CONGRUENCE BETWEEN PARENTS' AND EARLY CHILDHOOD EDUCATORS' PERSPECTIVES ABOUT CHILD DEVELOPMENT USING A MULTI-DIMENSIONAL SCREENING INSTRUMENT

By

CATHLEEN C. PASIA

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2013

ACKNOWLEDGEMENTS

Over the past six years, I have received support and encouragement from a great number of people. Dr. Patricia Snyder has been an outstanding mentor, who always believed in me when I did not, and pushed me further than I ever thought I could grow. Her guidance has made this a thoughtful and rewarding journey. To Dr. Tina Smith-Bonahue and Dr. Nancy Waldron, who not only helped to build the foundation for the practitioner I am today, but serve as models for connecting research and practice. This study would not have been possible without the support of Dr. Pamela Pallas. I also thank Dr. James Algina and Dr. Cynthia Garvan for your valuable statistical advice and guidance and to Dr. Fonda Davis Eyler for supporting me without hesitation.

I could not have accomplished the monumental feat of data collection without the assistance of the Center for Excellence in Early Childhood Studies and the personnel affiliated with Dr. Snyder's research lab. I appreciated your welcoming spirit and ongoing patience. I would like to thank the early childhood educators and the parents for generously taking the time out of their busy days to participate my study. Also, thank you to Jamie Runyons for helping me with data entry.

To Tara McLaughlin, Salih Rakap, and Crystal Crowe Bishop, I was fortunate to share this journey with all of you. I thank you for all the hours you spent listening to me talk about my research, reassuring me of my potential, and most of all, motivating me to work harder. I also could not have made it through this journey without my cohort working alongside me: Susan Craft, Jenny Heretick, Stacey Rice, Katrina Moore, Sally Moore, Suzie Long, and Angela Dobbins. I would like to thank Dieter Schaefer for always believing that I could do this. Finally, I have nothing but gratitude for my parents, who led by example through their hard work and dedication to their careers and family.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF DEFINITIONS

Caregiver	Any adult who spends extended time with a child. The adult may interact with the child within a home or child care setting.
Congruence	Agreement or correspondence between the judgments of two individuals as reflected in rating scale scores or observational codes (Dinnebeil & Rule, 1994; Snyder, Thompson, & Sexton, 1992).
Convergence	The process of synthesizing a wide base of information from multiple informants using the same or similar instrument in order to facilitate collaboration and teamwork (Bagnato 2007; Suen, Logan, Neisworth, & Bagnato, 1995).
Developmental screening	A form of assessment designed to be brief and used to help inform decisions about a child's developmental status relative to same-age peers as well as to identify if additional assessment, monitoring, or targeted support or intervention might be needed (Brassard & Boehm, 2007; Lichtenstein & Ireton, 1984; Nagle, 2007).
Early education and care center	A nonresidential facility that provides care and early learning experiences, often in classroom settings, for young children between the ages of 6 weeks through 5 years.
Early childhood educator	Adults who are employed by an early education and care center to provide care and early learning experiences for young children. The adults may be trained, certified, or licensed to provide care and early learning experiences for young children.
Examiner	A person trained to administer an assessment instrument and who does not interact regularly with the child.
Informant	An adult who completes a developmental screening instrument.

Microsystem	The immediate setting(s) in which a child develops and grows and is compromised of "a pattern of activities, roles, and interpersonal relations" (Bronfenbrenner, 1979, p. 22).
Mesosystem	Formed by the interrelationships among microsystems (Bronfenbrenner, 1979, p. 22).
Multi-dimensional developmental screening instrument	A type of assessment instrument designed to provide general information about child development in different domains, often including fine and gross motor, language and communication, problem-solving/adaptive behavior, and personal-social (Glascoe, 2010; Neisworth & Bagnato, 2005).
Parent	Refers to an adult who is a natural, adoptive, or foster parent of the child or is designated as their guardian. The adult is legally responsible for the child's welfare and may live with the child or may provide assistance for the child's ongoing primary care (Individuals with Disabilities Education Act [IDEA], 20 U.S.C. 1400 § 602(23), 2004).
Professional	A person who has expertise or training in the disciplines of medicine, psychology, education, or a related social science discipline.
Screening	An activity or procedure involving a brief sampling of behavior for the purpose of providing general information about whether or not additional diagnosis or assessment is needed (Harrington, 1984; Lichtenstein & Ireton, 1984; Scott & Carran, 1989).

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

CONGRUENCE BETWEEN PARENTS' AND EARLY CHILDHOOD EDUCATORS' PERSPECTIVES ABOUT CHILD DEVELOPMENT USING A MULTI-DIMENSIONAL SCREENING INSTRUMENT

By

Cathleen Pasia

August 2013

Chair: Tina Smith-Bonahue
Co-chair: Patricia Snyder
Major: School Psychology

Given the significant time many children spend with different caregivers, a need exists to identify efficient and effective strategies to support the reciprocal exchange of information about children's abilities, needs, and preferences. Developmental screening instruments are being used with increasing frequency in early education and care settings. While the purpose of developmental screening is to inform decisions about the need for additional assessment, supports, or services, screening might have an added value of serving as a method to gather and compare perspectives about child development from multiple caregivers who interact with the child in different contexts.

Previous research suggests that parent-completed developmental screening instruments are a time- and cost-effective approach that improves decision making, however, the methodology applied to examine the congruence between parent and professional perspectives using multi-dimensional developmental screening instruments limits the interpretations that can be made. The primary aim of the present study was to explore if parents and early childhood educators (ECEs) who spend significant time with young children shared similar perspectives about child

development when they each completed the same multi-dimensional developmental screening instrument. A secondary aim was to examine whether or not congruence varied based on the developmental domain assessed, child age, or length of relationship between children and ECEs.

Parent-ECE pairs (n = 126) were recruited from two university-based early education and care centers. Each member of the pair separately completed the same multi-dimensional developmental screening instrument for a child. Data obtained from parents and ECEs were analyzed descriptively by examining item agreement, classification agreement, and patterns of response selection. Exploratory analyses were conducted by examining intraclass correlation coefficients and beta regression. Parents and ECEs had highly similar perspectives as measured by overall item and classification agreement. Parent-ECE item agreement varied across developmental domains and age cohort, but variation was not associated with significant differences across informants on referral decisions. Data gathered from comparing similarities and differences in ratings on developmental screening instruments suggest the potential for the completion of routine developmental screenings to serve a dual-purpose of early identification and a foundation for rich discussion about child development between significant caregivers.

CHAPTER 1
INTRODUCTION

Statement of the Problem

Over the past several decades, the number of women in the workforce has been steadily increasing and current estimates are that 64.2% of women with children under the age of 6 years are employed (United States Department of Labor, 2011). As a result, children are spending an increasing amount of time in contexts outside of their home. The Children's Defense Fund (2005) estimated 12 million infants, toddlers and preschoolers are cared for by adults other than their parents, representing about half of all children in that age group. Of those 12 million children, approximately one-fifth are reported to spend time in early education and care centers. Children can enter an early education and care center as young as 6 weeks of age, for as many as 40 hours per week. This translates to almost a quarter of the 168 hours in a week being spent with early childhood educators (ECEs). These data suggest that many children are having extended contact with at least two types of caregivers: parents and ECEs. When primary caregivers share perspectives about a child's development, learning, and behavior within and across contexts, it helps each caregiver understand the child's unique abilities, needs, and preferences (Neisworth & Bagnato, 2005). When caregivers from different settings have shared understandings of children's abilities, needs, and preferences within and across contexts and people, they are better equipped to support children's development and learning (Bronfenbrenner, 1979).

Perspectives about children's development can be gathered and shared in formal and informal ways. It can be difficult for parents and ECEs to integrate regular communication about children's development as well as their abilities, needs, and preferences into their typical interaction patterns (Bailey & Wolery, 1992). One possible solution to this issue is to take

advantage of existing practices in the early education and care program and incorporate strategies that actively include parents and their perspectives about their child's development, learning, and behavior. Children are frequently screened throughout the early childhood years to determine the presence of early indications for delays in development and to identify children who might be in need of additional support (National Research Council, 2008). Although developmental screening has historically occurred in pediatric settings, screening is increasingly being conducted in early education and care programs (National Research Council, 2008). Periodic screenings provide an opportunity to gather and share perspectives from all caregivers, including parents, about children's development, learning, and behavior.

Although early childhood professional organizations recommend early education and care programs use multiple sources of information during the assessment process (Division for Early Childhood [DEC], 2007; National Association for the Education of Young Children [NAEYC], 2012), there is some flexibility for programs to determine how this is accomplished. Programs can gather information from parents using either informal or formal methods (NAEYC, 2012), leading to wide variability in the amount and quality of parental involvement. One common hesitation around incorporating parental perspectives about child development, learning, and behavior is the belief held by many professionals that parents are not reliable informants (Sheehan, 1988; Shelton, 1989). This belief persists despite promising evidence that shows when parents are provided opportunities to share perspectives about their child's development, learning, and behavior in the same ways as professionals using the same instruments there is not only convergence between parents and professionals but often high levels of congruence (Bricker & Squires, 1989b; Bricker, Squires, Kaminski, & Mounts, 1988; Snyder, Thompson, & Sexton, 1992; Squires & Bricker, 1991; Squires, Bricker, & Potter, 1997). A needs exists to examine

further issues related to congruence of parents' and ECEs' perspectives about children's development, learning, and behavior, particularly because early education and care programs are routinely using multi-dimensional development screening instruments as required by program policies and procedures (NAEYC, 2008, 2012). Given the significant time many young children spend in early care and education environments in addition to their home environments, it is important to examine further what differences, if any, exist between parents' and professionals' perspectives about children's development, learning, and behavior. Explicit opportunities to uncover and discuss consistency or variation in perspectives would be useful for ensuring young children experience nurturing and responsive relationships to facilitate their development and learning within and across contexts.

The purpose of the present study was to compare the perspectives of two primary caregivers (i.e., parent and ECE) about the development of a young child by having each of these caregivers independently complete the same multi-dimensional developmental screening instrument. In the remainder of this chapter, the conceptual framework and a definition for developmental screening are presented. In addition, the context for and significance of, the study is discussed. In chapter two, a review of the developmental screening literature is conducted, including studies focused explicitly on examining the convergence and congruence of parental and professional perspectives about child development using multi-dimensional developmental screening instruments. The methods used in the study are described in chapter three. In chapter four, findings are presented and chapter five includes discussion and implications of study findings for future research and practice.

Conceptual Framework

According to Bronfenbrenner's ecological systems model, development results from the interplay between an individual and the immediate settings in which the developing person

spends significant time (Bronfenbrenner, 1979). The immediate setting(s) in which a child develops and grows is called a microsystem and is compromised of "a pattern of activities, roles, and interpersonal relations" (Bronfenbrenner, 1979, p. 22). A child's home is typically the primary microsystem, but children who spend significant time in other settings, such as early education and care centers, experience additional microsystems. The transactional relationships between or among microsystems exert significant influences on the processes underlying human development and learning. Interrelationships among microsystems form a mesosystem.

The quality of the relationships that occur within the mesosystem is critical to children's development and learning (Bronfenbrenner, 1974). Mesosystems are said to be "multiply linked" when there is more than one person who actively participates in both settings (Bronfenbrenner, 1979). On the contrary, mesosystems that are connected only by the presence of the child in both settings are described as being "weakly linked." Within an ecological systems framework, children's developmental potential is purportedly enhanced through multiply linked mesosystems. The richness of a mesosystem is reflected in the number and quality of connections between the adults (i.e., parents and ECEs) from two microsystems (i.e., home and early education and care centers; Garbarino, 1990). When strong connections are formed through open communication channels between adults, children's development is enhanced because information, advice, and experience relevant to one setting are made available to the other (Bronfenbrenner, 1979). Similarly, open communication supports a shared understanding about children's development, abilities, needs, and preferences. Within a multiply linked mesosystem, there is a compatibility of expectations and activities for children. Continuity across settings contributes to the development of mutual goals, including mutual goals related to development,

learning, and behavior, thus increasing the frequency of opportunities for a child to experience developmentally enhancing learning opportunities across different contexts and people.

Mesosystems are created by the sharing of perspectives across settings, which can take a variety of forms and vary in the direction of flow (Bronfenbrenner, 1979). Three general types of information communicated between parents and ECEs are information about the child's day, descriptions of the child's behavior, and advice intended to provide adult support (Owen, Ware, & Barfoot, 2000). Parents and ECEs can share perspectives through the verbal exchange of information at arrival or pick-up, written exchanges through daily journals or behavior report cards, or visual exchanges through sample materials or weekly photographs (Bailey & Wolery, 1992). Developmental screening has potential to serve as a method to actively share information about children's development and learning across settings. For many early education and care centers, this information is already being gathered as part of ongoing decision-making about children's developmental status and potential need for additional support (NAEYC, 2008). By obtaining parent's perspectives about children's development at home, ECEs are better equipped to enhance children's classroom experiences because of the additional insight about the child's behavior and experiences in other settings (Owen et al., 2000). Likewise, ECEs can provide similar information to parents. This reciprocal flow of communication can either form or enhance a multiply linked mesosystem.

Studies focused on examining the congruence of parental and early childhood educator perspectives about child development can help inform how information should be gathered, shared, and used to form or enhance a multiply linked mesosystem. Although convergence and congruence are not always necessary or expected, knowing the similarities and differences in perspectives can serve as a starting point for conversations about why children might exhibit

different developmental skills in different contexts or ways all caregivers can support development, learning, and behavior across contexts. Developmental screening can serve as one method for strengthening mesosystems by systematically gathering, comparing, and sharing perspectives about children's abilities and needs.

Definition of Developmental Screening

Broadly defined, screening is an activity or procedure involving a brief sampling of behavior (Harrington, 1984; Lichtenstein & Ireton, 1984; Scott & Carran, 1989). Other authors specify the procedure or activity as separating a population into two groups: those at risk for later problems and those who are not (Barnett, 1984; Brooks-Gunn & Lewis, 1983; Glascoe, 2005; Green & Scriven, 1980; Kochanek, 1993; Meisels & Wasik, 1990; Sheldrick & Perrin, 2009). Both definitions focus on the procedure of screening, but more contemporary definitions of screening specify its role as an entry point into a larger decision-making process. In the latter type of definition, screening is described as a form of assessment to determine a child's developmental status relative to peers of the same age and to identify if additional assessment, monitoring, or targeted support or intervention might be needed (Brassard & Boehm, 2007; Harrington, 1984; Lichtenstein & Ireton, 1984; McLean & McCormick, 1993; Nagle, 2007; Taylor, 1993; Zehrbach, 1975).

Screening can occur at any age to aid in the detection or prevention of many types of illnesses or conditions (Lichenstein & Ireton, 1984). Developmental screening refers to the administration of a standardized assessment instrument or tool focused on sampling key developmental milestones or skills for the purposes of aiding in the identification of children who are in need of additional supports, evaluations, or services (American Academy of Pediatrics [AAP], 2001; Council on Children with Disabilities, Section on Developmental Behavioral Pediatrics, Bright Futures Steering Committee, & Medical Home Initiatives for

Children with Special Needs Project Advisory Committee, 2006). By definition, developmental screening alone should not result in a diagnosis or treatment plan. It is intended to identify areas in which a child's development may differ from the documented trajectory of same-age peers and should only be used to guide decisions as to whether a child may benefit from further evaluation, closer monitoring, or additional support or intervention.

Children are likely to encounter developmental screening multiple times before the age of 6 years as a result of recognition by professional organizations of the importance of early identification (Council on Children with Disabilities et al., 2006; NAEYC, 2012; National Association of School Psychologists [NASP], 2009; Neisworth & Bagnato, 2005). Consequently, developmental screening can serve as a method for gathering and exchanging information from multiple caregivers to determine children's developmental abilities and needs across various contexts (i.e., microsystems) and to form or enhance mesosystems. In addition to its applicability as a multi-informant tool, developmental screening instruments can serve as a multi-dimensional assessment. Practitioners might choose to use a multi-dimensional developmental screening instrument or combine multiple instruments to screen across different developmental domains. For the purpose of the present study, the focus was multi-dimensional developmental screening instruments because of legislative requirements and policy recommendations that children be assessed in five developmental areas. For example, the AAP and Division for Early Childhood (DEC) recommend that developmental screening instruments should be broad and examine all domains of development: fine and gross motor skills, language and communication, problem solving/adaptive behavior, and personal-social skills (Council on Children with Disabilities et al., 2006; Glascoe, 2010; Neisworth & Bagnato, 2005).

For most developmental screening instruments two methods are used to gather information: direct elicitation or parent/caregiver report (Boan, Aydlett, & Multunas, 2007; National Research Council, 2008). Direct elicitation involves the observation of child actions in response to the presentation of stimuli. When developmental screening instruments are administered through direct elicitation, examiners should be trained to administer the instrument reliably (Bradley-Johnson & Johnson, 2007; Brassard & Boehm, 2007; Kenny & Culbertson, 1993; Paget & Nagle, 1986). The second method is to gather information from a parent/caregiver about children's development either through a questionnaire or interview format (Boan et al., 2007). The second method can be used with any caregiver, including an ECE, provided the caregiver has spent enough time with the child to be able to answer all of the items. Most screening instruments specify approximately how much time a caregiver should spend with a child before the caregiver completes a questionnaire or participates in an interview.

Multi-dimensional developmental screening instruments can be administered as part of various assessment frameworks. Two common frameworks are developmental surveillance and multi-tiered screening. Developmental surveillance is most likely to occur in the pediatric setting (First & Palfrey, 1994; Glascoe, 2005) and is a "flexible, longitudinal, continuous, and cumulative process" for the purpose of gathering information about parents' concerns, children's developmental history, and risk and protective factors and integrate them with observations, physical examinations, and laboratory assessments. (Council on Children with Disabilities et al., 2006, p. 407). In an alternative framework, screening occurs within a multi-level or multi-tiered approach (Brooks-Gunn & Lewis, 1983; Dunbar & Reed, 1999; Jackson & Needleman, 2007; Ikeda, Neessen, & Witt, 2008; Teska & Stoneburner, 1980; Thurlow & Gilman, 1999; Zehrbach, 1975). One variant of the multi-tiered framework includes a pre-screening procedure intended to

limit the number of children for whom a multi-dimensional developmental screening is administered. In a second variant of the multi-tiered framework, developmental screening could be administered to all children as a universal screener and followed by targeted screeners or additional supports (Ikeda et al., 2008). Of these two frameworks, a multi-tiered framework for screening may be more appropriate for implementation of a multi-dimensional developmental screening instrument in an early education and care center.

Context for the Study

Three major pieces of legislation have mandated screening as a method of early identification of children in need of intervention. In the medical field, the Early and Periodic Screening, Diagnosis and Treatment (EPSDT) program was established as a part of a 1967 provision to the Medicaid Act to target children experiencing environmental risk due to low-income backgrounds (Frankenburg & North, 1974). In the field of education, Child Find was initially established in 1974 through the Elementary and Secondary Act (ESEA) as an effort to identify children with disabilities from birth through 21 years (Paget & Nagle, 1986). As Child Find and other legislative mandates have undergone successive reauthorizations other organizations have disseminated additional position statements and recommendations supporting the use of developmental screening: American Academy of Pediatrics (AAP, 2001; Council on Children with Disabilities et al., 2006), National Association for the Education of Young Children (2012), and National Association of School Psychologists (2009). Additionally, the Head Start Act (Improving Head Start for School Readiness Act, 42 U.S.C. § 1304.20, 2007) mandates developmental screening upon 45 days of entry into programs and NAEYC (2008, 2012) requires programs to conduct ongoing screening and assessment of child progress to obtain accreditation status.

Although developmental screening typically occurs within the pediatric setting because it serves as a primary context for professional contact with infants and toddlers (National Research Council, 2008), the increase in the number of children entering early education and care settings improves young children's access to screening (Children's Defense Fund, 2005). In many early education and care programs, regular and systematic developmental screening is being implemented to meet accreditation requirements (NAEYC, 2008, 2012). When conducting assessment in early childhood, recommended practices include involving parents as active team members in screening and subsequent assessment processes and procedures (Bagnato, 2007; Brassard & Boehm, 2007; Kochanek, 1993; Ireton, 1990; Moore, 1978; Neisworth & Bagnato, 2005; Snyder, McLaughlin, & McLean, in press; Thurlow & Gilman, 1999). As developmental screening becomes a universal practice across different contexts and settings, it fulfills multiple functions by helping to (a) identify children at-risk, (b) monitor children's development, (c) incorporate parent input, and (d) support sharing of perspectives about child development across caregivers and settings

Significance of the Study

Reviews conducted by Snyder, Thompson, and Sexton (1992) and Dinnebeil and Rule (1994) have supported the validity of eliciting parental perspectives about children's developmental status. These reviews only included studies that examined relationships between parent report (typically elicited through a questionnaire or interview) with a comprehensive, developmental assessment completed by a professional. In both of these reviews, the term "congruence" was used to describe the agreement or correspondence between parent and professional reports of child developmental status. To date, there has not been a published review of studies examining congruence when parents and early childhood educators complete multi-dimensional developmental screening assessment instruments. Through the use of systematic

search procedures described in detail in Chapter 2, 10 studies were identified that compared parent and professional perspectives about child development using multi-dimensional developmental screening instruments. Findings from these studies that are significant for the present study are summarized below.

Length of Relationship between Professionals and Children

Eight of the 10 studies compared parents' scores on multi-dimensional developmental screening instruments to those of professionals who did not typically interact with the child. The two remaining studies involved comparing parents' scores on multi-dimensional developmental screening instruments to day care providers' scores (Coghlan, Kiing, & Wake, 2003; Frankenburg, Fandal, &Thornton, 1987; Frankenburg, van Doornick, Liddell, & Dick, 1976). Across all ten studies, no descriptions were provided about the length of the time that the professional or day care provider knew or interacted with the child prior to completing the developmental screening instrument. Lack of congruence might be due to the limited opportunities that professionals have to observe or elicit skills or behaviors. A need exists for additional studies that examine congruence of parents' and ECEs' perspectives about child developmental status using a multi-dimensional developmental screening instrument when completed by two individuals who interact often with a child in the home or early education and care microsystems.

Comparing Parent and Professional Responses

Across all 10 studies, moderate to high levels of congruence were found between parent and professional responses on multi-dimensional developmental screening instruments. Often reported as interrater score reliability, congruence was examined using the following methods: (a) agreement on the classification of children related to those who require additional evaluation and those who do not; (b) item-level agreement; or (c) statistical evaluation of score

agreements/differences or relationships (e.g., Pearson-product moment correlation coefficients, F-tests, and kappa statistics).

Eight of the 10 studies calculated percent agreement between parents' and professionals' overall classification of children (e.g., not at risk, at risk). Although this method provides a gross estimate of congruence, it is limited with respect to similarities or differences in perspectives across the two respondents on specific items. Likewise, calculating agreement by using an overall average masks potential differences by developmental domain. Chiu and DiMarco (2010) and Coghlan et al. (2003) demonstrated that parent and professional responses might differ depending on the skill being addressed. Both of these studies showed high congruence across respondents for gross motor skills and low agreement across the same respondents for language skills. It is plausible that agreement or score congruence on items included on multi-dimensional developmental screening instruments might also differ depending on the child's age or the amount of time the professional has known the child. None of the reviewed studies, however, explored these two variables, suggesting a need for additional studies focused on examining correlates of congruence.

In seven studies, parent and professional perspectives were gathered by having both informants use the same instruments. When developing the Denver Prescreening Developmental Questionnaire (PDQ) and its revised edition, the R-PDQ, authors compared parent perspectives to those of day care providers and examiners, but only day care providers completed the same screening instrument as parents (Frankenburg et al., 1987; Frankenburg, van Doornick, Liddell, & Dick, 1976). Use of the same instrument to gather and quantify perspectives eliminates potential measurement bias that might result from using two different instruments. A need exists for more studies that gather perspectives of parents and early childhood educators about child

development using the same multi-dimensional developmental screening instrument. In addition, individuals who regularly interact with the child should complete these instruments. Employing these procedures would eliminate at least one form of measurement bias, potentially strengthen the inferences that could be made about congruence, and adhere to recommended practices in early childhood assessment (Bagnato, 2007; Neisworth & Bagnato, 2005).

Developmental screening instruments used in empirical investigations of congruence should meet standard psychometric criteria. The primary focus in all but two of the 10 studies was an examination of interrater score reliability for a new or revised developmental screening instrument. Interrater score reliability was calculated by comparing scores of parents and professionals. These studies provided some information about similarities and differences in parents' and professionals' perspectives about children's development, however, interrater score reliability is viewed by some researchers as a different construct than congruence. Conventional statistical evaluations of interrater score reliability operate under the assumption that informants need to be interchangeable and are randomly selected (Suen, Logan, Neisworth, & Bagnato, 1995). Interrater score reliability is a reflection of the stability of a score across informants and demonstrates that similar scores would be obtained independent of the informant used. When informants are not assumed to be equivalent, perhaps as in the case of parents and professionals, assumptions related to interrater score reliability might not apply.

