall, however, previous research indicates that family characteristics are related to patterns of utilization of early care and education, with findings suggesting that more advantaged parents tend to select higher quality early care and education, use more early care and education, and use higher quality settings (NICHD ECCRN, 2006). Given the evidence of selection effects, a high priority should be placed on the use of experimental designs. If these are not possible, the design phase of this project should carefully examine the options for well designed quasi-experimental approaches such as natural experiments or regression discontinuity designs, or use econometric methods such as fixed effects, to examine the issues of dosage, thresholds, and quality features in a way that more fully accounts for selection effects. The need for such approaches is especially important when factors such as attendance, level of overall quality, and exposure to specific quality practices are studied due to the role that parents play in determining how much children attend early care and education and whether they participate in programs with a focus on particular instructional areas.

- The present literature review points especially to the potential importance of joint consideration of the core constructs. Design work should include consideration of whether and how it might be possible to manipulate and evaluate, through experimental or quasi-experimental designs, the dosage of exposure to ECE that meets specific thresholds for quality on interaction-specific or domain-specific quality features.
- The design phase of this project will also provide an opportunity to consider ways to address key gaps in the characteristics of samples studied and in the distribution of quality in existing datasets. We have noted the relative lack of studies focusing on infants and toddlers in center-based settings as an impediment to reaching any clear conclusions in our review as to whether the patterns identified vary or are consistent across infancy/toddlerhood and preschool-age. It is very important to assess whether dosage, thresholds, and features function similarly for these age ranges. Consideration of this issue will require planning for adequate sample sizes in both age ranges between birth and school entry. This review also cautions that existing datasets may not have sufficient samples of low quality care to permit a full assessment of the nature of the relationship between quality and child outcomes, and whether this relationship is linear or nonlinear. The design phase of the Q-DOT project provides an important opportunity to develop strategies for sampling children across the full age range from birth to school entry, and sampling settings that span the full quality range.
- In the design phase of the Q-DOT project, it will be possible to give consideration to the dependent variables. The strength of association of quality and child outcomes rests in part on child outcome measures included in analyses: both the content and focus of the child outcomes and their reliability and validity. It will be possible in the design phase of the Q-DOT project to extend the consideration of child outcomes to potentially important outcomes not yet included in major ECE datasets available for secondary data analyses. Such outcomes might include, for example, measures of children's executive function. As in the discussion of using new data collection to both extend but also assess the strength of new quality measures, design work for new data collection might build in opportunities both to extend the range of child outcomes considered (thereby increasing the opportunities to assess alignment) and to examine the psychometric functioning of the measures.

REFERENCES

- Administration for Children and Families. (2003). Head Start FACES 2000: A Whole-Child Perspective on Program Performance. Office of Planning, Research and Evaluation. Washington, DC.
- Blau, D. M. (1999). Effect of child care characteristics on child development. *The Journal of Human Resources*, 34(4), 786-822.
- Broberg, A. G., Wessels, H., Lamb, M. E., & Hwang, C. P. (1997). Effects of day care on the development of cognitive abilities in 8-year-olds: A longitudinal study. *Developmental Psychology*, 33, 62-69.
- Burchinal, M. R., & Cryer, D. (2003). Diversity, child care quality, and developmental outcomes. *Early Childhood Research Quarterly*, 18, 401-426.
- Burchinal, M. R., Kainz, K., & Cai, Y. (in press). How well are our measures of quality predicting to child outcomes: A meta-analysis and coordinated analyses of data from large scale studies of early childhood settings. In M. Zaslow, I. Martinez-Beck, K. Tout & T. Halle (Eds.), *Measuring quality in early childhood settings*. Baltimore: Brookes Publishing.
- Burchinal, M. R., Kainz, K., Cai, Y., Tout, K., Zaslow, M., Martinez-Beck, I., & Rathgeb, C. (May 2009). Early care and education quality and child outcomes. Research-to-Policy Research-to-Practice Brief: OPRE Research-to-Policy Brief #1. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, US DHHS, and Child Trends.
- Burchinal, M. R., Peisner-Feinberg, E. S., 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. *Journal of Applied Developmental Science*, 4, 149-165.
- Burchinal, M. R., Vandergrift, N., Pianta, R., & 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.
- Campell, F. A., Pungello, E. P., Miller-Johnson, S., Burchinal, M. R., & Ramey, C. (2001). The development of cognitive and academic abilities: Growth curves from an early intervention educational experiment. *Developmental Psychology*, 37(2), 231-242.
- Campell, F. A., Ramey, C. T., Pungello, E. P., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adults outcomes from the Abecedarian project. *Applied Developmental Science*, 6(1), 42-57.
- 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.

- Early, D., Maxwell, K. L., Burchinal, M. R., Alva, S., Bender, R. H., Bryant, D., Cai, K., Clifford, R. M., Ebanks, C., Griffin, J. A., Henry, G. T., Howes, C., Iriondo-Perez, J., Jeon, H., Mashburn, A. J., Peisner-Feinberg, E. S., Pianta, R. C., Vandergrift, N., & Zill, N. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development, 78*(2), 558-580.
- Gallagher, P. A., & Lambert, R. G. (2006). Classroom quality, concentration of children with special needs, and child outcomes in Head Start. *Exceptional Children, 73*(1), 31-52.
- Halle, T., Vick Whittaker, J. E., & Anderson, R. (2010). *Quality in early childhood care and education settings: A compendium of measures, second edition*. Office of Planning, Administration for Children and Families, U.S. Department of Health and Human Services, Washington, DC: Child Trends.
- Harms, T., & Clifford, R. M. (1989). *The Family Day Care Rating Scale*. New York: Teachers College Press.
- Harms, T., Clifford, R. M., & Cryer, D. (2005). *Early Childhood Environmental Rating Scale-Revised*. New York: Teachers College Press.
- Harms, T., Cryer, D., & Clifford, R. M. (1990). *Infant/Toddler Environmental Rating Scale*. New York: Teachers College Press.
- Harms, T., Cryer, D., & Clifford, R. M. (2007). *Family Child Care Environmental Rating Scale-Revised*. New York: Teachers College Press.
- Harms, T., Jacobs, E. V., & White, D. R. (1996). *School Age Care Environment Rating Scale*. New York: Teachers College Press.
- Hill, J. L., Brooks-Gunn, J., & Waldfogel, J. (2003). Sustained effects of high participation in an early intervention for low-birth-weight premature infants. *Developmental Psychology, 39*(4), 730-744.
- Howes, C. (1997). Children's experience in center-based care as a function of teacher background and adult:child ratio. *Merrill-Palmer Quarterly, 43*, 404-425.
- Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., & Clifford, R. (2008). Ready to learn? Children's pre-academic achievement in pre-kindergarten programs. *Early Childhood Research Quarterly, 23*, 27-50.
- Howes, C., Phillips, D. A., & Whitebook, M. (1992). Thresholds of quality: Implications for the social development of children in center-based child care. *Child Development, 63*, 449-460.
- Hubbs-Tait, L., Culp, A. M., Huey, E., Culp, R., Starost, H., & Hare, C. (2002). Relation of Head Start attendance to children's cognitive and social outcomes: Moderation by family risk. *Early Childhood Research Quarterly, 17*(539-558).
- Hynes, K., & Habasevich-Brooks, R. (2008). The ups and downs of child care: Variations in child care quality and exposure across the early years. *Early Childhood Research Quarterly, 23*(559-574).
- Karoly, L. A., Kilburn, M. R., & Cannon, J. S. (2005). *Early childhood interventions: Proven results, future promises*. Santa Monica, CA: The RAND Corporation.

- Kontos, S., & Wilcox-Herzog, A. (1997). Influences on children's competence in early childhood classrooms. *Early Childhood Research Quarterly*, *12*, 247-262.
- Lambdin, D. J. (1996). Evidence of student attendance as an independent variable in education production functions. *The Journal of Educational Research*, *89*(3), 155-162.
- Loeb, S., Fuller, B., Kagan, S. L., & Carrol, B. (2004). Child care in poor communities: Early learning effects of type, quality, and stability. *Child Development*, *75*, 47-65.
- Mashburn, A. J., & Pianta, R. C. (in press). Opportunity in early education: Improving teacher-child interactions and child outcomes. In A. Reynolds, A. Rolnick, M. Englund & J. Temple (Eds.), *Cost Effective Programs in Children's First Decade: A Human Capital Integration*. New York: Cambridge University Press.
- Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarian, O. A., Bryant, D., Burchinal, M. R., Early, D., & Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, *79*(3), 732-749.
- McCartney, K. (1984). Effect of quality of day care environment on children's language development. *Developmental Psychology*, *20*, 244-260.
- McCartney, K., Burchinal, M. R., Clarke-Stewart, A., Bub, K. L., Owen, M. T., Belsky, J., & NICHD Early Child Care Research Network. (2010). Testing a series of causal propositions relating time in child care to children's externalizing behavior. *Developmental Psychology*, *46*(1), 1-17.
- McCartney, K., Scarr, S., Rocheleau, A., Phillips, D., Abbott-Shim, M., Eisenberg, M., Keefe, N., Rosenthal, S., & Ruh, J. (1997). Teacher-child interaction and child-care auspices as predictors of social outcomes in infants, toddlers, and preschoolers. *Merrill-Palmer Quarterly*, *43*(426-450).
- Nelson, G., Westhues, A., & MacLeod, A. (2003). A meta-analysis of longitudinal research on preschool prevention programs for children [electronic version]. *Prevention and Treatment*, *6* from <http://journals.apa.org/prevention/>.
- NICHD Early Child Care Research Network. (1998). Early child care and self-control, compliance, and problem behaviors at twenty-four and thirty-six months. *Child Development*, *69*(4), 1145-1170.
- NICHD Early Child Care Research Network. (2000). The relation of child care to cognitive and language development. *Child Development*, *71*(960-980).
- 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*, 199-206.
- NICHD Early Child Care Research Network. (2006). Child-care effect sizes for the NICHD Study of Early Child Care and Youth Development. *American Psychologist*, *61*(2), 99-116.
- NICHD Early Child Care Research Network, & Duncan, G. (2003). Does quality of child care affect child outcomes at age 4 1/2? *Developmental Psychology*, *39*(3), 451-469.