Directions for Research

A typical misconception of the goal of research focused on congruence is to provide evidence for parental estimates of child ability being similar to professionals (see discussions by Dinnebeil & Rule, 1994; Snyder et al., 1992). These authors and other proponents for multi-informant approaches to the assessment process recognize the need to incorporate information from both practitioners and parents. Striving for exact agreement in judgments or ratings would

negate the benefits of including multiple perspectives in assessment (Suen et al., 1995). Likewise, discussing parent-professional congruence results with the terminology “parents’ overestimations” or “professionals’ underestimations” devalues both roles as team members (Sheehan, 1988). Parents and professionals bring different sets of frameworks and expertise to the assessment process and each perspective is essential to making decisions as a team. Thus, while the studies reviewed were interested in levels of agreement, it may be more appropriate to investigate congruence, similarities in estimations that preserve the richness of each perspective by removing the expectation of absolute agreement (Dinnebeil & Rule, 1994; Snyder et al., 1992; Neisworth & Fewell, 1990). The interest in investigating congruence is to understand the relationship between the judgments or ratings of parents and professionals, specifically one who also knows the child (e.g., ECE), and how this is changes with developmental domain measured, child age, and length of relationship between ECE and child. Exploring these relationships may provide valuable information for the formation of multiply linked mesosystems. Ultimately, if congruence is explored across multiple microsystems so that several adults (e.g., parents, ECE providers, and other professionals) are sharing information, then convergence might occur. Convergence is the process of synthesizing a wide base of information from multiple informants using the same instrument in order to facilitate collaboration and teamwork (Bagnato 2007; Suen et al., 1995). Understanding the nature of the relationship between different perspectives could inform the way that information is gathered and shared across settings in which children learn and grow.

Purpose of the Present Study

Previous researchers have shown support for eliciting parent report about child development (Bricker & Squires, 1989b; Bricker et al., 1988; Snyder et al., 1992; Squires & Bricker, 1991; Squires et al., 1997). Few empirical studies have investigated parent and

professional congruence as measured by a multi-dimensional developmental screening instrument completed in the same way by each informant. As will be discussed further in Chapter 2, research that employs defensible methods for gathering parent and professional report, measuring child development, and analyzing correspondence between parent and professional judgments is needed. To address this gap in the literature, the following research questions will be addressed in the present study:

1. When using the same multi-dimensional developmental screening instrument, what is overall congruence between two informants who interact with the child regularly?

2. Does congruence between parents and early childhood educators vary based on child age cohort?

3. Does congruence between parents and early childhood educators vary depending on the child developmental domain assessed (i.e., communication, gross motor, fine motor, problem solving, personal-social)?

4. Are there differences in the pattern of response selection across informant types (i.e., parents and ECEs)?

5. Does child age and length of relationship with the early childhood educator predict item-level congruence between parents and ECE?

In summary, the increasing number of women entering the workforce has resulted in children spending more time in contexts with other caregivers. Children may spend up to 40 hours a week being cared for in other settings, such as an early education and care program. Given that young children spend significant time in at least two microsystems, the reciprocal exchange of information about child development and learning could facilitate connections between caregivers (i.e., parents and ECEs) and ensure each is informed about children's development, learning, and behavior within and across contexts. Open communication enhances caregivers' ability to support children's development and provides opportunities for children to practice skills across contexts and people.

Developmental screening instruments are increasingly being used in early education and care settings as a result of recommendations by prominent organizations in the medical, educational, and early childhood fields (AAP, 2001; Council on Children with Disabilities et al., 2006; NAEYC, 2012; NASP, 2009; National Research Council, 2008). While the primary purpose of developmental screening instruments is to inform decisions about whether children need additional supports or services, they potentially have an added value of serving as a formal method of gathering and comparing perspectives of primary caregivers from different settings. When microsystems are connected through frequent and high quality connections between adults, a multiply linked mesosystem is formed and children's developmental potential is enhanced (Bronfenbrenner, 1979).

CHAPTER 2
LITERATURE REVIEW

To provide a rationale for the need for the present study and situate it within the larger body of literature focused on developmental screening, a systematic descriptive review of the literature was conducted. This review explored three main topics: issues and trends in developmental screening, examination of studies that involved parents and professionals in developmental screening, and empirical studies focused on the caregiver-completed developmental screening instrument used in the present study.

Identification of Articles for Review

General Procedure

To address each of the topic areas outlined above, three separate electronic searches were conducted. The same procedure was applied to each search. EBSCOhost was used to search simultaneously Academic Search Premier, PsycInfo, and Professional Development Collection databases. Acceptable sources were limited to periodicals and books. The age of participants was limited to infancy (2-23 months), preschool (3-5 years), and childhood (birth-12 years). The results of each search were exported to RefWorks™ to identify duplicate sources. After the removal of duplicate sources, the titles and abstracts of each source were reviewed for specific criteria related to each topic. If an article met all specified criteria, the reference list was used to conduct an ancestral search.

Topic One: Issues and Trends in the Developmental Screening Literature

The following search terms were entered into the electronic databases: (a) developmental screen*, screen* and young children; (b) screen* and infants; screen* and preschool*; and (c) screen* and "developmental delay." The titles and abstracts of each of the 3,204 sources were reviewed for the presence of five criteria: (a) written in the United States or sampled participants

exclusively from this country; (b) categorized as either a research study, position statement, book chapter, or book; (c) primarily focused on screening; (d) described a multi-dimensional developmental screening instrument that measured at least four domains of development; and (e) was not an investigation of the psychometrics of one or more screening instruments. Of the 3,204 abstracts reviewed, 93 articles/books/book chapters met criteria for inclusion in this review.

Topic Two: Multi-Dimensional Developmental Screening Conducted by Parents and Professionals

The following search terms were entered into EBSCOhost: (a) developmental screening tests and parents; and (b) congruence* and development*. To be included in the present review, the articles from peer-reviewed journals had to meet all of the following criteria: (a) a professional (e.g., examiner, pediatrician, nurse, teacher, psychologist) completed a developmental screening instrument; (b) a parent/caregiver completed a developmental screening instrument by report or interview; (c) instruments used by professionals and parents were multi-dimensional developmental screening instruments, meaning that items sample at least four domains of development (e.g., motor, cognitive, communication, social, adaptive); and (d) the selected multi-dimensional developmental screening instrument(s) was used in its English version and had psychometric data reported in either the published report or the test manual. The titles and abstracts of these articles were reviewed for pre-specified inclusion criteria. A total of 10 articles were identified for this topic area.

Topic Three: Empirical Studies on Ages and Stages Questionnaires

The Ages and Stages Questionnaires - Third Edition (ASQ-3) was created for use in screening or monitoring programs in settings such as health clinics, pediatric practices, early education and care settings, home visiting programs, and community-based programs (Squires, Twombly, Bricker, & Potter, 2009). The purpose of the ASQ-3 is to gather information about

child development from individuals who interact regularly with the child. It is used in many early education and care programs as a universal screener, including the program in which the present study was conducted.

To identify articles related to the ASQ, the following search terms were entered in EBSCOhost: (a) “Infant Monitoring Questionnaires” or “Infant Monitoring System” or “Ages and Stages Questionnaires” and (b) Squires, Jane or Bricker, Diane. Articles were screened for four criteria: (a) the source must be available in English, (b) the source reports on empirical data collected by the listed authors, (c) the data were published in a journal or test manual, and (d) the primary research question(s) were focused on the application of the ASQ (or Infant Monitoring Questionnaire). After applying these four criteria, 45 of 186 sources were identified for inclusion in this review. As sources were reviewed, studies were classified as either an investigation of the technical properties of the ASQ or of implementation of the ASQ in applied settings. Within each of these categories, sources were separated if they were conducted with children from countries other than the United States. The present review only briefly summarizes the studies conducted outside of the United States, as these studies were primarily translations of the ASQ and subsequent investigations of applicability with samples outside the United States.

Scope and Sequence of Literature Review

Using the search procedure described above, articles that met inclusion criteria were reviewed in their entirety. The remainder of Chapter 2 is organized to review the history of developmental screening, issues and trends in the contemporary literature on developmental screening, empirical research on the completion of multi-dimensional developmental screeners by parents and professionals, and support for the use of the Ages and Stages Questionnaire as the primary study instrument of interest.

Historical Context for Developmental Screening of Children

The 1960s were an era of progressive reform because of growing concern for the poor educational and health outcomes of families who were described as being disadvantaged[1] because of limited financial resources (Children's Defense Fund, 1978; Frey & Walker, 2006). The achievement gap between students considered to be advantaged and disadvantaged was a prominent educational issue, prompting legislation for federal funding of Head Start to improve educational opportunities for young children from disadvantaged backgrounds (Frey & Walker, 2006). Studies conducted during this time highlighted the prevalence of medical conditions or diseases and secondary conditions that could be prevented among this population. In 1967, a provision to the Medicaid Act established the Early and Periodic Screening, Diagnosis, and Treatment (EPSDT) program with the primary purpose of early identification of and intervention with children who have "significant deviations in psychological, neurological, or emotional development" (Frankenburg & North, 1974, p. 139).

In addition to the growing concern for children whose families were characterized as disadvantaged, the 1960s and 1970s represented a time of heightened attention to the conditions of state institutions for individuals with disabilities (United States Department of Education, 2010). As part of the Elementary and Secondary Education Act (ESEA), the Education of Handicapped Amendments of 1974 (PL 93-380) established Child Find efforts for identifying children with disabilities from birth through 21 years, but there was no requirement for the provision of services (Paget & Nagle, 1986). Similar to EPSDT, Child Find arose from a response to growing awareness of the benefits of early identification and intervention for children with or at risk for disabilities. Unlike EPSDT, Child Find was intended to identify *any*

[1] The term "disadvantaged" was commonly used during the 1960s to refer to persons experiencing environmental risk due to limited finances. This terminology is used in this section to preserve historical accounts.

child who might have a disability. Although developmental screening arose from two different pieces of legislation often governing different disciplines (i.e., medicine and education), professionals in both fields faced similar implementation issues.

Early and Periodic Screening and Diagnosis Program

EPSDT was created to address the gaps in children's health care and placed an obligation on Medicaid that went beyond paying bills to assuring that eligible children received periodic screening and if a "diagnosis" was established, the child was to receive services early and as needed (Children's Defense Fund, 1978). From birth through 21 years, children from families characterized as disadvantaged were entitled under EPSDT to periodic physical examinations and screenings. EPSDT was the first major federal commitment to reach large numbers of children in need of basic health services as well as developmental and medical screenings.

Several years after the initial implementation of EPSDT, program evaluations conducted separately by the American Orthopsychiatric Association (AOA) and Children's Defense Fund concluded EPSDT was not adequately meeting the needs of the children who it was created to serve. Many barriers were identified at administrative and professional levels (AOA, 1978; Children's Defense Fund, 1978; Moore, 1978). EPSDT model demonstration programs heavily emphasized screening for cognitive development, primarily in relation to preparation for school and learning. Although this developmental area was essential, it did not take into account the importance of examining more than one domain of child development given the widely acknowledged recognition of the interrelationships among developmental domains (Shonkoff & Phillips, 2000) and the relationships of multiple developmental domains to later school success (Children's Defense Fund, 1978). This early narrow focus on cognitive development within the EPSDT program mirrored the goals of Head Start at that time and not the broader goal of illness prevention that EPSDT was created to address.

Another challenge in the implementation of EPSDT was the little value medical professionals placed on developmental screening as part of comprehensive health services (AOA, 1978). Even if developmental screening occurred under EPDST, professionals were often faced with a lack of available resources to conduct additional in-depth diagnostic assessments with a developmental focus or to offer services and supports to families if a child was identified as being at risk or having a developmental delay or disability (AOA, 1978; Children's Defense Fund, 1978). As discussed in the EPSDT guidelines, screening alone did not equate to offering help (Frankenburg & North, 1974). If families were not being connected to further evaluations or resources, then screening to identify risk, the need for additional diagnostic assessment, or the procurement of resources becomes questionable from an ethical perspective.

After conducting evaluations of EPSDT programs across the nation, the AOA (1978) and Children's Defense Fund (1978) concluded that the most significant area for improvement was informing and involving parents in the screening processes and information gathering related to making informed decisions. Overall, parents were reported to be minimally included in the screening process. In some EPSDT sites, for example, parents were not present for the screening (AOA, 1978; Children's Defense Fund, 1978; Moore, 1978). Information about screening was provided at times deemed inappropriate by the program evaluators. Within the primary health care setting, little or no time was allocated to inform parents of the screening results, provide recommendations for what to do next, or advise on how to address screening data that indicated risk or the need for additional assessment. Thus, the lack of continuity among screening, assessment, and intervention was due to insufficient resource allocation at the broad administrative level and within individual clinical or primary care practices.

Individuals with Disabilities Education Act

Several legislative acts were passed to assist states in educating children with disabilities as an alternative to home schooling, institutions, or non-public segregated schools. As part of ESEA, the Education for All Handicapped Children's Act (EHA; PL 94-142) was passed in 1975 to guarantee a free, appropriate public education to children with disabilities in every state, ages 6 to 18 (U.S. Department of Education, 2010). Subsequent amendments made to EHA included changing its name to the Individuals with Disabilities Education Act (IDEA) in 1990. The passage of EHA brought incentives for the identification of preschoolers with disabilities through the Preschool Incentive Grant and guaranteed access to special education and related services following the determination of eligibility for services (Kenny & Culbertson, 1993; Paget & Nagle, 1986). Child Find was the term designated for the process for identifying children, ages birth through 21 years, with delayed development and who subsequently, may benefit from special education services (Katz, 1989; Paget & Nagle, 1986). Although screening was not specifically described in this legislation, a national survey of implementation of early school screening requirements with lead educational agencies reported 33 states mandated some form of screening for preschoolers for physical, emotional, behavioral or learning problems (Gracey, Azzara, & Reinherz, 1984). The majority of these states reported conducting screenings primarily one time, upon entry into preschool or kindergarten, while only eight states conducted periodic (ongoing) screening. More than half of the states collected screening information from parents, demonstrating recognition of the important information parents have about their child's growth and development (Diamond, 1993; Diamond & Le Furgy, 1988; Diamond & Squires, 1993).

In 1986, the reauthorization of EHA through the passage of Public Law 99-457 contained provisions to expand the availability of early intervention services to eligible infants and toddlers

and established the preschool program for children with disabilities (IDEA, 20 U.S.C. 1419 § 300, 1986; U.S. Department of Education, 2010). With the new amendments to IDEA, free, appropriate public education was now available for children 3 through 21 years, and the Part H (now Part C) program was available for eligible children from birth through age 2 and their families. Included in the amendments was an increased emphasis on the family's role as fully participating members of the assessment and intervention team (Katz, 1989). Thus, legislation was increasingly acknowledging the importance of including parents in the decision-making process of determining whether a child needed early intervention or special education services.

Kochanek (1993) conducted a review of state policies present in 1991 pertaining to early identification and Child Find. He reviewed written policies for the presence of 10 exemplary screening indicators: multiple data sources, multiple data reporters, multivariate decision making, coordination with other screening efforts, family as informant, cultural sensitivity, screening as services, periodicity, score reliability and validity, and training. The last four indicators were not specified in federal regulations, but were considered recommended practices in assessment (DEC, 1993). Kochanek concluded that there was considerable variability in state policies on the inclusion of knowledge-based, exemplary components, and notably, state policies were less likely to make reference to screening indicators if they were not explicit in federal regulations. The majority of states did not reference periodicity, psychometric properties of screening instruments, or screening as an initial step in comprehensive assessment that was linked to intervention. More importantly, approximately 30% of states did not make reference to families being major informants regarding child and family status, needs, and goals. The majority of states made reference to this indicator, but the policies were often vague and did not provide specific procedures and implementation strategies related to family involvement.

Approximately 20% of states had written policies that defined families' roles, provided explicit procedures, and outlined quality assurance standards. Although it is possible that written policies did not reflect actual implementation in the field, findings from the Kochanek study indicated the need for guidance at the state level to support the use of recommended practices in screening.

Parallel Themes in Medical and Educational Legislation

The passing of EPSDT and EHA/IDEA reflected the enactment of federal policies that explicitly acknowledged the importance of early identification of children at risk for or with disabilities. Identification of these young children was facilitated by the implementation of developmental screening by professionals properly trained in administration and interpretation. While there was widespread acceptance of the concepts of early identification and the use of developmental screening, there was not universal implementation by all EPSDT programs or local education agencies in states. There was evidence of the early acknowledgement of improved accuracy in decision-making when professionals include parent input (AOA, 1978; Paget & Nagle, 1986). However, professionals in the health care and education fields had struggled to implement this practice effectively. Policies typically offered few guidelines on how to include parents as informants, leading to minimal or non-existent parent involvement in developmental screening conducted in health care clinics and schools. To improve practices that engage parents in the developmental screening process, a need existed for clarity and specificity in policies and recommendations.

Recommended Practices in Developmental Screening

Several organizations have offered recommended assessment practices for use with young children: American Academy of Pediatrics (2001; Council on Children with Disabilities et al., 2006), Division for Early Childhood (2007; Neisworth & Bagnato, 2005), National Association of School Psychologists (2009), and the National Association for the Education of Young

Children (2003). Of these, AAP and NASP make specific reference to screening while the remaining statements refer broadly to various assessment purposes and processes. Notably, although NAEYC (2003) does not specifically include screening in its recommended practices, it is a required practice for early education and care programs that are applying for accreditation status (NAEYC, 2012). Collectively, these organizations unanimously support the gathering of multiple sources of evidence to inform screening or assessment decisions, and specifically the inclusion of parent involvement in the decision-making process. Specific recommended practices and guidelines will be reviewed in conjunction with other relevant literature.

Criteria for Selecting Developmental Screening Instruments

The *Standards for Educational and Psychological Testing* (1999), a joint publication by the American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME) is commonly considered to be the prominent authority of standardized test construction. Some researchers have applied these standards to specifically address developmental screening instruments (Brassard & Boehm, 2007; Glascoe, 1991; Glascoe, 2007a; Salvia & Yssledyke, 1985). Jointly, the recommendations of these associations and researchers outline definitions and recommended criteria for developmental screening instruments for standardization, score reliability, score validity, sensitivity, and specificity (Table 2-1).

Three reviews have been conducted on the psychometric properties of commercially available developmental screening instruments. The first review, conducted by Katoff and Reuter (1980), included 21 developmental screening instruments. A decade later, Glascoe, Martin, and Humphrey (1990) selected a review panel of pediatricians and a special educator to examine the psychometric properties of 19 developmental screening instruments available at that time. Most recently, Macy (2012) reviewed published reports from peer-reviewed journals on 14

developmental screening instruments. Comparisons of the results of these three reviews show an increase in the number of developmental screening instruments that meet appropriate test standards (Glascoe et al., 1990; Katoff & Reuter, 1980; Macy, 2012). The selection of an appropriate developmental screening instrument requires the knowledge to critique the characteristics of an instrument using a set of recommended criteria.

Training

A quality developmental screening program begins with professionals who have been properly trained in administration and interpretation of developmental screening instruments (Kenny & Culbertson, 1993). In order to be effective, professionals who are using developmental screening instruments must receive training about developmental differences of infants, toddlers, and preschoolers that could affect the interpretation and use of screening results in conjunction with other information (Brassard & Boehm, 2007; Nagle, 2007; Paget & Nagle, 1986). An understanding of early childhood assessment practices is beneficial regardless of whether the professional is relying on direct elicitation methods or caregiver report. If direct assessment is conducted by the professional, additional training on developmentally appropriate assessment techniques would become necessary to ensure score validity and reliability (Bradley-Johnson & Johnson, 2007; Brassard & Boehm, 2007; Paget & Nagle, 1986) Using caregiver report to complete a developmental screening instrument would eliminate the necessity of specialized training, which could potentially translate to savings in cost and personnel time.

Time for Administration

Multi-dimensional developmental screening instruments should include items that reflect a broad sampling of skills across multiple domains of development (Boan et al., 2007; Council on Children with Disabilities et al., 2006; Glascoe, 2010; Meisels & Wasik, 1990; Neisworth & Bagnato, 2005). Taking this premise into consideration, instruments should also be relatively

brief in order to fit within the constraints of a busy clinical practice, large caseload, or early education and care program (Brooks-Gunn & Lewis, 1983; First & Palfrey, 1994; Meisels & Wasik, 1990). This includes considerations related to ease of scoring and interpretation of results (Brooks-Gunn & Lewis, 1983; Scott & Carran, 1989). While many developmental screening instruments take, on average, about 15 minute to complete (Council on Children with Disabilities et al., 2006), there is often limited time for a professional to build adequate rapport with the child to gather information in a direct elicitation format. This limitation might preclude valid inferences from screening data gathered using direct assessments. Pediatricians frequently report that although they consider early identification of developmental problems important, administration time deters them from routinely using developmental screening instruments (Glascoe & Dworkin, 1993; Smith, 1978). As a result, pediatricians will frequently rely on clinical judgment when deciding whether to refer a child for additional evaluation (Glascoe & Dworkin, 1993; Sand et al., 2005; Scott Lingaraju, Kilgo, Kregel, & Lazzari, 1993; Smith, 1978). This approach is heavily prone to error and might lead to the under-identification of children, particularly those with more subtle delays (Ysseldyke & O'Sullivan, 1987).

The use of standardized screenings is recommended over informal assessments (Glascoe, 1991; Marks, Hix-Small, Clark, & Newman, 2009; Meisels & Wasik, 1990; Squires, Nickel, & Eisert, 1996; Ysseldyke & O'Sullivan, 1987). Standardized screening instruments can take the form of parent- or professional-completed questionnaires about the child's activities, achievement of developmental milestones, or parent concerns, or by direct elicitation and observation of the child's ability to complete specified tasks. Parent-completed instruments can be completed prior to arriving at a scheduled appointment, leaving more time during the appointment for discussion or other evaluations. Although parent-completed instruments are

typically written at a functional reading level (Squires et al., 2009), practitioners should be sensitive to the needs of some parents who may require verbal administration or clarification of questions.

Serial Use

Developmental screening should be repeated over time to take into account age-related manifestations of delay (Council on Children with Disabilities, et al., 2006; Glascoe, 1991; Glascoe, 2005; Glascoe, 2007a; Jackson & Needelman, 2007; Kenny & Culbertson, 1993; Kochanek, 1993; NASP, 2009). When developmental screening is used as a one-time activity, such as screening upon entry into a program, children whose difficulties are not evident to informants until a later point might be overlooked. Repeated screenings provide other opportunities to identify children who have received false negative results on prior screens, meaning that they has a "negative" screen and were not referred for further evaluation when they should have been.

Multi-Source and Multi-Method

The screening process should use an ecological framework that accounts for the multiple contexts in which children grow: the home, community, and school (Bagnato, 2007; Boan et al., 2007; Colligan 1981; DEC, 2007; Glascoe, 1991; Hall, Bramlett, Barnett, & Cox, 1994; Kochanek, 1993; NASP, 2009; Neisworth & Bagnato, 2005; Paget & Nagle, 1986; Taylor 1993). Information should be gathered by using multiple modes of assessment (i.e., observation, interview, direct elicitation), multiple sources (i.e., early childhood educators, family members, professionals), and sample behaviors from multiple contexts (i.e., community, school, home). Taking into account other dimensions of the child's life and the potential interaction between those contexts improves the ecological validity of screening decisions.

Parent Involvement

DEC (2007) specifies that assessment should be a "shared experience between families and professionals" (p.10) in which there is constant communication of ideas and experiences with the intent of promoting a child's growth and development. It is recommended that throughout the screening process, communication with parents is frequent, clear, and culturally sensitive (Eddey, Robey, Zumoff, & Malik, 1995; Moore, 1978; Neisworth & Bagnato, 2005; Scott & Carran, 1989). Parents would benefit from having a clear understanding of the process, having their concerns attended to, and receiving immediate feedback about results.

If the results of a screening indicate the child may potentially have issues that require further evaluation, parents should be linked to assistance, either by a referral to the appropriate agency or by a list of recommendations for follow-up (Dworkin, 2006; Moore, 1978). As screening serves as the entry point into comprehensive assessment processes (Bagnato, 2007), interactions with parents set the tone for expectations of future encounters with other health and educational professionals (Thurlow & Gilman, 1999). Potential delays in a child's development can produce anxiety for parents and it is important to be sensitive to their needs during this process.

Parents should be invited to choose their preferred role in the assessment of their child, allowing for active engagement as team members in screening and subsequent procedures when desired (Bagnato, 2007; Brassard & Boehm, 2007; Kochanek, 1993; Ireton, 1990; Moore, 1978; Neisworth & Bagnato, 2005; Thurlow & Gilman, 1999). Parents can be included by asking them to observe and report on their child's behaviors (Diamond, 1993; Diamond & Le Furgy, 1988; Diamond & Squires, 1993; Marks et al., 2009). Given that screening is a relatively brief procedure, there often is limited time for a professional to build adequate rapport with the child to gather information in a direct elicitation format. This limitation might preclude valid

inferences from screening data. Parents have broader knowledge about their child's behaviors across time, contexts, and people than the professional is able to observe and elicit during a short period of time. Accurate information is likely to be gathered when parents are asked about observable behaviors, particularly behaviors that occur frequently. When asking parents about behaviors, it is recommended that parents select response choices rather than respond using open-ended formats (Diamond & Squires, 1993).