- Owen, M. T., Klausli, J. F., Mata-Otero, A., & Caughy, M. O. (2008). Relationship-focused child care practices: Quality of care and child outcomes for children in poverty. *Early Education and Development, 19*(2), 302-329.
- Peisner-Feinberg, E. S., & Burchinal, M. R. (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. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., & Yazejian, N. (2001). The relation of preschool child-care quality to children's cognitive and social development trajectories through second grade. *Child Development, 72*, 1534-1553.
- Pianta, R., La Paro, K., & Hamre, B. (2007). *Classroom Assessment Scoring System*. Baltimore: Brookes Publishing.
- Schliecker, E., White, D. R., & Jacobs, E. (1991). The role of day care quality in the prediction of children's vocabulary. *Canadian Journal of Behavioral Science 23*, 12-24.
- Swenson-Klatt, D. (2008). Using quality measures in Minnesota child care quality initiatives. Paper presented at the 2008 Annual Meeting of CCDF State and Territory Administrators, Washington, DC.
- Sylva, K., Sirai-Blatchford, I., & Taggart, B. (2006). *Assessing quality in the early years: The Early Childhood Environment Rating Scale-Extended*. London: Trentham Books.
- Tout, K., Zaslow, M., Halle, T., & Forry, N. (2009). Issues for the next decade of quality rating and improvement systems. Issue Brief: OPRE Issue Brief #3. Washington DC: Office of Planning, Research and Evaluation, Administration for Children and Families, US DHHS, and Child Trends.
- Tran, H., & Weinraub, M. (2006). Quality, stability, and multiplicity in nonmaternal child care arrangements during the first 15 months of life. *Developmental Psychology, 42*(3), 566-582.
- Vandell, D., Belsky, J., Burchinal, M., Steinberg, L., Vandergrift, N., & NICHD Early Child Care Research Network. (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.
- Volling, B. L., & Feagans, L. V. (1995). Infant day care and children's social competence. *Infant Behavior and Development, 18*(2), 177-188.
- Votruba-Drzal, E., Coley, R. L., & Chase-Lansdale, P. L. (2004). Child care and low-income children's development: Direct and moderated effects. *Child Development, 75*, 296-312.
- Whitebook, M., Howes, C., & Phillips, D. (1990). Who cares? Child care teachers and the quality of care in America. Final Report: The National Child Care Staffing Study. Oakland, CA: Child Care Employee Project.
- Witte, A. D., & Queralt, M. (2004). What happens when child care inspections and complaints are made available on the internet? No. 10227 NBER Working Paper Cambridge, MA: National Bureau of Economic Research.

Yazejian, N., & Bryant, D. M. (2009). Promising early returns: Educare implementation study data, March 2009. Chapel Hill: FPG Child Development Institute, UNC-CH.

Zaslow, M., Anderson, R., Redd, Z. Wessel, J., Tarullo, L. & Burchinal, M. (2010). Quality Dosage, Thresholds, and Features in Early Childhood Settings: Literature Review Tables, OPRE 2011-5a. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Zaslow, M., Martinez-Beck, I., Tout, K., & Halle, T. (Eds.). (forthcoming). *Measuring quality in early childhood settings*. Baltimore: Brookes Publishing.