Historical and Contemporary Issues in Developmental Screening

The value of developmental screening in identifying children who may benefit from further evaluation and, if needed, early prevention or intervention, has been recognized in legislation and in position statements disseminated by several national organizations. There has been an increase in the specification of rigorous standards and guidelines that should be used to evaluate developmental screening instruments and their implementation within a developmental screening model. Prominent organizations, particularly AAP and DEC, have published policy statements and recommended practices to guide practitioners in selecting and using developmental screening instruments. Improvements have been made in the number of available instruments with adequate psychometric qualities. Since the early implementation of EPSDT and IDEA legislation, many challenges related to implementing parent involvement practices in developmental screening remain.

From the 1970s to the present day, experts have recognized that parent involvement throughout the assessment process is important in promoting ecological validity of results and forming positive relationships with families (Bagnato 2007; DEC, 2007; Glascoe, 1991). Over-reliance on professionally administered developmental screening instruments has resulted in inconsistent implementation of parent involvement in the assessment process mainly due to time constraints in busy clinical practices or early education and care programs. Having parents

complete developmental screening instruments addresses many obstacles to consistent implementation. Parents are able to observe a wider range of behaviors, across time, and across multiple contexts and people and can provide this information to professionals in a shorter time. Including parents at this early stage of the assessment process is an opportunity to provide awareness and knowledge about important developmental milestones, regardless of the results of the screening.

Many children are spending large amounts of time in early education and care environments with non-parental caregivers. Developmental screening instruments can serve as a method to systemically gather and compare observations of children's skills across different informants to assist in discussion about children's potential needs and strengths. Information gathered from parents and other significant caregivers in children's lives provides insight that a professional unfamiliar to the child or family is unable to observe in a short period of time. As a result of the growing need to integrate information from multiple contexts, it becomes increasingly important to examine relationships between different informants' judgments about child development when measured with the same developmental screening instruments.

Multi-Dimensional Developmental Screening Conducted by Parents and Professionals

The "accuracy" of parent report of the developmental status of their children has been described as congruence between parent and professional judgments. Congruence has been defined as the degree of correspondence between individuals' judgments or ratings (Dinnebeil & Rule, 1994; Snyder et al., 1992). The term *congruence* is preferred over *accuracy* as the latter implies an evaluative statement of one individual, typically the professional, being the "gold standard" to which a parent's estimate of development is measured. Congruence between parents' and professionals' judgments can be measured by: comparison of global ratings, percentage of agreement or disagreement in identification, and item analysis (Dinnebeil & Rule,

1994). As discussed in Chapter 1, the term congruence is used to capture similarities in ratings without the implication that agreement is the ultimate goal. Reviews conducted by Snyder et al. (1992) and Dinnebeil and Rule (1994) have supported the validity of eliciting parent report, however, the studies reviewed by these authors typically examined the relationship between parent report as elicited through a questionnaire or interview and a professionally completed, comprehensive, developmental assessment that was often directly administered to the child. To date, there has not been a published review of studies examining congruence when both informants complete the same multi-dimensional developmental screening instrument.

Using the search procedures and criteria described earlier, 10 studies were identified that examined the relationship between parent and professional (e.g., examiner, pediatrician, nurse, day care provider, psychologist) perspectives of child development across multiple domains using a developmental screening instrument. Findings from a descriptive analysis of these 10 studies are reported below. These 10 studies were categorized as either studies that primarily investigated interrater score reliability or studies that primarily investigated congruence. To provide clarification when discussing these studies, the following terminology will be applied: (a) "relationship between parent and professional ratings" or "parent-professional relationship" will be used to refer broadly to the group of 10 studies, regardless of the stated purpose and (b) "interrater score reliability" and "congruence" will be used to distinguish studies by their primary purpose.

Participants

Seven of 10 articles compared parent report to an examiner who did not regularly interact with the child (i.e., Bricker et al., 1988; Bricker & Squires, 1989b; Burgess, Asher, Doucet, Reardon, & Daste, 1984; Chiu & DiMarco, 2010; Glascoe, 2002; Squires & Bricker, 1991; Squires, Bricker, & Potter, 1997). Examiners represented the disciplines of medicine,

psychology, and education or were reported to be professionals trained in assessment. In two studies, the disciplines or training of examiners were not described. One of the studies sent screening forms to day care providers or kindergarten teachers and made comparisons to parent report (Coghlan et al., 2003). Two of the studies used both of these approaches by making parent-examiner comparisons and parent-teacher comparisons (Frankenburg et al., 1987; Frankenburg, van Doornick, Liddell, & Dick, 1976). None of the studies provided information on the practitioners' length of relationship or contact with the child.

The characteristics of each study sample are shown in Table 2-2. Across all studies, children's ages ranged from 2 weeks to 72 months. However, none of the studies described the distribution of ages within the sample or explored if results varied by age or age group. Half of the studies sampled children with identified risk factors (Bricker & Squires, 1989b; Bricker et al., 1988; Chiu & DiMarco, 2010; Squires et al., 1997; Squires & Bricker, 1991). Risk factors were either medical (e.g., remaining in the neonatal intensive care unit for a minimum of three days) or environmental (e.g., extreme poverty). In four studies, only a subsample of participants had screening forms completed by parents and a professional (Bricker & Squires, 1989b; Bricker et al., 1988; Squires et al., 1997). In two of the 10 studies, authors compared parent-examiner and parent-teacher pairs, but only a subsample of children had received teacher-completed ratings (Frankenburg et al., 1987; Frankenburg, van Doornick, Liddell, and Dick 1976). For six studies, it was not reported whether the characteristics of the subsample used to address study questions related to interrater score reliability differed from the larger study sample.

Methods to Examine Parent-Professional Relationship

Although all 10 studies had moderate to high levels of agreement or score congruence, a variety of approaches were used to explore the relationship between parent and professional judgments or ratings. In the sections that follow, the multi-dimensional developmental screening

instruments implemented in each of the 10 studies and the analyses used to interpret results are reviewed.

Measures of parent-professional relationship

The following multi-dimensional developmental screening instruments were represented in this review: Ages and Stages Questionnaire, 2nd Edition (ASQ; Bricker & Squires, 1999), previously known as the Infant/Child Monitoring Questionnaires (ICMQ; Bricker & Squires, 1989a), Brigance Infant and Toddler Screen (Brigance & Glascoe, 2002), Denver Developmental Screening Test (DDST; Frankenburg & Dodd, 1975), Denver II (Frankenburg et al., 1992), Denver Prescreening Developmental Questionnaire (PDQ; Frankenburg, van Doornick, & Liddell, 1976), the Revised Denver Prescreening Developmental Questionnaire (R-PDQ; Frankenburg, 1987) and Parents' Evaluation of Developmental Status (PEDS; Glascoe, 2007b). Table 2-3 shows the instruments used in each study.

Eight of the studies administered the same instrument to both parents and professionals (Bricker & Squires, 1989b; Bricker et al., 1988; Coghlan et al., 2003; Frankenburg et al., 1987; Frankenburg, van Doornick, Liddell, & Dick, 1976; Glascoe, 2002; Squires & Bricker, 1991; Squires et al., 1997). Frankenburg et al. (1987) and Frankenburg, van Doornick, Liddell, and Dick (1976) compared two pairs of scores (i.e., parent-teacher and parent-examiner). Parents and teachers completed the PDQ or R-PDQ, while the examiner administered the DDST. These eight studies were all standardization or validation investigations of new or revised screening instruments. Although the primary purpose of these studies was to establish interrater score reliability, analyses were conducted that permitted an examination of the relationship between parent and professional scores. The remaining two studies primarily focused on the congruence between parents and professionals when using multi-dimensional developmental screening

instruments (Burgess et al., 1984; Chiu & DiMarco, 2010), but neither article used the term congruence to describe these comparisons.

Analyses of parent-professional relationship

The analyses used in each study are shown in Table 2-3. The majority of studies evaluated congruence by calculating percent agreement on the classification of children (Bricker & Squires, 1989b; Bricker et al., 1988; Chiu & DiMarco, 2010; Frankenburg et al., 1987; Frankenburg, van Doornick, Liddell, & Dick, 1976; Squires & Bricker, 1991; Squires et al., 1997). Screening instruments classified children as either in need of further evaluation, at some risk of needing further evaluation, or no need for further evaluation at the present time. The remaining studies used other analysis procedures. To calculate parent-examiner agreement, Frankenburg, van Doornick, Liddell, and Dick (1976) calculated item-level agreement between the 10 PDQ questions with the corresponding DDST questions. Compared to the studies described above, Chiu and DiMarco (2010) reported agreement on classification for each developmental domain represented on the screening instrument. Similarly, Coghlan et al. (2003) reported item-level percent agreement, with each item representing a different developmental domain. Chiu and DiMarco (2010) and Coghlan et al. (2003) were the only two studies in which domain-specific results were reported.

Few studies used tests of statistical significance to evaluate the relationship between parent and professional scores. Burgess et al. (1984) used a formula for calculating a developmental quotient (DQ) to translate PDQ scores (Frankenburg, van Doornick, & Liddell, 1976) and DDST (Frankenburg & Dodd, 1975) in order to calculate correlations for three pairs of informants. Frankenburg, van Doornick, Liddell, and Dick (1976) tested for differences between the positive predictive values of parent- and teacher-completed screening forms when compared to the professionally administered DDST. In the standardization of the Brigance Infant

and Toddler Screen, Glascoe (2002) averaged scores across all domains for parent-completed and examiner-elicited screens. *F* test statistics were used to evaluate whether there were statistically significant differences between means. In addition to reporting item/domain-level percent agreement, Coghlan et al. (2003) calculated kappa values to correct for chance agreement.

Findings of Studies Primarily Investigating Congruence

Only two of the 10 studies reviewed had a primary research question concerning the comparison of parent and professional scores (Burgess et al., 1984; Chiu & DiMarco, 2010). In neither of these studies was the same screening instrument completed by parents and professionals. Burgess et al. (1984) and Chiu and DiMarco (2010) used the professionally administered DDST (Frankenburg et al., 1975) or Denver II (Frankenburg et al., 1992). Additionally, both of these studies compared parent report to examiners who did not typically interact with the child. Despite these limitations, these studies provide preliminary information about score congruence between parents and professionals when multi-dimensional developmental screening instruments are used.

Burgess et al. (1984) were interested in comparing the relationship between three pre-screening methods with a professionally administered DDST (Frankenburg et al., 1975). Two of the methods involved administering a pre-screen either completed independently by mothers (PDQ) or in an interview format (PDQ-M). The third group was prescreened using the Alpern-Boll Development Profile (Alpern, Boll, & Shearer, 1980).

PDQ and DDST results were translated into developmental quotient (DQ) scores to calculate correlations between screening instruments. The three prescreening tests correlated significantly with the DDST DQ (Burgess et al., 1984). The strongest correlation was between Alpern-Boll and DDST scores (r = .68), followed by the PDQ-M (r = .60) and the PDQ (r = .48).

These results are shown in Table 2-3. Using the DDST as the criterion, the number of correct referrals (positive results on the pre-screen and screen) and the number of under-referrals (negative prescreen and positive screen) were the highest for PDQ-M. The authors concluded that the PDQ-M was the most acceptable means for pre-screening children. Administering the PDQ in an interview format also addressed concerns of mothers' reading ability related to completing the PDQ.

The second study of congruence compared two developmental screening instruments, one completed by parents and the other completed by a nurse, using a sample of children who were homeless (Chiu & DiMarco, 2010). Children received screenings at a homeless shelter. Mothers independently completed an ASQ-2 (Bricker & Squires, 1999) for their child. A nurse administered the Denver II (Frankenburg et al., 1992). The results were compared using a calculation of percent agreement on the classification of children for further evaluation by each developmental domain. Agreement was highest across parents and nurses in personal-social and gross motor domains (95% each), followed by fine motor (71%). The lowest percent agreement was reported for the language domain (67%). The ASQ-2 alone classified children as needing further evaluation in the areas of personal-social, gross motor, and fine motor/adaptive. The Denver II identified more children for referral in the area of language than the ASQ-2. The authors hypothesized that this result might be attributed to mothers' ability to understand their children's verbalizations (i.e., parents crediting items on the ASQ-2) even when others cannot (i.e., nurses not crediting items on the Denver II).

Findings of Studies with a Focus on Interrater Score Agreement or Reliability

When examining overall agreement between those completing developmental screening instruments on the classification of children, results show moderate to high agreement. This methodology was used in six studies (Bricker & Squires, 1989b; Bricker et al., 1988;

Frankenburg et al., 1987; Squires & Bricker, 1991; Squires et al., 1997) and was included as one of the analyses by Frankenburg, van Doornick, Liddell, and Dick (1976). When agreement was examined at the item level, two studies showed moderate to high agreement between parents and professionals (Coghlan et al., 2003; Frankenburg, van Doornick, Liddell, & Dick, 1976). Statistical tests of significance were utilized in three studies and reported mixed findings (Coghlan et al., 2003; Frankenburg, van Doornick, Liddell, & Dick, 1976; Glascoe, 2002). Three of these studies compared parent reports to reports of teachers or day care providers (Coghlan et al., 2003; Frankenburg et al., 1987; Frankenburg, van Doornick, Liddell, & Dick, 1976), while the remaining studies conducted comparisons with professional examiners that did not regularly interact with the child. The results of these studies are summarized in Table 2-3 and described in detail below.

Frankenburg, van Doornick, Liddell, and Dick (1976) validated a newly created instrument, the PDQ (Frankenburg, van Doornick, & Liddell, 1976), which was developed through the selection and adaptation of 10 items from the DDST (Frankenburg et al., 1975). Parents completed the PDQ while a subsample (n = 202) was selected to receive a teacher-completed PDQ. Item-level agreement between parents and examiners was 93%, with a range of 68% – 100%. To determine parent-teacher agreement, researchers calculated classification agreement on the children, which was 59%. Based on the PDQ results, teachers referred more children than parents, while PDQs completed by parents had a slightly higher positive predictive value. When both of the PDQs were compared to the examiner-administered DDST, positive predictive value was 15.8% for parents and 10.8% for day care providers, but the difference between these values was not statistically significant. Comparisons between parent-examiner and teacher-examiner scores for all children, regardless of classification, were not available.

The PDQ was revised almost a decade later as the R-PDQ (Frankenburg, 1987). Instead of answering 10 questions as on the PDQ (Frankenburg, van Doornick, & Liddell, 1976), parents continued answering R-PDQ questions until a ceiling of three "No" answers was achieved (Frankenburg et al., 1987). Additionally, responses choices were simplified to either "yes" or "no," eliminating the response of "no opportunity." Similar to the procedures used in the validation of the original instrument, parents completed the R-PDQ while a subsample (n = 71) was selected to complete a teacher-completed R-PDQ. Classification agreement between parents and teachers was 83%. There were almost equal numbers of parent ratings identifying children as having delays as there were with teacher ratings identifying delays. In total, 193 children received R-PDQ screenings completed by parents and DDST screenings completed by examiners. Data available in tables published with the study indicated that classification agreement between parents and examiners was 74%. Comparing the classification agreement across all of the children shows that parents have higher agreement with another caregiver who regularly interacts with the child.

Using the IMQ/ASQ, agreement ranged from 87% - 97% between parents and professional examiners who had not previously met the child (Bricker & Squires, 1989b; Bricker et al., 1988; Squires & Bricker, 1991; Squires et al., 1997). In three of these studies, a portion of the original sample had to be eliminated from the analysis because one or more test domains had two or more uncompleted answers by the examiner (Bricker & Squires, 1989b; Squires & Bricker, 1991; Squires et al., 1997). Examiners reported that they had little opportunity to observe the child engaged in certain activities reflected in ASQ items.

The PEDS is comprised of 10 questions to elicit concerns in nine developmental domains (Coghlan et al., 2003; Glascoe, 2007b). Item/domain-level agreement between parents and

teachers and kappa values were calculated for each of these concerns. In all categories, parents reported more concerns than teachers. Percentages of agreement were considered to be high (>75%) for all of the categories except for behavior (64.9%). Kappa values calculated for each category were low to modest (-0.02 - 0.40); suggesting that the high percent agreement reported may have been due to chance. It is also possible that lower kappa values may be due to a low base rate of response in some of the categories (Streiner, 2003). These values may have differed if there were a higher number of at risk children included in the sample or if overall agreement for the PEDS had been examined.

During the initial investigation of the psychometric properties of the Brigance Infant and Toddler Screen (Brigance & Glascoe, 2002), versions were administered to parents to complete by self-report or interview and to examiners for completion by observation or direct elicitation (Glascoe, 2002). Scores for each screen were averaged across six domains and compared. There were no statistically significant differences in results using either form of screening administration after controlling for age and psychosocial risk ($F[5,170] = 0.543$).

Discussion and Limitations of Current Research

Developmental screening is increasingly becoming a part of common practice in health care and early education and care programs (National Research Council, 2008). While developmental screening has been traditionally used as a method of identifying children who are in need of additional supports, services, and evaluations, it can also serve as an avenue to share and exchange information about child development. When completed by caregivers representing the multiple contexts in which children spend time, developmental screening can serve as a formal and systematic method of exchanging information about children's development. When information is collected from parents and teachers, discussions are informed by screening results with enhanced ecological validity. A review of literature related to developmental screening

score congruence was conducted to understand what is known about the relationship between parent and professional scores when these two informant groups use multi-dimensional developmental screening instruments

Through a systematic search of the literature, 10 studies were identified in which parent and professionals completed a multi-dimensional developmental screening instrument. Overall, studies showed moderate to high levels of professional and parent agreement or congruence, however, there is a clear need for additional studies that administer the same screening instruments to two individuals who are in regular contact with the child. Only three studies paired parent responses with teachers from day-care centers or other settings (Coghlan et al., 2003; Frankenburg et al., 1987; Frankenburg, van Doornick, Liddell, & Dick, 1976). Using the PEDS (Glascoe, 2002), seven of 10 items had percentages of agreement above 80%, although kappa values were generally low (Coghlan et al., 2003). The study conducted by Frankenburg, van Doornick, Liddell, and Dick (1976) provided data for parent-examiner (PDQ-DDST) and parent-teacher (PDQ-PDQ) agreement, however, it is difficult to compare these numbers because of the different methods employed to calculate each of these percentages. When validating the Revised PDQ, similar methods were applied to measure parent-teacher (R-PDQ – R-PDQ) and parent-examiner (R-PDQ – DDST) classification agreement (Frankenburg et al., 1987). Although results showed higher agreement between caregivers who regularly interact with their child, parents and examiners were completing two instruments with differing item content and administration methods (i.e., parent report versus direct elicitation). Additional research is needed to examine congruence by having parents and professionals complete the same screening instrument using the same administration methods, as differences in agreement may be a

function of the items and structures of the instruments rather than reflecting differences in perspectives about child development.

Improved methods for examining congruence are needed. Studies need to be conducted that use analyses that provide information beyond classification agreement. Likewise, calculating agreement by using an overall average masks potential differences by domain. The results of Chiu and DiMarco (2010) and Coghlan et al. (2003) demonstrate that congruence can differ depending on the skill being addressed. Both of these studies had high agreement for gross motor skills and low agreement in the rating of language skills. It is possible that congruence on developmental screening items might also differ depending on the child's age or the amount of time the early care and education provider has known the child, but these differences are unknown because results in previous studies were not differentiated by age of child or length of contact.

The majority of the studies reviewed primarily focused on demonstrating the score reliability of a new or revised instrument. Interrater score reliability was calculated by comparing ratings of parents and professionals. These studies provided some information as to how similarly parents and professionals report children's development, however, interrater score reliability operates under the assumption that informants need to be interchangeable and are randomly selected (Suen et al., 1995). Interrater score reliability is a reflection of the stability of a score across informants and high interrater score reliability suggests similar scores would be obtained independent of the informant used. This concept does not apply when informants are not assumed to be interchangeable, as in the case of parents and professionals. Whether or not the professional is someone who interacts with the child regularly, parents and ECEs likely will consider child development with different sets of experiences, backgrounds, and perspectives.

Alternatively, it is possible that parents and ECEs will have similar perspectives about child development due to the significant amount of time each person spends with the child, sharing many of the caretaking responsibilities.

Ultimately, similarities and differences in perspectives about child development should be used to help inform how perspectives should be gathered, shared, and used to create a mesosystem. Children's developmental potential is enhanced when caregivers from different microsystems communicate information that promotes the compatibility of expectations and activities (Bronfenbrenner, 1979). The relationships that occur within the mesosystem are critical for supporting children's development (Bronfenbrenner, 1974). Thus, the relationship between scores obtained from the two informants can serve as a starting point for conversations about why children might exhibit skills differently based on context or how caregivers can support their generalization across contexts. In addition, understanding variables that affect congruence of perspectives about child development might help inform activities related to establishing and maintaining parent-professional partnerships.

Multi-Dimensional Developmental Screening with the Ages and Stages Questionnaires

The present study focused on the congruence between parent and professional perspectives about child development and the ASQ-3 (Squires & Bricker, 2009) was used as the multi-dimensional developmental screening instrument. The ASQ-3 was already being implemented as part of a universal screening program at the early education and care program that served as the site for the present study. Children's ECEs complete the ASQ-3 at designated intervals as part of ongoing activities at the center. Prior to implementing the developmental screening instrument with parents from the center as part of the present study, a review of the ASQ empirical literature was conducted to support its continued use.

Overview of ASQ Studies

Twenty-seven articles investigated the use of the ASQ in other countries; 17 sources sampled participants residing within the United States. Seven of the 17 sources included research questions concerning the technical properties of the ASQ. Seven sources implemented the ASQ in applied settings such as pediatric clinics or home visiting programs. Three sources were interested in the ability of ASQ score to describe the development of a special population of children.

Studies Examining the Technical Properties of the ASQ

Seven sources were identified that investigated the technical properties of the Infant Monitoring Questionnaire (IMQ, a precursor to the ASQ) and the ASQ (Bricker & Squires, 1989b; Bricker et al., 1988; Gollenberg, Lynch, Jackson, McGuinness, & Msall, 2010; Squires & Bricker, 1991; Squires et al., 1997; Squires, Potter, Bricker, & Lamorey, 1998; Squires et al., 2009). The majority included articles published in journals (n = 6). One source was the technical manual for the current edition of the ASQ (Squires et al., 2009). The authors of the ASQ-3 conducted seven of the eight psychometric investigations reviewed here.

Each study was reviewed and data were gathered on demographics of participants and findings of score reliability (i.e., test-retest, interrater, and internal consistency) and score validity (i.e., construct validity, sensitivity, specificity, false positive ratio, false negative ratio) of the IMQ or ASQ. Definitions of each of these psychometric terms and recommended criteria were adapted from a review of seminal literature in standards for assessment (AERA et al., 1999; Salvia & Ysseldyke, 1985) and literature on developmental screening instruments (Brassard & Boehm, 2007; Glascoe, 1991; Glascoe, 2007a). Detailed definitions and criteria are listed in Table 2-1.

Standardization sample

Revisions for the current edition of the ASQ (i.e., ASQ-3, Squires et al., 2009) were based on 18,572 questionnaires completed by parents of children between 1 and 66 months of age. Within this sample, 3,434 children had more than one completed questionnaire. Therefore, the total demographic sample consisted of 15,138 children. Data were collected across four years in either paper or electronic format via a project-website. Respondents represented a wide geographic distribution, from all 50 states, and several U.S. territories. When broken down by age, sample sizes for questionnaires ranged from 352 (2 months) to 2,088 (12 months) children (Squires et al., 2009). There were no questionnaire intervals that fell below the recommended criteria of 100 sample participants per age group (Brassard & Boehm, 2007; Glascoe, 1991; Salvia & Ysseldyke, 1985).

It is recommended that the normative sample be representative of the population for which a screening instrument is intended (AERA et al., 1999; Brassard & Boehm, 2007; Glascoe, 1991). The ethnicity of the ASQ-3 normative sample of children was comparable to the U.S. Census estimates for 2007, suggesting representativeness of the general population (Squires et al., 2009). Data were gathered on mother's highest level of education completed and family income, which were proportionately similar to the most recent U.S. Census. Children included in the sample were identified as having one or more medical or environmental risk factors. Although these terms were not defined in the current test manual, studies of previous editions defined medical risk factors of infants to be at least three days spent in the neonatal intensive care unit or having been referred for an evaluation due to medical or developmental concerns (Bricker & Squires; 1989b; Bricker et al., 1988; Squires et al., 1997). Environmental risk factors include: (a) household income below the federal poverty line, (b) maternal age younger than 19 years at time of child's birth, (c) highest level of education completed by the mother was below

the 12th grade, and (d) a history of parental involvement with Child Protective Services. The description of the normative sample suggests that the authors made adequate effort to support the use of ASQ-3 with children from diverse ethnic, geographic, and socioeconomic distributions.

Descriptions of participants from studies of previous editions of ASQ. Compared to past studies of the ASQ, revisions for the third edition constituted the largest sample, spanned across more age intervals, and provided the most detail on participant characteristics (Table 2-4). These differences were partly due to the addition of age intervals with each subsequent revision; the IMQ had eight questionnaires from 4 to 36 months and the current edition, the ASQ-3, has 21 questionnaires from 2 – 60 months (Bricker & Squires, 1989b; Squires et al., 2009).

Investigations of special populations. Three studies have examined whether ASQ scores correctly identify children within special populations. The ASQ has been used to explore the frequencies of detection for children conceived by intracytoplasmic sperm injection and in vitro fertilization (Squires, Carter, & Kaplan, 2003). The ASQ identified children within this population who were in need of referral at the same frequency of detection as the normative sample. Researchers investigated the implementation of the ASQ to identify developmental problems in children aged birth – 3 years, who have experienced child maltreatment and compared screening results when conducted by child welfare (CW) or early intervention (EI) providers (McCrae, Cahalane, & Fusco, 2011). The ASQ identified children in need of further evaluation at similar frequencies as a national sample. However, frequencies of referral by EI providers were more accurate, as compared to national data, with EI providers identifying children at over three times the frequency of CW providers. Authors suggested that CW providers might lack the knowledge or training in child development issues needed to conduct

developmental screenings effectively. Additionally, the authors suggested the differences might also be due to CW providers' caseload size and focus of services.

Children of teen mothers are also considered a special population because of the increased probability of environmental risks (Ryan-Krause, Meadows-Oliver, Sadler, & Swartz, 2009). When comparing mothers' classifications to classifications using a professionally administered standardized instrument, there was little to no agreement. A high frequency of delay was identified when using the standardized instrument, but not the parent-completed ASQs. The authors concluded that mothers' inability to accurately complete the ASQ was due to unrealistic and inaccurate information about child development. In a different study with at risk mothers, it was demonstrated that completing ASQs increased mothers' agreement with professionals on their child's developmental status (Squires & Bricker, 1991). Agreement was measured using scores from mothers' reports during an interview and professionally administered assessment. Mothers at-risk who completed at least one ASQ had significantly higher levels of agreement than mothers at risk who did not complete an ASQ. The low agreement reported in Ryan-Krause et al. (2009) may indicate the need for improved education about child development for mothers at risk before or as part of completing multi-dimensional developmental screening instruments.

Validity

All seven studies examining the technical properties of the ASQ reported some validity data. In each of these studies, validity was addressed by comparing the classification of children based on their performance on a standardized developmental assessment with their classification based on their parent-completed ASQ (Bricker & Squires, 1989b; Bricker et al., 1988; Gollenberg, et al., 2010; Squires & Bricker, 1991; Squires et al., 1997; Squires et al., 1998; Squires et al., 2009). Studies typically mailed or gave the age-appropriate ASQ to parents to complete at home. Evaluations were scheduled within 2 weeks of receiving the completed ASQ.

Criterion instruments included the Battelle Developmental Inventory (BDI; Newborg, Stock, Wnek, Guidubaldi, & Svinicki, 1984, 2004), Bayley Infant Scales of Development (BSID; Bayley, 1969, 1993), Gesell and Armatruda Developmental and Neurological Examination (Knobloch, Stevens, & Malone, 1980), McCarthy Scales of Children's Abilities (McCarthy, 1972), and the Stanford-Binet Intelligence Test (Thorndike, Hagen, & Sattler, 1985). Children's performance was typically classified as *fail* on the standardized test when scores were 1.5 *SD* below the mean (standard score <75), with the exception of one study where *fail* was designated for scores 2 *SD* below the mean (Gollenberg, Lynch, Jackson, McGuinness, & Msall, 2010).

Criterion validity. For all studies, criterion validity was reported as a percent agreement on classification of children as either *pass* or *fail*. Results are summarized in Table 2-5. Classification agreement between the ASQ-3 and the BDI was 86% (Squires et al., 2009). For prior editions of the ASQ, agreement ranged from 79 - 93% (Bricker & Squires, 1989b; Bricker et al., 1988; Gollenberg, et al., 2010; Squires & Bricker, 1991; Squires et al., 1997; Squires et al., 1998). In a study that investigated the effectiveness of the ASQ with mothers of low-income status versus middle-income status, there were no statistically significant differences in percent agreement with the criterion instrument between these two groups (Squires et al., 1998). One study reported correlations between the ASQ and criterion instrument (Gollenberg et al., 2010). Statistically significant correlations were shown between the BSID-II Mental Scale and ASQ communication ($r = .52$, $p < .001$) and personal-social ($r = .45$, $p < .01$) domains. The BSID-II Motor Scale significantly correlated with the gross motor domain ($r = 0.46$, $p < .01$) of the ASQ.

Sensitivity. The sensitivity of a screening instrument is defined as its ability to accurately refer children who are identified on a standardized instrument as having a delay (Glascoe, 1991; Glascoe, 2007a). Sensitivity is calculated by the formula: true positives/(true positives + false

negatives). In the reference to screening results, *positives* designate children who fail screening or standardized evaluation (i.e., children who are positive for a condition) and *negatives* designate children who pass screening or standardized evaluation (i.e., children who are negative for a condition). Sensitivity for the current edition of ASQ is .86 (Squires et al., 2009), exceeding the recommended criteria of .70 - .80 (Brassard & Boehm, 2007; Glascoe, 1991; Glascoe, 2007a). Studies investigating prior editions of the ASQ showed sensitivities ranging from .54 – 1.00 (Bricker & Squires, 1989b; Bricker et al., 1988; Gollenberg et al., 2010; Squires et al., 1997). Results are displayed in Table 2-5.

Specificity. The specificity of a screening instrument is an indicator of its ability to identify children (using the designation of *pass*) who are not identified as having a delay in a standardized evaluation (Glascoe, 1991; Glascoe, 2007a). Specificity is calculated by using the formula: true negatives/(false positives + true negatives). The ASQ-3 has demonstrated specificity of .86 (Squires et al., 2009). As shown in Table 2-5, studies of prior editions showed specificities of .80 - .95 (Bricker & Squires, 1989b; Bricker et al., 1988; Gollenberg et al., 2010; Squires et al., 1997; Squires et al., 1998). The recommended criteria for specificity of screening instruments is .70 - .80 (Glascoe, 1991; Glascoe, 2007a). All of the studies of the ASQ have demonstrated results that have met this criterion.

Other measures of accuracy. Another method to evaluate the accuracy of a screening instrument is to examine indicators that show the proportion of children who are misclassified: false positive ratio, false negative ratio, underreferral rates and overreferral rates. False positive ratios are the percentage of children who receive failing scores on a screening instrument, but receive passing scores on a standardized developmental assessment (Glascoe, 1991; Glascoe, 2007a). False negative ratio is the percentage of children who receive a passing score on a

screening instrument, but have failing scores on a standardized developmental assessment. Both of these numbers are used to calculate the proportion of children who are under- or overreferred. During the standardization of the ASQ-3, false positive and negative ratios were both reported to be 14% (Squires et al., 2009). This resulted in an underreferral rate of 6% and overreferral rate of 8%. Rates for prior editions of the ASQ are displayed in Table 2-5.

Reliability

For developmental screening instruments, the recommended criterion for interrater and test-retest score reliability is 80% (Glascoe, 1991; Glascoe, 2007a; Salvia & Ysseldyke, 1985). Five of the seven ASQ studies reported interrater or test-retest score reliability data (Bricker & Squires, 1989b; Bricker et al., 1988; Squires & Bricker, 1991; Squires et al., 1997; Squires et al., 2009). The results of these studies are summarized in Table 2-6.

Interrater score reliability. Parents were given the ASQ to complete at home and return by mail (Bricker & Squires, 1989b; Bricker et al., 1988; Squires & Bricker, 1991; Squires et al., 1997; Squires et al., 2009). Within 2 weeks, parents brought the children to an evaluation in which a standardized instrument was administered by a trained examiner who was blind to the results of the parent-completed ASQ. After administration was completed, examiners completed an ASQ. Interrater score reliability was measured as percent agreement on the classification of children. Based on 107 children, interrater score agreement on the ASQ-3 was 93% (Squires et al., 2009). The first and second editions showed levels of agreement ranging from 87 – 100% (Bricker & Squires, 1989b; Bricker et al., 1988; Squires & Bricker, 1991; Squires et al., 1997). In three of these studies, a portion of the sample could not be included in the analyses because examiners were unable to complete some of the items (Bricker & Squires, 1989b; Squires & Bricker, 1991; Squires et al., 1997). Agreement was not reported separately by domain.

Test-retest reliability. As mentioned above, parents were given the ASQ to complete at home. For a subsample of children, parents completed a second ASQ within 2 weeks during the child's standardized evaluation (Bricker & Squires, 1989; Bricker et al., 1988; Squires et al., 1997; Squires et al., 2009). Based on 142 children, the ASQ-3 has test-retest score reliability of 93% (Squires et al., 2009), while prior editions have reported between 91 – 99% (Bricker & Squires, 1989b; Bricker et al., 1988; Squires et al., 1997).

Validated for use in other countries

The ASQ-3 is available in Spanish for use in the United States (Guiberson & Rodriguez, 2010; Squires et al., 2009). The English version of the ASQ has been investigated for use in Canada and Australia (Elbers, Macnab, McLeod, & Gagnon, 2008; Frisk et al., 2009; Skellern, Rogers, & O'Callaghan, 2001). Additionally, the ASQ has been translated for investigation of utility in Denmark (Klamer, Lando, Pinborg, & Greisen, 2005), France (Dionne, Squires, Leclerc, Peloquin, & McKinnon, 2006; Flamant et al., 2011), Korea (Heo, Squires, & Yovanoff, 2008), Netherlands (Kerstjens et al., 2009), Norway (Janson & Squires, 2004; Richter & Janson, 2007), Spain (Campos, Squires, & Ponte, 2011), Taiwan (Tsai, McClelland, Pratt, & Squires, 2006) and Turkey (Kapci, Kucuker, & Uslu, 2010). Generally, the technical properties and the distribution of scores of children from other countries have been similar to findings with samples from the United States.

Studies that Examined the Utility of the ASQ within a Screening Process

Seven of the 45 studies implemented the ASQ in applied settings as part of a screening process. The primary research questions concerned its feasibility and acceptability with different contexts and children. The next several sections describe aspects of these studies in relation to settings, detection of children in need of referral, time and cost for implementation, potential

benefits for children and families, and implementation of the ASQ in screening programs located in other countries.

Settings

The ASQ has been primarily investigated as a screening instrument within a developmental surveillance model in primary care clinics (Earls, Andrews, & Hawy, 2009; Hix-Small, Marks, Squires, & Nickel, 2007; Jee et al., 2010; Marks et al., 2009). One study investigated the effect of systematic education about the use of ASQ on pediatric residents' knowledge and preferences (Thompson, Tuli, Saliba, DiPietro, & Nackashi, 2010). The ASQ has been implemented in a Healthy Start home visiting program that supports children and families from birth through 5-years-old (Squires, Katzev, & Jenkins, 2002). Home visitors assisted families with reading and answering questions.

Detection of children in need of referral

After implementing a parent-completed screening instrument as part of developmental surveillance, several studies reported increased rates of detection of children needing further evaluation (Earls et al., 2009; Hix-Small et al., 2007; Jee et al., 2010; Marks et al., 2009). These primary care centers typically relied on clinical judgment to determine if a referral was necessary. Prior to the use of the ASQ, physicians were most likely to refer only if the child had delays in multiple domains, only gross motor delays, or only communication delays (Earls et al., 2009; Hix-Small et al., 2007). Jee et al. (2010) reported increased referral rates in all domain areas, but the greatest increase was observed in fine motor and personal-social referrals. Systematically collecting information from parents about all domains of development improved physicians' ability to detect problems in skill areas in which they may have been less familiar. Another study examined identification rates in children born pre-term compared to a group of children born full-term and found increases in detection rates were greater for pre-term children

(Marks et al., 2009). Early identification within this at risk population is important and it appeared that incorporating information from parents was beneficial for the physician.

Time and cost for implementation

Besides considering the technical properties of a screening instrument, it is recommended that screening instruments are relatively brief in order to ease adoption into the constraints of a busy practice or other early childhood setting (Brooks-Gunn & Lewis, 1983; First & Palfrey, 1994; Meisels & Wasik, 1990; Scott & Carran, 1989). This includes the time it takes to score and interpret results. In two studies, it took very little time to implement the ASQ when totaling the amount of time it took to give the ASQ to parents, explain its purpose, provide instructions, score, interpret results, and enter results into the electronic medical chart (Hix-Small et al., 2007; Jee et al., 2010). Training office staff on the entire screening process took approximately 30 minutes. Additionally, the itemized cost per patient ranged from $1.61 to $2.43 (Hix-Small et al., 2007). Costs varied depending on whether the ASQ was completed in office or mailed back and the amount of time required for physician follow-up.

Benefits for children and families

In order for children and families to benefit from the implementation of the ASQ, health care clinics and early education and care centers need to be using it in practice. Despite the numerous criticisms on its technical adequacy, the DDST or Denver II remains the developmental screening instrument of choice, particularly by physicians (Glascoe et al., 1992; Thompson et al., 2010). To address this gap between research and practice, the ASQ was incorporated into a pediatric resident education program (Thompson et al., 2010). Residents received a didactic lecture about three screening tools: Denver II, ASQ, and PEDS (Glascoe, 2007b). Following the lecture, residents completed a 30 minute structured clinical observation on the administration of each instrument and completed four evaluations in which all three tools

were used with each child. Prior to the lecture, residents reported having little knowledge of the ASQ, whereas the majority reported having experience with the Denver II. Following the four evaluations, residents reported the highest preference for the ASQ, particularly for its quality results and time efficiency. Thus, increasing pediatricians' awareness of and experience with incorporating the ASQ into developmental surveillance improved the likelihood that they will use this practice.

The ASQ was incorporated into a home visiting program for first-time parents (Squires et al., 2002). Beginning when the child was approximately 4 months of age, home visitors assisted parents in reading and completing the items during sessions. After scoring the items together, a sheet of activities was given to the parents to try before the next screening interval. Staff reported incorporating the ASQ into their sessions assisted in providing structure for the visits and when suggesting games and activities for the parents. They shared that parents from low-income families were able to complete the items with little trouble. Staff noted the ASQ engaged parents in viewing their children using a strength-based perspective.

Implementation in other countries

International researchers have investigated the feasibility and acceptability of incorporating the ASQ in developmental screening programs or in the identification of special populations of children. Studies have been conducted in Australia (Gasson & Piek, 2003; Lindsay, Healy, Colditz, & Lingwood, 2008; Smith et al., 2012), Canada (Kim & O'Connor, 1996; Rydz et al., 2006), Denmark (Plomgaard, Hansen, & Greisen, 2006), Ecuador (Handal, Lozoff, Breilh, & Harlow, 2007a, 2007b), and France (Troude, Squires, L'Helias, Bouyer, & de La Rochebrochard, 2011). The ASQ was used in a multinational study that included 19 countries (Yu et al., 2007). The findings of these studies suggest that the ASQ was equally effective in these countries when incorporated into developmental screening programs.

Support for the Use of the Ages and Stages as a Multi-Dimensional Developmental Screening Instrument

Upon review of the empirical studies on the ASQ, the ASQ generally meets recommended standards for technical adequacy as a screening instrument. The normative sample of the ASQ-3 was a large, relatively representative sample of children and the technical manual reported adequate levels of concurrent score validity, specificity, sensitivity, and other measures of accuracy. Reports of interrater and test-retest score reliability showed the ASQ can yield similar information from different informants who have observed the child. Several studies examined the utility of the ASQ as a tool to systematically incorporate parents' observations into developmental screening programs in primary care and home visitation settings. Researchers reported improved rates of detection of children needing further evaluation. The ASQ has been shown to be a time and cost-efficient procedure for screening children. Most importantly, incorporation of the ASQ represented improvements in services for children and families. Parents were actively engaged in child assessment and, in Squires et al. (2002), use of the ASQ resulted in reports of improved structure of visits and recommendations for supporting children's development and learning.

Summary

Growing numbers of children are being cared for in contexts outside of their home, particularly in early education and care programs. As developmental screening is increasingly incorporated into these programs in response to recommendations by prominent organizations and changes in federal and state policies, screening can serve dual purposes of early identification and a method to share formally perspectives about children's development across contexts. A mutual understanding of children's abilities and needs can enhance the experiences provided across multiple settings. A review of the historical background of developmental

screening reveals that while parent involvement has been long regarded as a recommended practice in screening, it has not been implemented consistently. Difficulties with incorporating parent input have been partially due to vague policies and guidelines for practitioners.

In the past decade, numerous organizations have provided policy recommendations and position statements on either developmental screening or the assessment of young children, including screening. Collectively, these organizations unanimously support the gathering of multiple sources of evidence to inform assessment, specifically, the inclusion of parent input in the decision-making process. Using parent-completed developmental screening instruments fits within the constraints of busy medical practices, early education and care programs, and school districts and is likely to improve the ecological validity of results.

An investigation of the literature on the relationship between parent and professional ratings as measured with multi-dimensional developmental screening instruments identified 10 studies demonstrating moderate to high agreement or score congruence. However, several limitations were identified in these studies that might preclude valid inferences about parent-ECE congruence. Typically, professionals in the studies were individuals who had limited contact with the child, which might affect score validity. There were weaknesses in the methodology used to gather and interpret score congruence. The majority of studies used percent agreement between parents and professionals on overall classifications of children, without investigating differences by age, developmental domain, or length of contact with the child.

Understanding variables that affect similarities and differences in ratings can inform how information should be gathered, shared, and used between adults who spend significant time with young children in different contexts (e.g., home and school). A review of the empirical literature on the Ages and Stages Questionnaire has provided evidence of its utility as a multi-dimensional

developmental screening that can be used in different contexts. The literature review also highlighted the need to conduct additional investigations of congruence and to explore the potential of the ASQ as a tool for facilitating the sharing of information about children's development and learning across home and early education and care contexts. The current literature comparing developmental screening instruments completed by parents and professionals cannot adequately address these issues, as the professionals included in research studies typically were unfamiliar with the child outside of assessment context.

Table 2-1. Select technical properties and recommended criteria for screening instruments.

Technical property	Definition	Recommended criteria
Standardization	Clear description of groups included in standardization sample; standardization sample representative of the population for which the test is intended; children with disabilities are included in sample or special studies are reported in the manual.	100 subjects per age group.
Score reliability	The degree to which scores are free of measurement error for a given group.	Test-retest and interrater reliability of .80 or greater.
Score validity	The degree to which accumulated evidence and theory support the proposed interpretation of test scores.	Correlation of .70-.80 with a test battery that samples the same range of developmental tasks.
Sensitivity	The proportion of cases in which an instrument refers a child who is later determined to have a delay. Sensitivity is calculated by true positives/(true positives + false negatives).	.70 to .80
Specificity	The proportion of cases in which an instrument does not refer a child who does not have a delay. Specificity is calculated by true negatives/(false positives + true negatives).	.70 to .80

Table 2-2. Sample characteristics of studies examining the relationship between parent and professional informants using developmental screening instruments.

Citation	Sample characteristics	Ages	*n*
Bricker & Squires, 1989b	Infants with medical risk factors	4 - 24 months	39
Bricker, Squires, Kaminski, & Mounts, 1988	Infants with medical risk factors	4 - 24 months	73
Burgess, Asher, Doucet, Reardon, & Daste, 1984	Sampled from pediatric clinic	6 – 60 months	60
Chiu & DiMarco, 2010	Homeless children, none chronically ill or with developmental delay	4 – 60 months	21
Coghlan, Kiing, & Wake, 2003	Sampled from child care centers and kindergartens in Australia	18 – 72 months	262
Frankenburg, Fandal, & Thornton, 1987[a] (parent-examiner)	Sampled from pediatric clinics	2 weeks – 72 months	193
Frankenburg et al., 1987[a] (parent-teacher)	Sampled from day care centers	2 weeks – 72 months	71
Frankenburg, van Doorninck, Liddell, & Dick, 1976[a] (parent-examiner)	Sampled from community	3 – 72 months	1,115
Frankenburg, von Doorninck, Liddell, & Dick, 1976[a] (parent-teacher)	Sampled from day care centers	3 – 72 months	202
Glascoe, 2002	Sampled across United States	2 weeks – 24 months	408
Squires & Bricker, 1991	Children with environmental risk factors	4 – 30 months	14
Squires, Bricker, & Potter, 1997	Children with medical risk factors, environmental risk factors, and normative sample	4 – 36 months	112

Note. Medical risk factors include more than 3 days spent in the NICU or referred for evaluation of medical or developmental concerns. Environmental risk factors included household income below the federal poverty line, maternal age <19 years at time of birth, maternal education <12th grade, or history of parental involvement with Child Protective Services. [a]Studies selected a subsample of participants to receive additional screenings from day care providers. Subsamples are reported for parent-examiner agreement and parent-teacher agreement.

Table 2-3. Summary of studies examining parent-professional relationship on developmental screening instruments.

Citation	Professional role	Professional screener	Parent screener	Analysis	Results
Bricker & Squires, 1989b	Examiner	IMQ	IMQ	Classification agreement	Overall agreement – 87%
Bricker, Squires, Kaminski, & Mounts, 1988	Examiner	IMQ	IMQ	Classification agreement	Overall agreement – 97%
Burgess, Asher, Doucet, Reardon, & Daste, 1984	Examiner	DDST	PDQ or PDQ-M	Correlation of DQ	*r* = 0.478** (PDQ) *r* = 0.597** (PDQ-M)
Chiu & DiMarco, 2010	Examiner	Denver II	ASQ	Classification agreement	Personal-social – 95% Gross Motor – 95% Fine Motor - 71% Language - 67%
Coghlan, Kiing, & Wake, 2003	Day-care or Kindergarten provider	PEDS	PEDS	Item agreement; Kappa values	Other concerns – 95.8%, κ = -0.02 Gross motor – 94.7%, κ = 0.40** Global/cognitive – 92.4%, κ = -0.04 Fine motor – 91.6%, κ = 0.11 Receptive language – 90.8%, κ = 0.18** School – 90.8%, κ = 0.28** Self-help – 87%, κ = 0.23** Social-emotional – 79.8% κ = 0.37** Expressive language – 77.5%, κ = 0.26** Behavior – 64.9%, κ = 0.13*
Frankenburg, Fandal, & Thornton, 1987	Day-care provider	R-PDQ	R-PDQ	Classification agreement	Parent-teacher – 83%
	Examiner	DDST	R-PDQ	Classification agreement	Parent-examiner – 74%
Frankenburg, van Doornick, Liddell, & Dick, 1976	Day-care teacher	PDQ	PDQ	Classification agreement	Parent-teacher – 59%
	Examiner	DDST	PDQ	Item Agreement	Parent-examiner – 93%
Glascoe, 2002	Examiner	Brigance Infant and Toddler Screen	Brigance Infant and Toddler Screen	Comparison of mean scores	F[5,170] = 0.543
Squires & Bricker, 1991	Examiner	IMQ	IMQ	Classification agreement	Overall agreement – 100%
Squires, Bricker, & Potter, 1997	Examiner	ASQ	ASQ	Classification agreement	Overall agreement – 94%

Note. IMQ = Infant Monitoring Questionnaires. DDST = Denver Developmental Screening Test. R-PDQ = Revised Denver Prescreening Developmental Questionnaire. DDST = Denver Developmental Screening Test. PDQ = Denver Prescreening Developmental Questionnaire. PDQ-M = Denver Prescreening Developmental Questionnaire, by maternal interview. Denver II = Denver Developmental Screening Test II. ASQ = Ages and Stages Questionnaire. PEDS = Parents' Evaluation of Developmental Status. *p-value < 0.05 **p-value < 0.01

Table 2-4. Demographic information of participants described in psychometric studies of Ages and Stages Questionnaires.

		Child characteristic					Risk factor represented			Additional information	
Citation	*N*	*n*	2 – 24 months	25 - 36 months	37 - 60 months	Ethnicity	Medical	Environmental	Normative sample	Maternal education	Family income
Infant Monitoring Questionnaires											
Bricker, Squires, Kaminski, & Mounts (1988)	653	264	✓			✓	✓			✓	✓
Bricker & Squires (1989b)	743	265	✓				✓		✓	✓	✓
Squires & Bricker (1991)	18	18	✓	✓				✓		✓	✓
Ages and Stages Questionnaires – Second Edition											
Squires, Bricker, & Potter (1997)[a]	4,145	2,008	✓	✓		✓	✓	✓	✓	✓	✓
Squires, Potter, Bricker, & Lamorey (1998)	247	96	✓	✓		✓		✓	✓	✓	✓
Gollenberg, Lynch, Jackson, McGuinness, & Msall (2010)	40	40	✓							✓	
Ages and Stages Questionnaires – Third Edition											
Squires, Twombly, Bricker, & Potter (2009)	18,572	15,138	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note. N represents the total number of ASQs included in the analyses when children were sampled over time. Normative samples were defined as having no previous history of developmental or serious health problems, full-term (>37 weeks) at birth, and never assigned to a neonatal intensive care unit (NICU). Medical risk factors include more than three days spent in the NICU or referred for evaluation of medical or developmental concerns. Environmental risk factors included household income below the federal poverty line, maternal age <19 years at time of birth, maternal education <12th grade, or history of parental involvement with Child Protective Services. [a]Included ASQ sample from Bricker et al., 1988 and Bricker & Squires, 1989b.

Table 2-5. Results of validity studies of Ages and Stages Questionnaires (in percentages).

Citation	*N*	Criterion instrument	Agreement on classification	Sensitivity	Specificity	False positive	False negative	Under-referral	Over-referral
				Infant Monitoring Questionnaires					
Bricker & Squires (1989b)	743	Gesell, BSID, Stanford-Binet	79 - 92	63[a]	91[a]	-	-	7	6
Bricker, Squires, Kaminski, & Mounts (1988)	653	Gesell	79 – 94	54	95	-	-	3 - 16	1 - 9
Squires & Bricker (1991)	18	BSID	89	-	-				
				Ages and Stages Questionnaire – Second Edition					
Squires, Bricker, & Potter (1997)[b]	4,145	Gesell, BSID, Stanford-Binet, MSCA	84	75	86	14	23	4.0	11.9
Squires, Potter, Bricker, & Lamorey (1998)	247	BSID	80 – 93	-	85 - 89	-	-	0 - 5	6 - 20
Gollenberg, Lynch, Jackson, McGuinness, & Msall (2010)	40	BSID-II	88[c]	100	93	-	-	0	8
				Ages and Stages Questionnaire – Third Edition					
Squires, Twombly, Bricker, & Potter (2009)	18,572	BDI	86	86	86	14	14	6	8

Note. Children were classified as "fail" on the criterion instrument when scores were 1.5 SD below the mean. Gesell = Revised Gesell and Armatruda Developmental and Neurologic Examination; BSID = Bayley Scales of Infant Development; Stanford-Binet = Stanford-Binet Intelligence Test; MSCA = McCarthy Scales of Children's Abilities; BDI = Battelle Development Inventory. [a]Included sample from Bricker et al., 1988 in analysis. [b]Included samples from Bricker et al., 1988 and Bricker & Squires, 1989b in analyses. [c] "Fail" was designated by scores 2 *SD* below the mean.

Table 2-6. Results of score reliability studies of Ages and Stages Questionnaires.

Citation	Interrater Reliability (*n*)	Test-Retest (*n*)	Cronbach's α				
			C	GM	FM	PS	P
			Infant Monitoring Questionnaires				
Bricker & Squires (1989b)	87% (39)	91% (107)	-	-	-	-	-
Bricker, Squires, Kaminski, & Mounts (1988)	97% (73)	99% (68)	-	-	-	-	-
Squires & Bricker (1991)	100% (14)	-	-	-	-	-	-
			Ages and Stages Questionnaires – Second Edition				
Squires, Bricker, & Potter (1997)[a]	94% (112)	94% (175)	.63 - .75	.53 - .87	.49 - .79	.52 - .75	.52 - .68
			Ages and Stages Questionnaires – Third Edition				
Squires, Twombly, Bricker, & Potter (2009)	93% (107)	92% (145)	.57 - .83	.57 - .87	.51 - .83	.53 - .78	.51 - .71

Note. C = communication domain; GM = gross motor domain; FM = fine motor domain; PS = problem solving domain; P = personal-social domain. [a]Included sample from Bricker et al., 1988 and Bricker & Squires, 1989b.

CHAPTER 3
METHODS

Few empirical studies have been conducted to investigate parent and professional congruence as measured by a multi-dimensional developmental screening instrument when the same instrument is used and completed in the same way by each informant. The purpose of the present study was to contribute to the research on congruence between parent and professional judgment by recruiting a group of professionals (i.e., ECEs) who have consistent and frequent contact with the children. The following research questions were addressed in the present study:

1. When using the same multi-dimensional developmental screening instrument, what is overall congruence between two informants who interact with the child regularly?

2. Does congruence between parents and early childhood educators vary based on child age cohort?

3. Does congruence between parents and early childhood educators vary depending on the child developmental domain assessed (i.e., communication, gross motor, fine motor, problem solving, personal-social)?

4. Are there differences in the pattern of response selection across informant types (i.e., parents and ECEs)?

5. Does child age and length of relationship with the early childhood educator predict item-level congruence between parents and ECEs?

This chapter includes information about the design of study, setting, participants, measures, procedures, and data analyses. The description of procedures includes sampling and recruitment of subjects, instructions provided to participants, and data collection activities.

Design of Study

A cross-sectional, exploratory correlational research design was used. Parents and ECEs from two university-based early education and care centers were recruited at a single-point in time to compare their responses about child developmental status on a multi-dimensional developmental screening instrument. In the current study, congruence was defined as the

correspondence between the judgments of two informants. This terminology was selected to avoid using the value-laden terms of parent "over-estimation or under-estimation" of child development.

Setting

This study was conducted at two university-based early education and care centers. These centers provide early education and care to children ages 6 weeks through 5 years. The centers serve families who attend or are employed by a large university located in the southeastern United States. One center serves the entire university community, while the second center serves families affiliated with an academic health sciences center. At the time of recruitment, enrollment across the two centers was 268 infants, toddlers, and preschoolers.

Both centers have accreditation from the National Association for the Education of Young Children (NAEYC). The centers utilize a team teaching approach; two lead teachers collaboratively plan and teach all lessons and activities within each classroom. As a minimum requirement, ECEs must have a Child Development Associate (CDA) credential.

To fulfill the NAEYC accreditation requirement specifying that centers must conduct ongoing and systematic assessment of children's development (NAEYC, 2008), several different forms of assessment are used in the program. ECEs complete the Ages and Stages Questionnaire (ASQ-3; Squires & Bricker, 2009) for all children serially and at regular intervals to screen for developmental "red flags" and to inform decision-making related to the need to monitor more closely a child's development, to provide additional supports, or to refer for more comprehensive evaluation. All full-time staff are required to attend yearly training, provided by the student investigator of this study, on using and interpreting the ASQ-3. In the present study, ECEs and parents consented to complete ASQs for children enrolled in the early education and care

centers. Both informants completed ASQs for the same child, even if the child was not due for his or her regular ASQ screening interval at the center.

Participants

Early Childhood Educators

ECEs were eligible to participate in the study if they worked full-time (i.e., approximately 40 hours per week) in a classroom and had completed an ASQ-3 training provided by the center within the past 2 months. Fifty-one full-time early childhood educators were employed across the two centers at the time the present study was conducted, however, six of the ECEs were not regularly assigned to one classroom and were therefore not eligible for the present study. Forty-four of the 45 eligible ECEs consented to participate in the present study. During data collection, four teachers were removed from the pool of eligible ECEs because of maternity leave (n = 1) and reassignment to a new classroom within the center (n = 3). These ECEs were each able to complete 1 to 2 ASQs and therefore are included in the description of participants below.

Of the 44 participating ECEs, the majority were female (97.7%). ECE participants identified their race/ethnicity as European or Caucasian (68.2%), Latino or Latino American (15.9%), African or African American (9.1%), or mixed ethnicities (6.8%). The majority of ECEs had a bachelor's degree (56.8%), primarily in the areas of early childhood/early childhood special education (31.8%) or child development (20.5%). Forty-three percent of the ECEs reported having a child development associate (CDA) credential.

ECEs had been employed with the centers for an average of 26.2 months (range = 4 -150, SD = 32.4). Thirty-one ECEs had taught young children prior to working at the centers. Mean years of experience working with young children were reported to be 8.9 years (range = 1 - 40, SD = 8.1).

Parents and Children

Eligible parents for the present study were those who had already consented for their child to participate in regular screenings at the center and those not employed by the centers. In addition, the child of each eligible parent needed to be enrolled in an eligible ECE's classroom for at least 1 month before the parent-ECE pair completed the ASQ-3.

Some parents completed for ASQs for siblings, and subsequently, there were 110 parent participants in the study (detailed further below). The parents who completed the ASQ-3 forms represented 91 mothers/step-mothers (82.7%), 17 fathers/step-fathers (15.5%), 1 legal guardian (.9%), and 1 foster parent (.9%). Of these respondents, 83.6% were female (*n* = 92) and 16.4% were male (*n* = 18). Parents self-reported their affiliation with the university where the early education and care centers were based: 28.2% were faculty, 26.4% were staff, 26.4% were graduate or professional students, 8.2% were fellows, 7.3% were categorized as other (e.g., not affiliated due to merger with an off-campus center or other type of hospital affiliation), and 2.7% were undergraduate students. The majority of parents were born in the United States (68.2%). The race/ethnicity of the parents was reported as follows: 65.5% European or Caucasian, 17.3% Asian/Pacific Islander or Asian American, 7.3% Latino or Latino American, 4.5% mixed ethnicities, and 4.5% African or African American. Over one-quarter (*n* = 28) of the parents reported speaking a language other than English in the home.

There were 126 children in study sample. Table 3-1 shows the distribution of children across the four age cohorts. The largest age cohort represented was 42 – 60 months (*n* = 44), followed by 27 – 36 months (*n* = 38), 14 – 24 months (*n* = 27), and 2 – 12 months (*n* = 17). The mean age of the study sample was 32.9 months (range = 3.6 – 64.5, *SD* = 17.3). In the sample, 58.7% of the children were male and 41.3% were female. The race/ethnicity of the children were reported as follows: 57.9% European or Caucasian, 18.3% mixed ethnicity, 14.3% Asian/Pacific

Islander or Asian American, 6.3% Latino or Latino American, and 3.2% African or African American.

Thirty-six of the children were considered to be dual language learners and of these children, 12 spoke English as their second language. Data were collected on two risk variables identified as relevant for interpreting ASQ scores in the ASQ-3 manual (Squires et al., 2009). These risk variables were prematurity and having a history of concerns. There were 11 children (8.7%) who were born three or more weeks prematurely. Of these children, only one child received an ASQ that was adjusted for prematurity, as recommended in the ASQ-3 manual (Squires et al., 2009); the remaining children regularly received ASQ screenings based on their chronological age. Seventeen children (13.5%) had been evaluated at least once for medical or developmental concerns.

Measures

Demographic Forms

After giving written informed consent, ECEs and parents completed demographic forms. Parents' demographic forms were included in the packet that included the ASQ-3 forms. For each child who received an ASQ-3, the ECE completed an additional demographic form to collect information on the length of his or her relationship with the specified child.

Ages and Stages Questionnaire

The Ages and Stages Questionnaire - Third Edition (ASQ-3; Squires & Bricker, 2009) is a multi-dimensional developmental screening instrument for use with children 1 to 66 months of age. The ASQ-3 was designed with the intent that parents, or other caregivers who know the child well, complete the questionnaires. Parents or other caregivers complete the ASQ-3 by reading and responding to a series of questions about developmental skills appropriate for children at different age intervals. They are instructed to indicate if the child is doing the skill

regularly (as indicated by marking “yes”), “sometimes,” or “not yet.” There are 21 questionnaire intervals that are used for children from 1 – 66 months (Squires et al., 2009). Each questionnaire indicates the age range applicable for use. For example, the 12 month questionnaire is used for children between the ages of 11 months, 0 days through 12 months, 30 days. Each questionnaire has 30 items divided evenly between five developmental domains: communication, gross motor, fine motor, problem solving, and personal-social. The distribution of the number of items represented in each developmental domain remains consistent across the age intervals, but the pool of items varies for each age interval. Some specific items may be repeated across only two age intervals, while other items repeat more frequently.

The ASQ is scored by converting responses to points: each yes response is 10 points, each sometimes response is 5 points, and each not yet response is 0 points. Points are summed to obtain a score for each domain and transferred to a summary page. Age-based cut-offs are provided on the summary page and indicate whether scores for each developmental domain fall in one of the following categories: above the cutoff, close to the cutoff (referred to as monitoring zone or at-risk), and below the cutoff. Scores that are above the cutoff are interpreted as the child’s development is on schedule and no further action is needed. Scores close to the cutoff fall in the monitoring zone and it is suggested that learning activities and additional monitoring be conducted. If scores are below the cutoff, further assessment might be needed.

Validity of ASQ scores

A detailed review of score validity studies conducted for the three editions of the ASQ was provided in Chapter 2. This section focuses on the development of the ASQ questionnaires, comparison of performance on questionnaires by non-risk and risk groups, and the determination of cutoff scores. This section ends with a brief review of classification agreement between ASQ-

3 and the first and second editions of the Battelle Developmental Inventory (Newborg, et al., 1984, 2004).

Development of the ASQ. The content of the ASQ was informed by reviewing (a) standardized developmental tests published when the first version of the instrument (i.e., Infant Monitoring Questionnaire, IMQ) was developed, (b) nonstandardized tests focused on early development, (c) textbooks, and (d) literature about early developmental milestones (Squires et al., 2009). Items developed for the ASQ had to meet the following criteria: (a) skills could be easily observed or elicited by parents, (b) skills were highly likely to occur within homes and child care settings, and (c) skills reflected important developmental milestones. Developmental skills that met these criteria were written into items using concrete terms and vocabulary that did not exceed a sixth-grade reading level.

To narrow the large pool of potential items, a developmental quotient (DQ) was calculated for each item using Equation 3-1.

$$(\text{age equivalent})/(\text{age interval of ASQ item}) \times 100 = \text{DQ} \qquad (3\text{-}1)$$

In this formula, the age equivalent was derived from the original source of the item (Squires et al., 2009). For example, if the item was derived from the Battelle Developmental Inventory (Newborg et al., 1984, 2004), the age equivalent specified in the BDI for a skill was the one used in this formula. The final pool of items for each age ASQ interval questionnaire was selected if the DQ ranged from 75 to 100. This criterion was selected to restrict items for a particular questionnaire interval to those that targeted a skill at the middle to low end of the developmental range for that particular chronological age interval. The rationale was that a child who was not observed to use a skill calculated at a DQ of 75 to 100 might benefit from further assessment.

Comparison of non-risk and risk groups. A sample of 18,572 ASQ questionnaires for children aged 1 to 66 months was used to determine the cutoff scores for the ASQ-3 (Squires et al., 2009). At the time of ASQ-3 completion, families were asked to indicate whether the child had any environmental or medical risk factors. Environmental risk was defined as family income level below the federal poverty guideline, maternal age of 19 years or younger at the time of the child's birth, maternal education below the 12th grade, or a history of involvement with child protective services. Medical risk was defined as prematurity (< 39 weeks gestation) or infant birth weight less than 3 pounds, 5 ounces. Using these guidelines, 19% (n = 2,186) of the ASQ standardization sample (n = 18,572) had two or more risk factors. When mean scores were compared between the non-risk and risk groups, means for the non-risk group were generally higher than the means for the risk group. The *ASQ-3 User's Guide* (Squires et al., 2009) provides graphic displays of the means of non-risk and risk groups across all ages, for each developmental domain. The authors noted that there were some exceptions in which the risk group had consistently higher means than the non-risk group, although they attributed these anomalies to low representations of children at risk at certain age intervals.

Determination of cutoff scores. Non-risk and risk groups were combined for analyses to determine cutoff scores (Squires et al., 2009). The authors used two strategies to determine cutoff points: (1) receiver operating characteristic (ROC) curves and (2) comparing the percentage of children identified as needing further evaluation using three different cutoff scores. In both of these strategies, the authors compared cutoff scores for each questionnaire interval that were at 2, 1.5, and 1 standard deviations below the mean score.

ROC analyses were applied to compare conditional probabilities for sensitivity, specificity, false positive proportion, false negative proportion, under-identification, and over-

identification, for each of the cutoff points (Squires et al., 2009). Analyses were conducted for each questionnaire interval. The authors were interested in selecting a cutoff score that represented a reasonable balance between over-identification and under-identification rates. Over-identification was defined as the proportion of children identified as needing further assessment by the ASQ, but who received scores in the typical range (scaled score above 75) on the BDI or BDI-2. Under-identification was defined as the proportion of children identified as not needing further assessment by the ASQ, but receiving scores below the typical range (scaled score of 75 and below) on the BDI or BDI-2. When cutoff scores were set at 1 and 1.5 SD below the mean, the over-identification rate increased as the under-identification rate decreased. Cutoff scores were set at 2 SD below the mean because they were the most balanced in terms of the true positive and false positive proportions for all 21 questionnaires.

The second strategy entailed comparing the percentage of children that fell below each of the cutoff scores when set at 2, 1.5, and 1 SD below the mean (Squires et al., 2009). The authors referenced U.S. Census Bureau and Centers for Disease Control and Prevention prevalence data for young children with developmental disabilities to set target identification percentages. It was projected that questionnaires should identify 12% - 16% of children for referral in one developmental area, while 2% - 7% of children should be identified for referral in two or more areas. Using these percentages, 2 SD below the mean was selected to be the cutoff point for all 21 questionnaires.

Classification agreement. For the ASQ-3, classification agreement was measured by comparing the classification of children based on their performance on the Battelle Developmental Inventory first and second editions (BDI, Newborg et al., 1984; BDI-2, Newborg et al., 2004) with their classifications on the ASQ-3 (Squires et al., 2009). Agreement meant that

the child either received “typical” scores on both the BDI and ASQ-3 or received “eligible” (for referral) scores on the BDI and ASQ-3. A child’s performance on the BDI was designated as eligible if the child’s scaled score was equal to or less than 75 on any scale or subscale. A child’s performance on the ASQ-3 was designated as typical if the child’s classification was above the cutoff or monitoring zone and as eligible if the child’s classification was below the cutoff. A total of 18,572 questionnaires were used for analyses. Overall classification agreement between BDI and ASQ-3 was 85.8%. Classification agreement also was examined for four age cohorts: 2 -12 months, 14-24 months, 27 – 36 months, and 42 – 60 months. Across these intervals, classification agreement ranged from 82.6% (14-24 months) to 88.9% (2 – 12 months).

Reliability of ASQ scores

This section describes the internal consistency, test-retest, and interrater reliability of scores for the ASQ-3 as reported in the *ASQ-3 User’s Guide* (Squires et al., 2009).

Internal consistency score reliability. The internal consistency score reliability of all questionnaires was examined through correlational analyses and Cronbach’s coefficient alpha (Squires et al., 2009). For 20 ASQ intervals (9- and 10-month questionnaires were combined), Pearson product moment correlation coefficients were calculated for developmental domain scores with overall ASQ scores. For each developmental domain score, correlations with ASQ overall scores were as follows when all questionnaire intervals were combined: communication (r = .76), gross motor (r = .65), fine motor (r = .73), problem solving (r = .78), and personal-social (r = .79). All correlations were statistically significant at $p < .01$. These findings suggest moderate to strong internal consistency between developmental domain scores and total test score.

For each of the 20 questionnaire intervals, Cronbach’s coefficient alphas were calculated for developmental domain scores (Squires et al., 2009). Alphas ranged from .51 to .87 across all

developmental areas and age intervals. Table 2-6 shows the alpha ranges for each developmental domain. Alphas specific to each questionnaire are available in the *ASQ-3 User's Guide* (Squires et al., 2009). These alphas indicate that the ASQ items have good to acceptable internal consistency score reliability.

Test-retest score reliability. To examine test-retest score reliability, 145 parents were asked to complete a second questionnaire within 2 weeks of completing an initial questionnaire. Agreement was defined as having the same classification (i.e., typical or eligible) on both questionnaires. The percent agreement for the 145 parents was 92%. Intraclass correlation coefficients (ICCs) calculated for classification agreement ranged from .75 to .82. Similar results were found in other studies examining test-retest reliability of all editions of the ASQ and are displayed in Table 2-6.

Interrater score reliability. To examine interrater score reliability, 107 children were selected to have questionnaires completed by both parents and trained test examiners (Squires et al., 2009). Parents completed the ASQ at home and mailed in the forms. Within 2 weeks of completion, children were brought in for a standardized assessment with a trained test examiner who administered the BDI-2 to the child (Newborg et al., 2004). Immediately following the assessment, the examiner completed the age-appropriate ASQ for the same child. Interrater score agreement was defined as having the same classification (i.e., typical or identified) on both parent- and examiner-completed questionnaires. Across all questionnaires, the percent agreement between parents and examiners was 93% and the ICCs of developmental domain classification agreement ranged from .43 (communication) to .69 (personal-social). Results of the ICCs for other developmental domains were not provided.

Procedures

Sampling and Recruitment

For the present study, a statistical power analysis suggested a minimum of 30 Parent-ECE pairs, up to a maximum of 35 parent-ECE pairs, would be optimal to recruit within each of the following age cohorts: 2 - 12 months, 14 – 24 months, 27 - 36 months, 42 – 60 months. These age cohorts were selected to replicate the age intervals used in validity studies with the ASQ-3 (Squires et al., 2009). The age-cohort affiliation for each participant was based on the applicable ASQ-3 age interval for the child. For example, according to ASQ-3 forms, a child who is 25 months, 5 days old would receive the 24-month questionnaire (Squires et al., 2009), and therefore, the parent-ECE pair would be categorized within the 14 - 24 month age cohort.

ECEs were recruited prior to sending informational letters about the study home with children enrolled in their classrooms. The student investigator for the present study attended each center's monthly hallway meeting to inform ECEs about the study and to invite them to participate. Of the 45 eligible ECEs, 44 consented to study participation and signed informed consent forms.

Recruitment procedures for parents included sending home informational letters and informed consent forms. Distribution of these forms was accomplished by placing them in all eligible children's cubbies (n = 262), having blank consent forms available in the classrooms and at the front desk of each center, and greeting parents outside of classrooms to inform them about the study. Six families were not eligible to participate in the present study because they were employed by the centers. Signed consent forms were received for 145 children, representing 54.1% of eligible children across both centers. Refer to Table 3-2 for percentages of recruited eligible children by age cohort.

To be included in analyses, the ECE and parent must have completed the ASQ-3 within 2 weeks of each other. This time frame was selected based on interrater score reliability studies of previous and current editions of the ASQ (Bricker & Squires, 1989b; Bricker, et al., 1988; Squires & Bricker, 1991; Squires et al., 1997; Squires et al., 2009). Five parent-ECE pairs returned ASQ forms outside of the 2-week window and were not included in analyses. Fourteen parent-ECE pairs had incomplete forms, meaning that consent had been received from the ECE and parent, an ASQ-3 packet had been given, but either one or both of the packets were not returned. Data for 126 parent-ECE pairs across the two centers were included in analyses (Table 3-2).

Instructions for ECE Practitioners and Parents

ASQ-3 packets for ECEs and parents had identical sets of instructions describing the completion of the form and the scoring rubric to be used (Appendix). Parents were informed they were permitted to discuss the items with other caregivers, but could not consult with any staff from the early education and care center and vice versa for ECEs. The same respondent was asked to complete the demographic form and ASQ-3 form. The instructional form included clarification on the three scoring responses (yes, sometimes, and not yet) along with examples that illustrated how to select the appropriate response on the ASQ-3 questionnaire.

Data Collection

After parents and ECEs consented to participate in this study, each parent-ECE pair was given packets containing instructional forms (Appendix), demographic forms, and ASQ-3 forms to complete for the same child. The ASQ-3 summary sheets (used to determine cutoff scores and assign classification categories) were not included in either packet, so parents and ECEs were unable to determine scoring classification categories. To ensure that forms were returned within

the 2-week time period, parents and ECEs received a verbal and written reminder within 1 week of receiving the forms.

Of the 110 parents who completed an ASQ-3, 95 parents (86%) completed an ASQ-3 for 1 child, 14 parents (13%) completed an ASQ-3 for 2 children, and 1 parent completed an ASQ-3 for 3 children. ECEs completed ASQs for an average of 2.9 children (range = 1 - 6, modes = 1, 3). These centers utilize a team teaching model, subsequently, 44 ECEs were distributed across 24 classrooms. Within these classrooms, ASQs were completed for an average of 5.3 children (range = 1-10, mode = 4). The distribution of questionnaires across the available age intervals is displayed in Figure 3-1. The questionnaires were not administered at the 2-month or 9-month intervals, as there were no consented participants within the applicable age range for these two age intervals. For the remaining 19 age intervals, the number of questionnaires distributed ranged from 1 (at the 20-month interval) to 15 participants (at the 33-month interval).

Data Analyses

Using the scoring procedures described above, parent- and ECE-completed ASQ-3 forms were scored by the author and checked independently by a second person. Discrepancies in scoring were checked against hard copies of the forms and corrections were made on the form. Demographic form data were entered into Microsoft Excel™ and checked using a double-data entry procedure. ASQ-3 forms were entered into a REDCap database (Harris et al., 2009) by the author and checked by a second person. A function is available within REDCap by which forms can be designated as “incomplete,” “unverified” and “verified.” After the initial data entry, a form was marked as unverified. After a second person looked over the electronic entry and verified all responses had been correctly transferred from the hard copy form into the database, the form was marked as verified. After this procedure was completed for all parent and ECE

forms, the RedCap database was exported into an Excel spreadsheet. Demographic and ASQ Excel sheets were exported into SPSS Version 20 for analyses.

The following questions were used to guide analyses conducted in the present study:

1. When using the same multi-dimensional developmental screening instrument, what is overall congruence between two informants who interact with the child regularly?
    a. What is the item-level agreement between parents and ECEs across all ASQ items?
    b. What is the classification agreement between parents and ECEs across all 5 developmental domains?
2. Does congruence between parents and early childhood educators vary based on child age cohort?
    a. What is the item-level agreement between parents and ECEs for each age cohort?
    b. What is the classification agreement between parents and ECEs for each age cohort?
3. Does congruence between parents and early childhood educators vary depending on the child developmental domain assessed (i.e., communication, gross motor, fine motor, problem solving, personal-social)?
    a. What is the item-level agreement between parents and ECEs for each developmental domain?
    b. What is the classification agreement between parents and ECEs for each developmental domain?
    c. What is the item-level agreement between parents and ECEs within each developmental domain for each age cohort?
    d. What is the classification agreement between parents and ECEs within each developmental domain for each age cohort?
    e. What is the relationship between parent and ECEs developmental domain raw scores?
4. Are there differences in the pattern of response selection across informant types (i.e., parents and ECEs)?
    a. As a group of informants, do parents select responses (i.e., yes, sometimes, not yet, omitted) at similar frequencies to ECEs across all ASQ items?

b. As a group of informants, do parents select responses at similar frequencies to ECEs across all ASQ items for each age cohort?

c. As a group of informants, do parents select responses at similar frequencies to ECEs within each developmental domain?

d. As a group of informants, do parents select responses at similar frequencies to ECEs within each developmental domain for each age cohort?

5. Does child age and length of relationship with the early childhood educator predict item congruence between parents and ECEs?

Congruence Examined using Percent Agreement

Congruence was examined by calculating percent agreement between parents and ECEs using two approaches. The first approach examined parent and ECE congruence on ASQ items. The second approach examined parent and ECE congruence on classification decisions.

Congruence on ASQ items

Item-level agreement between parents and ECEs was calculated across ASQ items to generate a percentage. When parents and ECEs selected the same response, agreement was coded as 1; when different responses were selected, agreement was coded as 0. If one informant omitted the item, this was also coded as 0. Item-level agreement was calculated in two ways: (1) by using the three original response choices (yes, sometimes, not yet) provided in the ASQ-3 (Squires & Bricker, 2009) and (2) by using two response choices generated by considering yes and sometimes as agreement. When calculating agreement for two response choices, agreement was coded as 1 when both informants selected either yes or sometimes. Research questions for which item-level agreement was calculated are shown in Table 3-2.

To examine overall congruence on ASQ items, the total number of agreements was summed and divided by the total number of items (v = 30) to generate a percentage (Equation 3-2).

Overall item agreement = (Total number of agreements)/30 X 100 (3-2)

Overall item agreement was calculated for each parent-ECE pair using Equation 3-2 and then averaged across all 126 parent-ECE pairs. To examine differences by child age cohort, children were divided into one of four age cohorts based on ASQ interval: birth – 12 months, 14 – 24 months, 27 -36 months, and 42 – 60 months. Agreement was calculated by averaging Equation 3-2 over the parent-ECE pairs in each age cohort.

To examine item-level agreement by developmental domain, Equation 3-2 was adjusted to account for the number of items in each domain. The 30 ASQ items are divided evenly among the five domains so that each domain is comprised of six items (Squires et al., 2009). The formula is shown below (Equation 3-3).

Developmental domain agreement = (Number of agreements)/6 X 100 (3-3)

To examine developmental domain differences within each age cohort, Equation 3-3 was applied for each of the five developmental domains, and then percentages were averaged across the number of parent-ECE pairs in each age cohort.

Congruence on classification decisions

Classification decisions were determined by age-based cutoffs for developmental domain scores specified on the summary sheets (Squires et al., 2009). Classification agreement was calculated in two ways: (1) using either the three original descriptive categories (above the cutoff, monitoring zone/at-risk, and below the cutoff) provided in the ASQ-3 (Squires & Bricker, 2009) or (2) using two categories generated by considering above the cutoff and monitoring zone as agreement. For the latter method, the two classification categories were renamed to typical (i.e., either above the cutoff or monitoring zone) and eligible (i.e., below the cutoff) to parallel terminology used in validity studies of the ASQ-3 (Squires et al., 2009). When parent and ECE domain scores translated to the same classification category, agreement was coded as 1; when domain scores translated to different classification categories, agreement was coded as 0. When

calculating classifications using two categories, agreement was coded as 1 if either above the cutoff or monitoring zone was selected. Research questions for which classification decision congruence was calculated are shown in Table 3-2.

To calculate overall classification agreement, the total number of agreements on each classification decision was added for all 5 domains and divided by the total number to generate a percentage (Equation 3-4). Percentages were averaged across all parent-ECE pairs in the sample.

Overall classification agreement =

(Total number of agreements on classification decisions)/5 X 100 (3-4)

To examine classification agreement for each age cohort, overall classification agreement was calculated and averaged across the number of pairs in each age cohort using Equation 3-4. This equation was simplified when calculating classification agreement by developmental domain. Classification decisions are based on the total score within each developmental domain. Therefore the only possible values for each parent-ECE pair's agreement on classification decision for each developmental domain is either 0% or 100%. The number of parent-ECE pairs who agreed for the classification decision were summed and divided by the total number of children (N = 126) to represent the proportion of pairs with agreement on domain classification (Equation 3-5). This formula was applied for each developmental domain.

Developmental domain classification agreement =

(n pairs with agreement)/N X100 (3-5)

When examining developmental domain differences within each age cohort, N represented the total number of pairs within a specific age cohort.

Congruence on Developmental Domain Raw Scores

Many researchers recommend the use of more than one index to examine score reliability (Portney & Watkins, 2000). In addition to calculating percent agreement on items and

classification decisions by developmental domain and age cohort, an intraclass correlation coefficient (ICC) was calculated using each informant's score on each of the five developmental domains. ICC is calculated using variance estimates obtained through analysis of variance, thus it reflects both degree of correspondence and agreement among ratings. The ICC ranges from 0.00 to 1.00, with higher numbers indicating that two groups of informants rate similarly. For the purposes of this study, parents were considered one group of informants and ECEs are the comparison group of informants. Several models are available to calculate ICCs and the selection of model is based on the nature of the study. Equation 3-6 was used to calculate ICC because each child is assessed by a set of different informants (i.e., unique sets of parent-ECE pairs) and informants were considered as if they were chosen randomly (Portney & Watkins, 2000). This model is based on a one-way analysis of variance in which children are treated as the independent variable. Total variance is composed of differences between children and error variance, which represents the variation within a child across informants. Some of the error component could be due to true scores changing across informants, some could be due to informant error, and some is unexplained. The ANOVA does not discriminate between these sources of error. In Equation 3-6, BMS is the between-subjects mean square, EMS is the error mean square, RMS is the between-informants mean square, and *k* is the numbers of informants.

$$\text{ICC} = \frac{\text{BMS} - \text{EMS}}{\text{BMS} + (k - 1)\text{EMS}} \tag{3-6}$$

Item Response Pattern of Parents and ECEs

For each ASQ item, informants could select one of the three provided response choices or omit the item. Across the ASQ items, the total number of times each possible response choice (e.g., yes, sometimes, not yet, or omitted) was selected was totaled for each informant type (i.e., parent and ECE). Research questions for which item response patterns were calculated are shown

in Table 3-2. To calculate the overall item response pattern for each informant type, Equation 3-7 was applied using $n = 126$.

Percentage of items with response choice = (3-7)

(Total number of items response choice was selected)/(n * 30) X 100

The divisor in the formula represented the total number of possible items across all participants, totaling 3780 items. To examine overall item response patterns by age cohort, Equation 3-7 was adjusted so that n represented the number of participants in each informant type for each age cohort.

To calculate item response patterns within each developmental domain, the total number of times each possible response choice (yes, sometimes, not yet, or omitted) was selected across the 6 items was totaled for each informant type (i.e., parent and ECE) using the formula shown in Equation 3-8.

Percentage of developmental domain items with response choice =

(Total number of items response choice was selected)/(n * 6) X 100 (3-8)

The divisor in the formula represented the total number of possible items in each developmental domain across all participants, totaling 756 items. Calculations were repeated for each developmental domain. This equation was also applied to examine developmental domain differences for each age cohort. Equation 3-8 was adjusted so that n represented the number of participants within each age cohort.

Predictors of Congruence

To address research question five, beta regression was used. Regression analysis is used when testing whether an independent variable (IV) predicts a dependent variable (DV; Younger, 1979). Ordinary least squares (OLS) regression is commonly applied to test the linear relationship between two variables. One of the primary assumptions in OLS regression is that the

conditional distribution of the DV is normal. This assumption is violated when the DV is a proportion (Smithson & Verkuilen, 2006), as is the case with percentages. When the DV is a proportion, the function relating the DV to the IV is not likely to be linear. In cases such as this, it is more appropriate to use beta regression in which the predicted proportions are between 0 and 1 (Simas, Barreto-Souza, & Roucha, 2010; Smithson & Verkuilen, 2006). Beta regression assumes that the residuals follow a beta distribution that is bounded by 0 and 1, which can be symmetric or skewed. Beta regression consists of two sub-models. The first sub-model tests the conditional mean of the DV while the second sub-model tests the precision parameter (Smithson & Verkuilen, 2006). Testing the conditional mean of the DV permits examination of whether the IV scores (e.g., child age and length of relationship) predict the proportions of the DV (e.g., item-level percent agreement). The test for the precision parameter asks whether the IVs predict the variance of the conditional beta distributions. Including the second sub-model allows for the fact that if the DV is related to the IV, the variance will also be related. As opposed to other analyses used in research questions two and three, child age was defined by number of months to provide a comparative unit of analysis to length of relationship with the ECE, also defined in months.

Table 3-1. Comparison of ASQ-3 standardization sample distribution with the present sample distribution by age cohort.

Age cohort	ASQ-3 standardization sample *n*(%)	Sample *n*(%)
2 – 12 months	7,158 (39%)	17 (14%)
14 – 24 months	5,743 (31%)	27 (21%)
27 – 36 months	3,064 (17%)	38 (30%)
42 – 60 months	2,607 (14%)	44 (35%)
Total *N*	18,572	126

Note. There are 21 ASQ-3 questionnaire intervals that are used for children from 1 – 66 months (Squires, Bricker, Twombly, & Potter, 2009). Each questionnaire indicates the age range applicable for use. Age cohorts were formed by referencing the grouping of age intervals used in validity studies of the ASQ-3.

Table 3-2. Number of children who were eligible for inclusion in the present study, whose parents provided consent, and who were included in analyses.

	Eligible Children *N*	Consent Received *n*(% of eligible)	Included in Analyses[a] *n*(% of eligible)
0 – 12 months	32	22 (69%)	17 (53%)
14 – 24 months	57	31 (54%)	27 (47%)
27 – 36 months	66	40 (61%)	38 (58%)
42 – 60 months	107	52 (49%)	44 (41%)
Total across ages	262	145 (55%)	126 (48%)

Note. [a]To be included in analyses, the parent and ECE must have completed the ASQ-3 within 2 weeks of each other. Five parent-ECE ASQ-3 pairs returned ASQ forms outside of the 2-week window and were not included in analyses. Fourteen parent-ECE pairs had incomplete forms, meaning that consent had been received from the ECE and parent, an ASQ-3 packet had been given, but either one or both of the packets were never returned.

Table 3-3. Summary of research aims by ASQ unit of analysis, variable of interest, and analysis procedure.

	ASQ Unit			Variable of Interest			Analysis			
Research Aim	Items	CD	Raw Score	Child Age	DD	LR	% agree	ICC	BR	% of RS
1a	✓						✓			
1b		✓					✓			
2a	✓			✓			✓			
2b		✓		✓			✓			
3a	✓				✓		✓			
3b		✓			✓		✓			
3c	✓			✓	✓		✓			
3d		✓		✓	✓		✓			
3e			✓		✓			✓		
4a	✓									✓
4b	✓			✓						✓
4c	✓				✓					✓
4d	✓			✓	✓					✓
5	✓			✓		✓			✓	

Note. CD = classification decisions; DD = developmental domain; LR = length of relationship between child and ECE; % agree = percent agreement between parent and ECE; ICC = intraclass correlation coefficient; BR = beta regression; % of RS = percent of response selection.

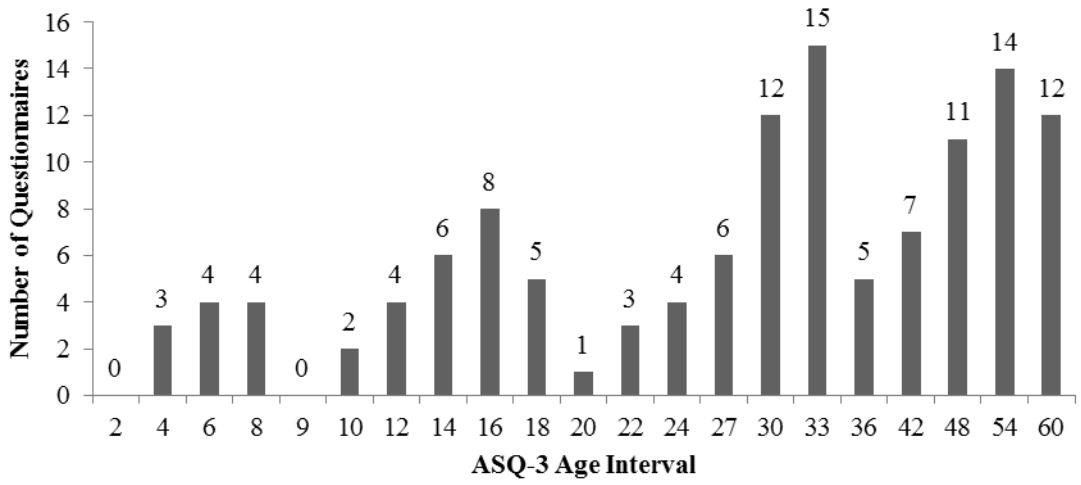

Figure 3-1. The number of completed questionnaires for each ASQ-3 age interval. There are 21 ASQ-3 questionnaire intervals that are used for children from 1 – 66 months (Squires, Bricker, Twombly, & Potter, 2009). Each questionnaire indicates the age range applicable for use.

CHAPTER 4
RESULTS

The purposes of the present study were to (a) examine overall congruence in the judgment-based ratings of two informants (i.e., parent and early childhood educator) when they each completed the same multi-dimensional developmental screening instrument for a child (Ages and Stages Questionnaire – 3 [ASQ-3]; Squires & Bricker, 2009), (b) identify whether congruence between the two informants varied based on child age cohort, (c) examine if congruence between the two informants varied by developmental domain included on the screening instrument (e.g., communication, fine motor, gross motor, problem solving, and personal-social), (d) compare patterns of responses between the two types of informants, and (e) explore whether child age and length of relationship with the early childhood educator predicted congruence. To conduct the study, parents and early childhood educators (ECEs) from two university-based early education and care centers were recruited to complete separately a developmental screening instrument for the same child. The multi-dimensional developmental screening instrument used was the ASQ-3. Of the 262 eligible children enrolled in the centers, signed informed consent forms were received for 145 children (54.1%). Questionnaires were included for analyses if parent-ECE pairs completed the forms within 2 weeks of each other. Five parent-ECE pairs returned ASQ-3 forms outside of the 2-week window and were not included in analyses. Fourteen parent-ECE pairs had incomplete forms, meaning consent had been received from the ECE and parent, an ASQ-3 packet had been given, but either one or both of the packets were not returned. ASQ-3 data for 126 parent-ECE pairs across the two centers were used for data analyses.

Overall Congruence Between Parent and ECE Ratings

Descriptive analyses were conducted to address the first research question related to congruence of parent and ECE judgment-based ratings on the ASQ-3. One analysis involved examining overall item-level agreement and the other overall classification agreement.

Overall Item-Level Agreement

Overall item-level agreement between parents and ECEs was calculated across all 30 ASQ items to generate a percentage agreement score. Item-level agreement was calculated in two ways: (1) by using the three original response choices (yes, sometimes, not yet) provided in the ASQ-3 (Squires & Bricker, 2009) and (2) by using two response choices generated by considering yes and sometimes as agreement. When parents-ECE pairs selected the same response for each ASQ item, a score of 1 was applied and when they selected responses that differed from each other a score of 0 was applied. If one informant omitted the item, this was also scored as 0. Scores were summed and then divided by 30 and multiplied by 100 to obtain a percent agreement score for each informant pair.

When examining congruence using the three response choices, overall percent agreement averaged across 126 parent-ECE pairs was 78% (*SD* = 14%). When calculating congruence using two response choices (i.e., yes/sometimes, not yet), overall percent agreement averaged across 126 parent-ECE pairs was 92% (*SD* = 9%).

Overall Classification Agreement

On the ASQ-3, a total domain score can be derived based on informant responses to each item (Squires et al., 2009). If an ASQ-3 item is marked yes a score of 10 is given, if the item is marked sometimes a score of 5 is given, and if the item is marked not yet a score of 0 is given. Scores for the six items in each domain are summed and can range from 0 to 60. Different total domain scores correspond to one of three classification categories (above the cutoff, monitoring

zone, below the cutoff) using age-based cutoffs provided on the summary forms. The classification categories are used to provide guidance about whether or not the child should receive further evaluation or monitoring.

In the present study, congruence about overall classification decisions referred to the average agreement between parents and ECEs for the classification categories across the five developmental domains included on the ASQ-3. Classification decision congruence was calculated for each developmental domain by applying a score of 1 when parents and ECEs agreed on the classification category and applying a score of 0 when they did not agree on the classification category. These scores were summed and then divided by five and multiplied by 100 to obtain an average classification percent agreement score across all ASQ domains. When calculating classification agreement using three categories (i.e., above the cutoff, monitoring zone, and below the cutoff), average percent agreement for the 126 parent-ECE pairs was 86%, *SD* = 21%. When calculating agreement using two categories (i.e., typical, eligible) created by considering above the cutoff and monitoring zone as agreement, average percent agreement for the 126 parent-ECE pairs was 93%, *SD* = 13%.

Parent and ECE Congruence by Child Age Cohort

Two types of descriptive analyses were conducted to address the second research question related to whether ASQ-3 score congruence between parents and ECEs varied based on child age cohort: overall item-level agreement by age cohort and overall classification agreement by age cohort. Similar to the analyses described for research question one, overall item-level agreement and overall classification agreement were examined separately. Variation in item-level and classification agreement based on child age was examined using four age cohorts, which were consistent with those used in validity and reliability studies reported in the ASQ-3

manual (Squires et al., 2009): (a) birth – 12 months, (b) 14-24 months, (c) 27-36 months, and (d) 42-60 months.

Overall Item-Level Agreement by Age Cohort

The distribution of the child sample in the present study across the four age cohorts is shown in Table 4-1. When examining congruence of parent-ECE ASQ-3 ratings across age cohorts using three item-level response choices (i.e., yes, sometimes, not yet), overall item-level agreement for each age cohort was (a) *M* = 83%, *SD* = 11% for children 42-60 months of age (*n* = 44); (b) *M* = 82%, *SD* = 13% for children 27-36 months of age (*n* = 38); (c) *M* = 71%, *SD* = 12% for children 14-24 months of age (*n* = 27); and (d) *M* = 69%, *SD* = 14% for children 0-12 months of age (*n* = 17). When examining congruence using two item-level response choices (i.e., yes/sometimes; not yet), overall item-level percent agreement for each age cohort was (a) *M* = 96%, *SD* = 6% for children 42-60 months of age; (b) *M* = 93%, *SD* = 9% for children 27-36 months of age; (c) *M* = 88%, *SD* = 8% for children 14-24 months of age; and (d) *M* = 86%, *SD* = 10% for children 0 to 12 months of age (Table 4-1).

Overall Classification Agreement by Age Cohort

To examine congruence related to classification agreement for each of the four age cohorts, the analytic procedures for overall classification agreement described above were applied to each of the four age cohorts. When classification decision congruence was examined using the three classification categories (i.e., above the cutoff, monitoring zone, and below the cutoff), the classification agreement percentages by age cohort were (a) *M* = 91%, *SD* = 15% for children 42-60 months of age; (b) *M* = 91%, *SD* = 16% for children 27-36 months of age; (c) *M* = 79%, *SD* = 23% for children 14-24 months of age; (d) *M* = 69%, *SD* = 27% for children 0-12 months of age. When congruence was calculated based on two classification categories (i.e., typical, eligible), classification agreement percentages by age cohort were (a) *M* = 96%, *SD* = 10% for

children 42-60 months of age; (b) *M* = 94%, *SD* = 12% for children 27 – 36 months of age; (c) *M* = 90%, *SD* = 17% for children 14-24 months of age; (d) *M* = 88%, *SD* = 16% for children 0 – 12 months of age. Table 4-2 shows data for classification agreement by age cohort.

Parent and ECE Congruence for Five Developmental Domains

Five descriptive analyses were conducted to address research question three related to whether congruence across parent-ECE judgment-based ratings on the ASQ-3 varied across developmental domains: (a) item-level agreement for each developmental domain, (b) classification agreement for each developmental domain, (c) item-level agreement by developmental domain within age cohorts, (d) classification agreement by developmental domain within age cohorts, and (e) intraclass correlation coefficients (ICCs) by developmental domain.

Item-Level Agreement by Developmental Domain

For each of the five developmental domains on the ASQ-3, congruence between parent-ECE ratings was evaluated by calculating a percent agreement score for the six items in each domain. The analytic methods used for overall item-level agreement were applied but the divisor was 6 rather than 30 (Equation 3-3). Agreement was calculated for three response choices (i.e., yes, sometimes, not yet) and two response choices (i.e., yes/sometimes, not yet). For the three response choices, average percent agreement for the 126 parent-ECE pairs for each domain were: (a) gross motor, *M* = 86% (*SD* = 16%), (b) problem solving, *M* = 79% (*SD* = 22%), (c) communication, *M* = 78% (*SD* = 22%), (d) fine motor, *M* = 75% (*SD* = 23%), and (e) personal-social, *M* = 73% (*SD* = 24%). For two response choices, average percent agreement for the 126 parent-ECE pairs for each domain were: (a) gross motor, *M* = 96% (*SD* = 9%), (b) problem solving, *M* = 92% (*SD* = 13%), (c) fine motor, *M* = 91% (*SD* = 15%), (d) personal-social, *M* = 91% (*SD* = 14%), and (e) communication, *M* = 90% (*SD* = 16%). Considering the response

choices of yes and sometimes as agreement resulted in the smallest increase in percent agreement across informant pairs for the gross motor domain when comparing with increases made in other developmental domains (Table 4-1).

Classification Agreement by Developmental Domain

For each developmental domain score on the ASQ-3, age-based cutoffs provided on the summary forms (Squires & Bricker, 2009) are used to determine the equivalent classification category (i.e., above the cutoff, monitoring zone, below the cutoff). At the domain level, percent agreement for parent-ECEs classification decision was either 0% or 100% for each parent-ECE pair. To examine classification congruence by domain, agreement was averaged across 126 parent-ECE pairs (Equation 3-5). When three classification categories were used, percent agreement for classification decision by domain was (a) problem solving, *M* = 88% (*SD* = 33%); (b) fine motor, *M* = 88% (*SD* = 33%); (c) gross motor, *M* = 87% (*SD* = 33%); (d) personal-social, *M* = 83% (*SD* = 38%); and (e) communication, *M* = 83% (*SD* = 38%). When two classification categories were used, percent agreement for each domain was (a) problem solving, *M* = 97% (*SD* = 18%); (b) fine motor, *M* = 95% (*SD* = 21%); (c) personal-social, *M* = 94% (*SD* = 23%); (d) communication, *M* = 90% (*SD* = 30%); and (e) gross motor, *M* = 89% (*SD* = 32%). Data related to classification congruence by domain are shown in Table 4-2.

Item-Level Agreement by Developmental Domain and Child Age Cohort

For the youngest age cohort (0 – 12 months), there were 17 parent-ECE pairs. For this age cohort, item-level percent agreement (three response choices) by developmental domain was (a) gross motor, *M* = 81% (*SD* = 13%); (b) fine motor, *M* = 73% (*SD* = 25%); (c) problem solving, *M* = 71% (*SD* = 21%); (d) communication, *M* = 62% (*SD* = 20%); and (e) personal-social, *M* = 58% (*SD* = 28%). When calculating item-level percent agreement for two response choices, percent agreement increased across all developmental domains, but the order of the last two

developmental domains changed, with communication having the lowest percent agreement. Complete results for item-level agreement by developmental domain and child age cohort are shown in Table 4-1.

The age cohort of 14 – 24 months had 27 parent-ECE pairs. For this age cohort, item-level agreement (three response choices) by developmental domain was (a) gross motor, *M* = 87% (*SD* = 18%); (b) fine motor, *M* = 75% (*SD* = 22%); (c) communication, *M* = 70% (*SD* = 20%), (d) problem solving, *M* = 61% (*SD* = 25%), and (e) personal-social, *M* = 61% (*SD* = 25%). When calculating item-level agreement for two response choices, percent agreement increased across all developmental domains, but the third (communication) and fourth (problem solving) developmental domains changed order from a-e above (Table 4-1).

The age cohort of 27 – 36 months had 38 parent-ECE pairs. For this age cohort, item-level agreement (three response choices) by developmental domain was (a) communication, *M* = 90% (*SD* = 16%); (b) gross motor *M* = 85%, (*SD* = 18%); (c) problem solving, *M* = 80% (*SD* = 18%); (d) personal-social, *M* = 80% (*SD* = 18%); and (e) fine motor, *M* = 74% (*SD* = 25%). When calculating item-level agreement for two response choices, percent agreement increased across all developmental domains and the order of developmental domains remained the same as a-e above (Table 4-1).

The age cohort of 42 – 60 months had 44 parent-ECE pairs. For this age cohort, item-level agreement (three response choices) by developmental domain was (a) problem solving, *M* = 92% (*SD* = 12%); (b) gross motor, *M* = 86% (*SD* = 15%); (c) personal-social, *M* = 81% (*SD* = 21%); (d) communication, *M* = 79% (*SD* = 23%) and (e) fine motor, *M* = 78% (*SD* = 23%). When calculating item-level agreement for two response choices, percent agreement increased across

all developmental domains and the order of developmental domains remained the same as a-e above (Table 4-1).

Classification Agreement by Developmental Domain and Child Age Cohort

For the 0 – 12 month age cohort, there were 17 pairs. For this age cohort, classification agreement (three categories) by developmental domain was (a) fine motor, *M* = 82% (*SD* = 39%); (b) communication, *M* = 76% (*SD* = 44%); (c) problem solving, *M* = 76% (*SD* = 44%); (d) gross motor, *M* = 65% (*SD* = 49%); (e) and personal-social, *M* = 47% (*SD* = 51%). When calculating classification agreement for two categories (i.e., typical, eligible), percent agreement by developmental domain was (a) problem solving, *M* = 100% (*SD* = 0%); (b) personal-social, *M* = 94% (*SD* = 24%); (c) fine motor, *M* = 88% (*SD* = 33%); (d) communication, *M* = 82% (*SD* = 39%); and (e) gross motor, *M* = 76% (*SD* = 44%). Classification agreement for two categories was higher across all developmental domains, but the order of the first (problem solving) and third (fine motor) developmental domains changed from a-e above.

For the age cohort of 14 – 24 months, classification agreement (three categories) by developmental domain was (a) gross motor, *M* = 89% (*SD* = 32%); (b) fine motor, *M* = 89% (*SD* = 32%); (c) problem solving, *M* = 74% (*SD* = 45%); (d) personal-social, *M* = 74% (*SD* = 45%); and (e) communication, *M* = 70% (*SD* = 47%). When calculating classification agreement for two categories, percent agreement by development domain was (a) fine motor, *M* = 93% (*SD* = 27%); (b) problem solving, *M* = 89% (*SD* = 32%); (c) communication, *M* = 89% (*SD* = 32%); (d) personal-social, *M* = 96% (*SD* = 19%); and (e) gross motor, *M* = 81% (*SD* = 40%). Classification agreement for two categories was higher across all developmental domains, but the order of the developmental domains changed when considering highest to lowest classification agreement percentages.

For the 27 – 36 month age cohort, classification agreement (three categories) by developmental domains was (a) communication, *M* = 95% (*SD* = 23%); (b) problem solving, *M* = 95% (*SD* = 23%); (c) personal-social, *M* = 92% (*SD* = 27%); (d) gross motor, *M* = 87%, (*SD* = 34%); and (e) fine motor, *M* = 87% (*SD* = 34%). When calculating classification agreement for two categories, percent agreement by developmental domain was (a) fine motor, *M* = 100% (*SD* = 0%); (b) problem solving, *M* = 97% (*SD* = 16%); (c) communication, *M* = 95% (*SD* = 23%); (d) personal-social, *M* = 92% (*SD* = 27%); and (e) gross motor, *M* = 87% (*SD* = 34%). Classification agreement for two categories was higher across all developmental domains, but the order of the developmental domains changed when considering highest to lowest classification agreement percentages.

For the 42 – 60 month age cohort, classification agreement (three categories) by developmental domain was (a) gross motor, *M* = 95% (*SD* = 21%); (b) problem solving, *M* = 95% (*SD* = 21%); (c) personal-social, *M* = 93% (*SD* = 26%); (d) fine motor, *M* = 91% (*SD* = 29%); and (e) communication, *M* = 82% (*SD* = 39%). When calculating classification agreement for two categories, percent agreement by developmental domain was (a) gross motor, *M* = 100% (*SD* = 0%); (b) problem solving, *M* = 100% (*SD* = 0%); (c) personal-social, *M* = 95% (*SD* = 21%), (d) fine motor, *M* = 95% (*SD* = 21%); and (e) communication, *M* = 91% (*SD* = 29%). Classification agreement for two categories was higher across all developmental domains and the order remained the same as a-e above. Results for classification agreement by developmental domain and age cohort using three and two classification categories are shown in Table 4-2.

Relationships between Parent and ECE Raw Scores on Developmental Domains

As noted previously, raw scores can be calculated for each ASQ-3 developmental domain by summing the item-level scores. A yes response to an item receives a score of 10, a sometimes response receives a score of 5, and a not yet response receives a score of 0. Scores on each

domain, therefore, range from 0 to 60 given there are six items associated with each of the five ASQ-3 domains. An intraclass correlation coefficient (ICC) was calculated for each developmental domain using the raw scores for parents and ECEs. A one-way ICC model for single measurement was selected for this analysis (Equation 3-6). Several models are available to calculate ICCs and the selection of model is based on the nature of the study. A one-way ICC model for single measurement was selected because each child is assessed by a set of different informants (i.e., unique sets of parent-ECE pairs) and informants were considered as if they were chosen randomly (Portney & Watkins, 2000). ICCs range from 0 to 1.00.

ICCs were as follows: (a) gross motor domain = .73, $F(1, 125) = 6.06$, $p < .001$; (b) communication domain =.64, $F(1, 125) = 5.32$, $p < .001$; (c) problem solving domain ICC = .56, $F(1, 125) = 3.73$, $p < .001$; (d) personal-social domain ICC = .54, $F(1, 125) = 3{,}34$, $p < .001$; and (e) fine motor domain ICC = .46, $F(1, 125) = 2.71$, $p < .001$. These results differ from those reported for overall item-level percent agreement by domain because the ICC accounts for chance agreement between parents and ECEs (Portney & Watkins, 2000) and also is based on correspondence between domain scores rather than on item-level responses.

Comparison of Frequencies and Percentages of Response Selection between Types of Informant

For each of the 30 items on the ASQ-3, there are three available response choices: yes, sometimes, and not yet. The informant may only omit 1 to 2 items in each domain. In these cases, the ASQ-3 provides adjusted scoring guidelines for each domain. If more than 2 items are omitted, the developmental domain cannot be scored (Squires et al., 2009). To address research question four related to frequency of response selections for each type of informant (i.e., parent or ECE), the number of times one of the four possible responses (i.e., yes, sometimes, not yet, and omitted*)* was selected was totaled. Frequencies were calculated by totaling the number of

times that a response choice was selected. Using these frequencies, percent of response selection was calculated across all of the 30 items, by age cohort, by developmental domain, and by developmental domain within age cohorts using Equations 3-7 or 3-8 as described in Chapter 3. In the following sections, percentages of response selection are reported to facilitate comparisons across age cohorts and the reader may refer to Table 4-3 for frequencies of response selection.

Frequency and Percentage of Response Selection for All ASQ Items

The frequency and percentage of item response selection was calculated for all 30 ASQ-3 items (Table 4-3 and Table 4-4). When comparing the frequency and percentages of responses selected across parents and ECEs, parents selected sometimes more often than ECEs (+2%) and omitted 20 more items (+.5%). ECEs selected *yes* more often than parents (+2%). Both parents and ECEs selected not yet 7% of the time (Table 4-4).

Frequency and Percentage of Response Selection by Age Cohort

The frequency and percentage of responses selected by each informant type within each age cohort was calculated using an adjusted version of Equation 3-7 by totaling the number of ASQ-3 items and multiplying by the number of participants in the age cohort (30 x *n* children in each age cohort). As shown in Tables 4-3 and 4-4, ECEs selected not yet more often than parents for children 0 – 12 months (+2%), 14 – 24 months (+1%), and 42 – 60 months (+1%) and parents selected not yet more often than ECEs for children 27 – 36 months (+2%). ECEs selected sometimes more often than parents for children 42 – 60 months (+1%) and parents selected sometimes more often than ECEs for children 0 – 12 months (+3%), 14 – 24 months (+4%), and 27 – 36 months (+3%). ECEs selected yes more often than parents for three age cohorts: children 0 – 12 months (+1%), 14 – 24 months (+3%), and 27 – 36 months (+5%). Both parents and ECEs selected *yes* for 88% of the 30 response opportunities (i.e., 30 ASQ-3 items) for children 42 – 60 months.

Parents were more likely than ECEs to omit items across all four age cohorts. For children ages 0 – 12, parents omitted 2 items and ECEs did not omit any. In the 14 – 24 month age cohort, parents omitted 4 items and ECEs did not omit any. For children ages 27 – 36 months, parents omitted 8 items and ECEs omitted 1 item, while for the 42 – 60 month age cohort, parents omitted 10 items and ECEs omitted 3 items. The patterns of parent and ECE omissions are shown in Table 4-3.

Frequency and Percentage of Response Selection by Developmental Domain

Frequencies and percentages associated with response selection patterns were calculated based on the number of items in each domain multiplied by the number of parent-ECE pairs (6 x 126 = 756) using Equation 3-8. As shown in Tables 4-3 and 4-4, ECEs selected not yet more often than parents in the communication (+4%) and gross motor (+1%) domains. Parents selected not yet more often than ECEs in the fine motor domain (+3%). Both parents and ECEs selected not yet for 8% of response opportunities in the problem solving domain. Parents selected sometimes more often than ECEs in four domains: communication (+2%), gross motor (+2%), problem solving (+3%), and personal-social (+3%). In the fine motor domain, parents and ECEs both selected sometimes for 12% of the response opportunities. In terms of selecting yes, parents chose this response more often than ECEs in the communication (+3%) domain. ECEs selected yes more often than parents in four domains: gross motor (+1%), fine motor (+5%), problem solving (+4%), and personal-social (+3%).

Overall, across developmental domains, parents omitted 24 items while ECEs omitted 4 items (Table 4-3). Parents and ECEs each omitted one item in the communication domain. ECEs omitted one item in fine motor and two items in personal-social. There were no similarities in the patterns of omissions by ECE informants. Comparatively, parents omitted more items in gross motor (+2 items), fine motor (+12 items), problem solving (+6 items), and personal-social (+3

items) domains. The most commonly omitted item among parents was from the fine motor domain: whether the child was stringing small items (4 times) and able to use child-safe scissors (2 times).

Frequency and Percentage of Response Selection by Developmental Domain across Age Cohorts

When examining the frequency of responses and percentages of response selection categories for the 0 – 12 month age cohort across all domains, ECEs selected not yet and yes more frequently than parents and parents selected sometimes more frequently than ECEs (Tables 4-3 and 4-4). When considering each domain, ECEs selected not yet more often than parents in the communication (+12%), problem solving (+2%), and personal-social domain (+5%). ECEs selected *yes* more often than parents in the gross motor (+5%) and fine motor (+9%) domain. The overall pattern of responses changed in the following ways when examined by developmental domain: parents selected not yet more often than ECEs in gross motor (+3%) and fine motor (+4%) and selected *yes* more often than ECEs in communication (+2%) and personal-social (4%). Overall percentage results and percentage results by developmental domain within this age cohort are shown in Table 4-4.

When examining the pattern of responses for the 14-24 month age cohort across all domains, parents selected sometimes more frequently than ECEs and ECEs selected not yet and yes more frequently than parents (Table 4-3). When examining response pattern percentages by developmental domain, the overall pattern changed in the following ways: (a) ECEs selected sometimes more often than parents in fine motor (+1%); (b) parents selected not yet more often than ECEs in fine motor (+1%) and problem solving (+4%); (c) parents selected yes more often than ECEs in gross motor (+1%) and communication (+5%); and (d) parents and ECEs selected not yet at equal percentages in personal-social (Table 4-4).

When examining frequencies and percentages of responses for the 27-36 month cohort across all domains, parents selected not yet and sometimes more often than ECEs and ECEs selected yes more often than parents (Tables 4-3 and 4-4). When examining response patterns by developmental domain, this pattern was consistent for fine motor and personal-social domains. The overall pattern changed in the following ways for the other developmental domains: ECEs selected not yet more often in gross motor (+1%) and problem solving (+4%) and parents and ECEs selected sometimes at equal percentages in the communication domain (Table 4-4).

When examining frequencies and percentages of response patterns across all developmental domains for the 42-60 month age cohort, ECEs selected not yet more often than parents and parents selected sometimes more often than ECEs. Parents and ECEs selected yes equally. As shown in Tables 4-3 and 4-4, ECEs selected not yet more often than parents in communication (+3%) and gross motor (+1%), ECEs selected sometimes more often in communication (+1%), fine motor (+1%), and problem solving (+2%), and parents and ECEs selected yes at equal percentages in the gross motor domain. Parents selected not yet more often than ECEs in fine motor (+1%) and personal-social (+1%), parents selected sometimes more often than ECEs in personal-social (+3%) and parents selected yes more often than ECEs in communication (+4%) and problem solving (+2%). Parents and ECEs selected sometimes at equal percentages in the gross motor domain.

Child Age and Length of ECE-Child Relationship as Predictors of Congruence

The fifth research question focused on whether child age and length of time the early childhood educator had known the child when the ASQ-3 was completed were noteworthy predictors of parent-ECE item agreement. To conduct exploratory analyses related to this question, beta regression was used and consisted of two sub-models. Model 1 tested whether the conditional mean of the dependent variable (i.e., overall item agreement between informants

expressed as a proportion) increased as the categories of the independent variable (i.e., age and length of time ECE knew child when ASQ-3 completed) increased. Model 2 tested the precision parameter: whether the dependent variable predicted the variance of conditional beta distributions. Child age and length of ECE-child relationship (independent variables, IVs) were measured in months and congruence (dependent variable, DV) was a proportion between 0 and 1.0. When length of ECE-child relationship was not included in the model, child age significantly predicted the conditional mean of parent-ECE congruence, $b = .016$, $t(124) = 3.94$, $p < .01$, and was a non-significant predictor of the conditional variance of the beta distribution, $b = .007$, $t(124) = 0.96$, $p = .337$. When child age was not included in the model, length of ECE-child relationship significantly predicted the conditional mean of parent-ECE congruence, $b = .028$, $t(124) = 3.50$, $p < .01$, and significantly predicted the conditional variance of the beta distribution, , $b = .060$, $t(124) = 3.57$, $p < .01$. When child age and length of ECE-child relationship were tested as predictors of parent-ECE congruence, child age ($b = .013$, $t(124) = 3.00$, $p < .01$,) and length of child-ECE relationship ($b = .018$, $t(124) = 3.03$, $p = .040$) significantly predicted the conditional mean of parent-ECE congruence. Estimating the precision parameter model with both variables showed that length of ECE-child relationship was a significant predictor ($b = .049$, $t(124) = 3.02$, $p < .01$) and child age was a non-significant predictor of the conditional variance of the beta distribution (Table 4-5).

Summary

The purposes of the present study were to examine overall congruence in the judgment-based ratings of two informants (i.e., parent and ECE) when they each completed the same multi-dimensional developmental screening instrument for a child (ASQ-3; Squires & Bricker, 2009) and whether there were variations in congruence due to developmental domain, child age cohort, or child age and length of relationship with the ECE. Overall item-level agreement between

parents and ECEs was generally high as was classification agreement. Variability was noted in item and classification agreement across child age cohorts. Overall item-level percent agreement was higher for parents and ECEs of preschool-aged children relative to infant and toddler age cohorts. Findings showed differences in item agreement across developmental domains and agreement was consistently lower for fine motor skills. Examination of response patterns showed some differences across parent and ECE informants, with parents selecting sometimes more often than ECEs and ECEs selecting yes and not yet more often than parents. Findings showed that the overall response pattern was generally consistent across developmental domains and age cohorts, but some changes in response pattern by developmental domain were noted. Child age and length of child-ECE relationship significantly predicted the conditional mean of parent-ECE congruence, expressed as a proportion of item agreement, but only length of child-ECE relationship was a significant predictor of the conditional variance of the beta distribution.

Table 4-1. *M*(*SD*) of item-level percent agreement by age cohort and developmental domain.

Developmental Domain	0 – 12 month age cohort (*n* = 17)	14 – 24 month age cohort (*n* = 27)	27 – 36 month age cohort (*n* = 38)	42 – 60 month age cohort (*n* = 44)	Total across age cohorts (*n* = 126)
			3 Response Choices		
Communication	62 (20)	70 (20)	90 (16)	79 (23)	78 (22)
Gross Motor	81 (13)	87 (18)	85 (18)	86 (15)	86 (16)
Fine Motor	73 (25)	75 (22)	74 (25)	78 (23)	75 (23)
Problem Solving	71 (21)	61 (25)	80 (18)	92 (12)	79 (22)
Personal-Social	58 (28)	61 (25)	80 (18)	81 (21)	73 (24)
Total across Domains	69 (14)	71 (12)	82 (13)	83 (11)	78 (14)
			2 Response Choices		
Communication	78 (18)	83 (17)	96 (11)	93 (15)	90 (16)
Gross Motor	91 (10)	96 (10)	95 (11)	97 (6)	96 (9)
Fine Motor	89 (16)	90 (14)	89 (17)	93 (13)	91 (15)
Problem Solving	87 (13)	85 (17)	93 (12)	98 (6)	92 (13)
Personal-Social	83 (20)	86 (14)	90 (14)	97 (7)	91 (14)
Total across Domains	86 (10)	88 (8)	93 (9)	96 (6)	92 (9)

Table 4-2. *M*(*SD*) classification agreement by age cohort and developmental domain.

Developmental Domain	0 – 12 month age cohort (*n* = 17)	14 – 24 month age cohort (*n* = 27)	27 – 36 month age cohort (*n* = 38)	42 – 60 month age cohort (*n* = 44)	Total across age cohorts (*n* = 126)
			3 Classification Categories		
Communication	76 (44)	70 (47)	95 (23)	82 (39)	83 (38)
Gross Motor	65 (49)	89 (32)	87 (34)	95 (21)	87 (33)
Fine Motor	82 (39)	89 (32)	87 (34)	91 (29)	88 (33)
Problem Solving	76 (44)	74 (45)	95 (23)	95 (21)	88 (33)
Personal-Social	47 (51)	74 (45)	92 (27)	93 (26)	83 (38)
Total across Domains	69 (27)	79 (23)	91 (16)	91 (15)	86 (21)
			2 Classification Categories		
Communication	82 (39)	89 (32)	95 (23)	91 (29)	90 (30)
Gross Motor	76 (44)	81 (40)	87 (34)	100 (0)	89 (32)
Fine Motor	88 (33)	93 (27)	100 (00)	95 (21)	95 (21)
Problem Solving	100 (0)	89 (32)	97 (16)	100 (0)	97 (18)
Personal-Social	94 (24)	96 (19)	92 (27)	95 (21)	94 (23)
Total across Domains	88 (16)	90 (17)	94 (12)	96 (10)	93 (13)

Table 4-3. Frequency of response selection by parents (P) and early childhood educators (ECE) on the Ages and Stages Questionnaire – 3.

	0 – 12 months (*n* = 17)		14 – 24 months (*n* = 27)		27 – 36 months (*n* = 38)		42 – 60 months (*n* = 44)		Total across Ages (*n* = 126)	
Response Selected	P	ECE	P	ECE	P	ECE	P	ECE	P	ECE
Communication Domain										
Not Yet	6	18	27	41	16	14	5	13	54	86
Sometimes	18	9	20	14	11	11	26	28	75	62
Yes	78	75	115	107	200	203	233	222	626	607
Omitted					1			1	1	1
Gross Motor Domain										
Not Yet	14	11	7	12	14	15	3	6	38	44
Sometimes	10	8	14	10	18	8	18	19	60	45
Yes	78	83	141	140	196	205	241	239	656	667
Omitted							2		2	0
Fine Motor Domain										
Not Yet	7	3	10	8	27	14	10	7	54	32
Sometimes	14	10	19	21	28	22	33	38	94	91
Yes	80	89	130	133	170	191	216	219	596	632
Omitted	1		3		3	1	5		12	1
Problem Solving Domain										
Not Yet	6	8	18	11	19	20	4	6	47	45
Sometimes	14	13	34	16	25	15	11	15	84	59
Yes	81	81	109	135	181	193	248	243	619	652
Omitted	1		1		3		1		6	0
Personal-Social Domain										
Not Yet	17	22	17	18	21	18	4	1	59	59
Sometimes	21	20	31	29	32	24	33	27	117	100
Yes	64	60	114	115	174	186	225	234	577	595
Omitted					1		2	2	3	2
Total across domains										
Not Yet	50	62	79	90	97	81	26	33	252	266
Sometimes	77	60	118	90	114	80	121	127	430	357
Yes	381	388	609	630	921	978	1163	1157	3074	3153
Omitted	2		4		8	1	10	3	24	4

Table 4-4. Percentage of response selection by parents (P) and early childhood educators (ECE) on the Ages and Stages Questionnaire–3.

	0 – 12 months (*n* = 17)		14 – 24 months (*n* = 27)		27 – 36 months (*n* = 38)		42 – 60 months (*n* = 44)		Total across Ages (*n* = 126)	
Response Selected	P	ECE	P	ECE	P	ECE	P	ECE	P	ECE
Communication Domain										
Not Yet	6	18	17	25	7	6	2	5	7	11
Sometimes	18	9	12	9	5	5	10	11	10	8
Yes	76	74	71	66	88	89	88	84	83	80
Omitted					<1			<1	<1	<1
Gross Motor Domain										
Not Yet	14	11	4	7	6	7	1	2	5	6
Sometimes	10	8	9	6	8	4	7	7	8	6
Yes	76	81	87	86	86	90	91	91	87	88
Omitted							1		<1	
Fine Motor Domain										
Not Yet	7	3	6	5	12	6	4	3	7	4
Sometimes	14	10	12	13	12	10	13	14	12	12
Yes	78	87	80	82	75	84	82	83	79	84
Omitted	1		2		1	<1	2		2	<1
Problem Solving Domain										
Not Yet	6	8	11	7	8	9	2	2	6	6
Sometimes	14	13	21	10	11	7	4	6	11	8
Yes	79	79	67	83	79	85	94	92	82	86
Omitted	1		1		1		<1		1	
Personal-Social Domain										
Not Yet	17	22	11	11	9	8	2	<1	8	8
Sometimes	21	20	19	18	14	11	13	10	15	13
Yes	63	59	70	71	76	82	85	89	76	79
Omitted					<1				<1	<1
Total across domains										
Not Yet	10	12	10	11	9	7	2	3	7	7
Sometimes	15	12	15	11	10	7	9	10	11	9
Yes	75	76	75	78	81	86	88	88	81	83
Omitted	<1		<1		1	<1	1	<1	1	<1

Note. Percentages were calculated based on the number of items in each domain multiplied by the number of participants in each age group.

Table 4-5. Predictors of parent-ECE item agreement.

	Model 1 Conditional Mean		Model 2 Precision Parameter	
Variable	*b*	95% CI	*b*	95% CI
Child Age	.016**	[.008, .024]	.007	[-.007, .021]
Length of ECE-Child Relationship	.028**	[.012, .044]	.060**	[.027, .093]
Combined Model				
Child Age	.013**	[.005, .022]		
Length of ECE-Child Relationship	.018*	[.001, .035]	.049*	[.017, .082]

Note.
* $p < .05$
** $p < .01$

CHAPTER 5
DISCUSSION

The primary aim of the present study was to evaluate congruence between ratings of two informants who spend significant amounts of time with young children (i.e., parent and ECEs) when they each completed the same multi-dimensional developmental screening instrument, the Ages and Stage Questionnaire-3 (ASQ-3; Squires & Bricker, 2009). A secondary aim was to examine whether congruence varied based on the developmental domain assessed, child age (by cohort or in months), or length of relationship between children and ECEs. To explore these aims, descriptive and exploratory analyses were used.

This chapter interprets findings of the present study and discusses them in the context of related literature. Limitations of the present study are described and implications for future research and practice are presented.

Overall Congruence Between Parents and ECEs Using the ASQ-3

Overall Item-Level Congruence

Overall item-level agreement was 78% for three response choices (yes, sometimes, and not yet) and 92% when calculated for two response choices (yes/sometimes, not yet). The latter agreement percentages are similar to those of Frankenburg, van Doornick, Liddell, and Dick (1976), which was the only research study located that examined congruence using item-level agreement between parents and examiners and focused on children of similar age to those in the present study. Item agreement, based on two response choices (yes, no), between the PDQ (Frankenburg, van Doornick, & Liddell, 1976) and the DDST (Frankenburg et al., 1975) was reported to be 93%. Unlike in the present study where both informants completed the ASQ-3 in the same way, in the Frankenburg et al. study, parents and examiners completed two different developmental screening instruments that used different administration methods (i.e., parent

questionnaire versus direct elicitation). The PDQ was created, however, through the selection of essential items from the DDST (Frankenburg, van Doornick, Liddell, & Dick, 1976). Item agreement in the Frankenberg et al. study was calculated by comparing the responses on the PDQ only to the corresponding item on the DDST.

In the present study, when using three versus two response choices to evaluate congruence, the majority of disagreements between parents and ECEs occurred between the response categories of yes and sometimes. Comparing the increase in overall item-level percent agreement when the response categories of yes and sometimes were considered an agreement revealed this response pattern. The increase from 78% to 92% agreement indicates that 14% of the disagreements between parent and ECE informants were between yes and sometimes. Only 8% of disagreements between these informant pairs were between response choices of yes/sometimes versus not yet. This finding shows almost twice as many disagreements occur between yes and sometimes as between yes/sometimes and not yet.

One possible explanation for the finding that agreement is higher when the response categories of yes and sometimes are combined could be that a child's phase of learning in relation to a particular developmental skill reflected in an ASQ-3 item might influence whether a parent or ECE scores yes versus sometimes. Four phases of learning have been described in the literature: acquisition, fluency, generalization, and adaptation (e.g., Haring, Lovitt, Eaton, & Hansen, 1978). In the acquisition phase, children are learning a new skill and are not yet able to use the skill consistently or with accuracy. During the fluency phase, children are accurate in their responses but need repeated practice to improve the speed of responding. When children are acquiring or becoming fluent with a skill, these skills might be characterized as developing skills and, on the ASQ-3 might be scored as either not yet or sometimes occurring (Squires et al.,

2009). Fluency sets the occasion for the generalization phase of learning (Haring et al., 1978). In this phase, children are accurate and fluent in using the skill and are beginning to use the skill consistently across contexts or people. Developmental skills in this phase might be indicated on the ASQ-3 as disagreements between yes and sometimes (Squires et al., 2009), because a skill is emerging and beginning to generalize across contexts or people. Developmental skills in the last phase, adaptation, reflect a level of mastery where the child is learning how to adjust the skill for novel tasks or situations (Haring et al., 1978). Developmental skills in this phase may be indicated on the ASQ-3 by responses of *yes* by both informants (Squires et al., 2009), because the child has mastered the skill across contexts or people and so both informants should share the same perspective.

Given the possible relationship of the phases of learning to ASQ-3 (Squires & Bricker, 2009) response choices, the finding that the majority of disagreements between parents and ECEs occur between yes and sometimes suggest that differences in perspectives might be primarily occurring for emerging developmental skills. Although disagreements about child development may often instigate concern, it is important to obtain multiple perspectives to inform judgments or decisions about child development and learning across contexts and people, rather than ensuring exact agreement about skill performance across informants (Suen et al., 1992). Pursuing exact agreement could negate the rich perspectives of each informant. Thus, the knowledge that parents and ECEs typically do not agree on children's emerging developmental skills can inform the basis for important conversations between parents and ECEs about children's current phase of learning and how opportunities for learning can be provided across contexts.

Using an instrument such as the ASQ-3 completed by parents and ECEs to inform discussions about similarities or differences in skills demands or expectations across at least two contexts where many young children spend significant time (i.e., home and early care and education program) might lead to the formation of a multiply linked mesosystem and support young children's continued skill development (Bronfenbrenner, 1974, 1979). It is particularly beneficial for these types of conversations between parents and ECEs to be shared discussions in which each participant offers perspectives about children's skill development and provides information about similarities as well as differences in developmental skill demands and expectations across contexts, reaching consensus when appropriate. For example, a parent might score on the ASQ-3 that a child sometimes drinks from a cup unassisted, while an ECE might score that a child drinks from a cup unassisted (i.e., response choice of yes). Discussions between the parent and ECE reveal that at home the child is not always given an opportunity to drink from a cup unassisted given other parental and child demands or expectations at mealtimes. In the classroom, however, children are always provided a cup during mealtimes and they are expected to drink independently from the cup. In this particular example, the parent and ECE could discuss whether or not it is appropriate to align expectations for independent cup drinking across contexts but there is no need to be concerned about the child's ability to drink from a cup independently.

Overall Classification Congruence

The ASQ-3 is primarily used to determine which children might need monitoring or further evaluation. This purpose focuses on classification decisions. In the present study, overall classification agreement between parents and ECEs was 86% for three classification categories (above the cutoff, monitoring zone, below the cutoff) and was 93% when two classification

categories were used (eligible, typical). The terms eligible and typical were used to facilitate comparisons to validity studies reported in the *ASQ-3 User's Guide* (Squires et al., 2009), and correspond to recommendations for whether or not to refer a child for further monitoring or evaluation. Eligible indicates that a child's scores fall below the normative cutoff and it is recommended that the child should be referred for further evaluation. Typical indicates that a child's scores fall in the normative range and no additional action is needed. Although overall item agreement congruence (described above) was lower when using three response choices (78%), the disagreements that occurred between parents and ECEs when rating discrete developmental skills did not result in clinically significant differences in whether or not to refer a child, as shown by 93% agreement between parents and ECEs in identifying children as eligible versus typical.

Using three versus two categories to evaluate classification congruence, comparisons showed that parent-ECE disagreements occur at similar percentages for children broadly demonstrating lower developmental skills (i.e., disagreements between eligible and typical) and higher levels of ability on developmental skills (i.e., disagreements between above the cutoff and monitoring zone). Comparing the increase in percent agreement when the classification categories of above the cutoff and monitoring zone were considered an agreement revealed this classification pattern. The increase from 86% to 93% agreement indicates that 7% of the disagreements were between above the cutoff and monitoring zone, reflecting disagreements on children who broadly demonstrate higher developmental abilities or a greater number of emerging or mastered skills (Haring et al., 1978). The remaining 7% of disagreements were between eligible and typical, which might reflect children who may broadly demonstrate fewer

developmental skills in one context versus another or a greater number of developing or emerging skills that are not seen consistently across contexts.

In the validity studies reported in the *ASQ-3 User's Guide* (Squires et al., 2009), two classification categories (typical, eligible) were used to examine classification agreement between parents and examiners. The authors reported 93% classification agreement when administering the same instrument to both informants. The classification agreement percentage was the same in the present study when two classification categories were used. In the congruence literature reviewed in Chapter 2, classification agreement between two informants ranged from 59% (Frankenburg, van Doornick, Liddell, & Dick, 1976) to 97% (Bricker et al., 1988). When considering only studies that used informant procedures similar to those in the present study and that were conducted with the ASQ, classification agreement ranged from 87% – 97% (Bricker & Squires, 1989b; Bricker et al., 1988; Squires et al., 1997). Despite the fact that these ASQ studies were based on a comparison between parents and examiners who did not have extended contact with the child, classification agreement percentages in these studies were similar to those reported in the present study. Other than differences in the sample of professionals and the length of time they had known the child, there were not differences between the present study and prior ASQ studies with respect to the procedures used to gather data (i.e., an ASQ completed by two different informants within 2 weeks of each other). Similarities in overall classification decision agreement when two categories (typical, eligible) were used in previous studies and in the present study suggest that this type of agreement might not be significantly affected by the length of time an ASQ informant has known the child.

Two studies included in the Chapter 2 literature review examined overall classification agreement between parent and teacher: studies on the PDQ (Frankenburg, van Doornick, &

Liddell, 1976) and the R-PDQ (Frankenburg, 1987). Both studies used two classification categories (fail and pass) to indicate whether or not a child should receive further evaluation or support. Parent-teacher classification agreement on the PDQ was reported at 59% (Frankenburg, van Doornick, Liddell, & Dick, 1976) while parent-teacher classification agreement on the R-PDQ was reported as 83% (Frankenburg et al., 1987). These results reflect lower classification agreement than was found in the present study for three and two classification categories (86% and 93%, respectively). The teachers sampled in both of the Frankenburg studies were from various day-care centers, Head Start programs, and kindergartens in the community and data describing the characteristics of the teachers, including their experience, were not available (Frankenburg et al., 1987; Frankenburg, van Doornick, Liddell, & Dick, 1976). It was therefore not possible to compare characteristics of ECEs in the present study with characteristics of teachers in the Frankenburg studies in relation to differences in overall classification agreement percentages.

Exploring Variations in Congruence of Parent-ECE Ratings on the ASQ-3

A secondary aim of the present study was to explore if parent-ECE congruence varied by developmental domain, child age (by cohort and in months), or length of child-ECE relationship. For congruence by developmental domain and child age cohort, the descriptive analyses used were item-level agreement, classification agreement, and examination of response patterns across informant types. Exploratory analyses included intraclass correlation coefficients to examine congruence by developmental domain raw scores and beta regression to examine predictors of congruence.

Item-Level Congruence by Developmental Domain

Congruence for each ASQ-3 developmental domain was examined by determining agreement or disagreement at the item level between parents and ECEs for the six items within each developmental domain, calculating percent agreement for each informant pair, and averaging percentages across the 126 parent-ECE pairs. Findings from the present study showed congruence varied across ASQ-3 developmental domains. Examining percentage agreement for the six items associated with each of the five ASQ-3 developmental domains, parents and ECEs had the highest agreement on the gross motor domain (86%), followed by the problem solving (79%), communication (78%), fine motor (75%), and personal-social (73%) domains, respectively. Of the score congruence studies reviewed in Chapter 2, only one study examined item-level agreement by developmental domain (Coghlan et al., 2003). In the Coghlan et al. study, parents and day care providers completed the PEDS (Glascoe, 2007b), which had 10 items that each represented a different domain of development. There were three possible response choices (no, a little, and yes) for each PEDS item. Coghlan et al. (2003) reported gross motor (95%) and problem solving (cognitive, 92%) had the highest item-level agreement compared to other developmental domains. Comparisons between PEDS items and other ASQ-3 domains are less straightforward because the PEDS reports concerns for expressive and receptive language separately whereas these are represented on the ASQ-3 as one developmental domain (e.g., communication). The percentages for item agreement reported in Coghlan et al. (2003) ranged from 78% to 94% across the 10 items or domains, which is slightly higher than the range reported in the present study across the five developmental domains represented on the ASQ-3 (73% - 86%). The differences might be attributed to having only 1-2 items comprise a developmental domain on the PEDS (Glascoe, 2007b) compared to six items per developmental

domain on the ASQ-3 (Squires & Bricker, 2009). Nevertheless, despite differences in the number of items and types of items per domain across the PEDS and ASQ-3, in both the present study and the Coghlan et al. study, agreement between parents and care providers/ECEs was highest for gross motor and problem solving (cognition) relative to other developmental domains.

In the present study, average overall item agreement was higher across all developmental domains when the response choices of yes and sometimes were considered an agreement. Comparing item agreement for three versus two response choices, the largest increases occurred for personal-social (+18%) and fine motor domains (+16%), which indicate the percent of disagreements on emerging skills (yes versus sometimes). Parents and ECEs disagreed on developing skills (yes/sometimes versus not yet) 9% of the time for both personal-social and fine motor domains. For these two developmental domains, parents and ECEs disagreed on emerging skills almost twice as often as on developing skills. As noted previously, emerging skills are likely to result in disagreements on the ASQ-3 item responses between sometimes and yes since children have acquired a skill but might not be consistently using the skill across contexts or people (Haring et al., 1978).

Classification Congruence by Developmental Domain

Congruence, as measured by classification agreement percentages, also varied across ASQ-3 developmental domains, Using three classification categories, agreement was 88% for fine motor and problem solving, 87% for gross motor, 83% for communication, and 83% for personal-social. When the classification categories of above the cutoff and monitoring zone were considered as agreement, classification agreement increased by 7% - 9% for all developmental domains except gross motor. For other developmental domains, calculating classification agreement using two categories increased percentages to 97% for problem solving, 95% for fine

motor, 94% for personal-social, and 90% for communication. Classification agreement for gross motor increased from 87% to 89%.

Although average item-level agreement percentages were lower for fine motor and personal-social domains compared to other ASQ-3 developmental domains, the lower item-level agreement percentages were not associated with lower classification agreement percentages for these same two domains (i.e., 95% classification agreement for fine motor and 94% for personal-social).

The results reported above were compared to the only score congruence study on classification agreement by developmental domain. Chiu and DiMarco (2010) compared the responses of mothers on the ASQ-2 (Bricker & Squires, 1999) to a nurse's administration of the Denver II (Frankenburg et al., 1992) and calculated classification agreement using the categories of pass and fail. Classification agreement was reported for four of the five ASQ-2 developmental domains given the Denver II did not have a comparable domain for problem solving. Classification agreement results reported by Chiu and DiMarco were 95% for personal-social and gross motor, 71% for fine motor, and 67% for language (communication).

The classification findings reported by Chiu and DiMarco for personal-social were the same as those reported for the personal-social domain in the present study, while results for gross motor were only slightly higher than the present study (95% compared to 89%). Two findings reported by Chiu and DiMarco (2010) were different from those reported in the present study. First, across the four developmental domains compared in the Chiu and DiMarco study, language (communication) had the lowest classification agreement at 67%, which is substantially lower than reported in the present study (90%). There are several possible explanations for this difference. The authors hypothesized that low classification agreement was observed for

language skills because mothers can understand their children's utterances even when adults who do not spend significant time with the child cannot. Higher classification agreement for communication might have been obtained in the present study given the study sample included ECEs who observe and interact with the children daily, sometimes up to 40 hours per week, compared to a nurse who is only interacting with the child at a single point in time. Second, Chiu and DiMarco reported classification agreement of 71% for fine motor, which is lower than reported in the present study (95%). They reported mothers' scores identified children for further evaluation of fine motor concerns more often than the nurse.

Divergent results from the present study may also be due to measurement differences between the ASQ-2 (Bricker & Squires, 1999) and the Denver II (Frankenburg et al., 1992) given that the ASQ-2 is completed as a parent questionnaire while the Denver II is completed by direct elicitation. Moreover, although the developmental domains represented across the two instruments are similar, the items that comprise each of the developmental domains are different, which likely precluded Chiu and DiMarco from examining item-level agreement as was done in the present study.

Item-Level Congruence by Age Cohort

Congruence, as defined by overall item-level agreement, may vary based on child age cohort. Interpretations offered in this section should be considered preliminary given the small numbers of infants and toddlers in the present study. Average percent agreement across ASQ-3 domains for the two preschool age cohorts was 82% and 83%, respectively and for the infant and toddler age cohorts was 69% and 71%, respectively. For each of the age cohorts, comparisons were made for overall item-level agreement percentages using three versus two response choices to evaluate congruence. When the response categories of yes and sometimes were considered as

agreement, item agreement increased by 17% for the infant and toddler age cohorts and increased by 11% and 13% for the preschool age cohorts. Thus, item agreement for two response choices was as follows: 86% for infants, 88% for toddlers, and 93% and 96% for the preschool age cohorts. Considering the large increase in percentages when yes and sometimes are considered as agreement, results show that parents and ECEs are not in agreement when rating emerging skills that are children are beginning to generalize across contexts (Haring et al., 1978). The large increase in item agreement when using two response choices is consistent across all age cohorts, but is more predominant for the infant and toddler age cohorts. None of the studies reviewed in Chapter 2 focused on congruence investigated differences in parent-professional item-level congruence by age cohorts.

Within each of the four age cohorts, congruence varied by developmental domain, with greater variability across developmental domains for infant and toddler age cohorts than preschool age cohorts. Average item agreement between parents and ECEs in the infant age cohort varied by 23% across ASQ-3 developmental domains (range = 58% – 81%) and by 26% for the toddler age cohort (range = 61% – 87%). For each of the preschool age cohorts, average item agreement varied by 16% for the 27 – 36 month age cohort (range = 74% - 90%) and by 9% for the 42-60 month age cohort (range = 78% to 92%).

The finding that there is greater variability across developmental domains for infants and toddlers is not unexpected when considering the progression of child development. Children's growth consists of multiple, complex processes and in order to effectively monitor the development of these processes, they are often classified into developmental domains (Bagnato, 2007). Although the use of categories implies discrete divisions between developmental domains, skills can involve processes across multiple domains. For example, the ability to

"…pick up a toy and put it in his mouth" is aimed to sample a problem solving skill related to exploration of objects (Squires & Bricker, 2009), but also samples the fine motor/perceptual-motor skills of grasping objects and hand-eye coordination. During the infant years, skill development appears undifferentiated and globalized due to the immaturity of the nervous system, sensory abilities, and motor functions (Bagnato, 2007). As children grow, the progression of skill development reflects an increasing refinement and specialization of behaviors, and thus, developmental skills are more discrete and might be easier to identify.

Compared to preschool age cohorts, infants and toddlers may have lower overall item congruence and greater variability across developmental domains because developmental domains are highly interrelated during infant and toddler years, and therefore, it may be more difficult to identify discrete skills. Additionally, for some developmental skills, there are precise differences that mark the progression of skill development that may be difficult for informants to identify or agree on the frequency of its occurrence. For example, on the ASQ-3, items associated with several age intervals ask informants to observe the way a child picks up a small piece of food (i.e., Cheerio), however, earlier age intervals ask about whether a raking motion with all fingers is being used while later age intervals ask about whether the tips of the thumb and fingers are being used (Squires & Bricker, 2009). As skills become more discrete or differentiated as children age, congruence between scores used by parents and ECEs might be higher than at earlier ages.

Within the infant and toddler age cohorts, item-level agreement was higher for the ASQ-3 gross motor domain (81% infant age cohort and 87% toddler age cohort), compared to other ASQ-3 developmental domains (see Table 4-1). This finding might be related to the types of skills reflected on ASQ-3 items in the gross motor domain for infants and toddlers. The skills are

discrete (e.g., sits up straight for several minutes) and might be easier to rate when compared to skills in other developmental domains (e.g., copying activities that you do in the personal-social domain). Additionally, differences across developmental domains and age cohorts may be due to the sequence and timing of development, with certain developmental domains assuming prominence during particular ages. Thus, parents and ECEs might be focused on the development of certain skills at particular ages leading to similar perspectives. For example, between approximately 6– 15 months of age, children are beginning to crawl, sit, pull themselves up, stand, and walk (Bagnato, 2007), which are developmentally relevant as well as observable skills reflected on the ASQ-3. For infant and toddler age cohorts, higher item-level agreement for the gross motor domain relative to other domains might be explained by parents and ECEs heightened sensitivity to the developmental milestones included in the gross motor items on the ASQ-3.

Item agreement between parents and ECEs for infant and toddler age cohorts in the personal-social skills domain was 58% and 61%, respectively. One possible explanation for this relatively low level of agreement across informants might be the content of the items in the personal-social domain in the ASQ-3, as well as the number of children cared for in an infant or toddler classroom versus a home environment. The personal-social domain includes items related to behaviors that occur before and during feedings, diaper changes, and dressing (Squires et al., 2009). Within a classroom, as many as 8 infants between the ages of birth – 15 months could be enrolled. NAEYC (2012) recommends a staff-to-child ratio for a group this size to be 1:4 and this ratio changes to 1:3 if there are 6 or fewer children in the classroom. Even when accounting for the required number of adults in the classroom, this is a large number of same-aged children to care for, particularly given the caretaking responsibilities and level of supervision needed for

infant and toddler age cohorts. Within the home, it is possible that there is only one child in the home, and even if there are siblings, a child may be the only one in his or her age range. Thus, it is possible that there are differing amounts of support and attention available to each child depending on the context. ECEs may be less focused on individual child behaviors related to personal-social skills as they are simultaneously conducting feedings, diaper changes, and dressing with other children.

Within the younger preschool cohort (27 – 36 months), item agreement was higher for the communication domain (90%) relative to the other ASQ-3 developmental domains (Table 4-1). Additionally, item agreement for communication was also higher than reported for the other three age cohorts of infants, toddlers, and older preschoolers (62%, 70%, and 79%, respectively). Relatively high congruence for the communication domain may be due to parents and ECEs heightened sensitivity to these developmental skills between the ages of 27 – 36 months because children at this age are becoming increasingly verbal, including the rapid acquisition of new vocabulary and increasing length of phrases and sentences (Bagnato, 2007). For this age cohort, one possible explanation for higher item-level agreement for the communication domain relative to other domains might be parents and ECEs heightened sensitivity to developmental milestones reflected in the communication domain, leading to more similar perspectives. It may also be that the content of the communication domain for this age cohort reflects discrete skills that are easier to observe and to arrive at similar perspectives.

For older preschoolers (42 – 60 months), item agreement between parents and ECEs for the ASQ-3 problem-solving domain was 92% compared to agreement for the infant, toddler, and younger preschool cohort (71%, 61%, and 80%, respectively). This finding might be explained by the heightened emphasis on pre-academic skills for this age cohort with items on the ASQ-3

focused on pre-academic concepts such as numbers, letters, shapes, and sizes (Squires & Bricker, 2009).

Across all age cohorts, item agreement percentages were relatively consistent for the gross motor domain (range = 81%-87%) and the fine motor domain (range = 73%-78%). For the preschool-age cohorts, fine motor was the developmental domain with the lowest item-level agreement. This may be attributed to differing demands and expectations across contexts for fine motor skills. The early education and care centers involved in the present study frequently planned activities in which to observe and support fine motor skill development (i.e., drawing, writing, stringing beads), particularly at the preschool ages when these skills are considered important for school readiness. There were several parents who omitted the ASQ-3 item asking about stringing beads (Squires & Bricker, 2009), which is not an activity that often occurs naturally in the home. Another parent omitted an item asking about how well a child cuts on a line (Squires & Bricker, 2009) and wrote in an explanation that they did not have child-safe scissors in the home. When considering the number of opportunities children have at early education and care centers to learn and refine fine motor skills reflected on the ASQ-3, the finding of slightly lower agreement percentages between parents and ECEs compared to other domains was not unexpected.

Classification Congruence by Age Cohort

Congruence, as measured by overall classification agreement, varied based on child age cohort, with higher percent agreement for preschool age cohorts than infant and toddler age cohorts. Preschool age cohorts each had 91% classification agreement between parents and ECEs, while agreement between these two sets of informants was 69% for infants and 79% for toddlers. Comparisons were made for classification agreement for three versus two classification

categories. Across all four age cohorts, calculating classification agreement using two categories (e.g., eligible and typical) resulted in congruence ranging from 88% - 96%. When comparing classification agreement of three versus two categories, the largest increase in percent agreement occurred in the infant cohort (69% to 88%), indicating that 19% of the disagreements occurred between above the cutoff and monitoring zone, with the remaining 11% of disagreements between eligible and typical. Referencing the phases of learning, this may indicate that parents and ECEs are primarily disagreeing on children who broadly demonstrate higher developmental abilities or have more emerging skills (Haring et al., 1978). As reported in the previous section, parents and ECEs of the infant age cohort had the lowest overall item-level agreement at 69%. While this result might warrant some concern, results of classification agreement show that low overall item-level agreement generally did not result in clinically significant differences in whether or not to refer a child, as shown by 88% classification agreement when using two classification categories.

For the toddler and preschool age cohorts, using three versus two classification categories to evaluate congruence shows that parent-ECE disagreements occur at similar percentages for children broadly demonstrating lower developmental skills (i.e., disagreements between eligible and typical) and higher levels of ability on developmental skills (i.e., disagreements between above the cutoff and monitoring zone). As described in previous sections, this was determined by comparing the increase in percent agreement when above the cutoff and monitoring zone were considered agreement. None of the studies reviewed in Chapter 2 focused on congruence investigated differences in parent-professional classification congruence by age cohorts.

Within each age cohort, there was greater variability in classification agreement across developmental domains for infant and toddler age cohorts (35% and 19%, respectively) than

preschool age cohorts (8% and 13%, respectively). As described earlier, skill development appears undifferentiated and globalized during the infant years due to the immaturity of the nervous system, sensory abilities, and motor functions (Bagnato, 2007). As children grow, the progression of skill development reflects an increasing refinement and specialization of behaviors, and thus, developmental skills are more discrete and may be easier for informants to identify. When developmental skills are discrete and easily observable, it is likely that perspectives across informants will be similar. As noted previously, however, these preliminary infant and toddler cohort findings should be interpreted with caution given the small number of infants and toddlers in the present study.

For each age cohort, variability in classification agreement percentages across developmental domains decreased when three versus two classification categories were used. Additionally, using two classification categories resulted in several instances in which parents and ECEs had perfect agreement on classification: problem solving for the infant (0 - 12 month) and older preschool (42 – 60 month) cohorts, fine motor for the young preschool (27 – 36 month) cohort, and gross motor for the older preschool cohort. Classification agreement for two categories was generally high across all age cohorts and developmental domains with one exception. Parents and ECEs of children 0 – 12 months disagreed on classification decisions (e.g., eligible versus typical) more often in the gross motor domain (76% classification agreement) compared to other developmental domains. Using three versus two classification categories to evaluate congruence, comparisons showed that parent-ECE disagreements occurred more often for children broadly demonstrating lower developmental skills than for children demonstrating higher levels of ability on developmental skills. Comparing increases in percent agreement when the classification categories of above the cutoff and monitoring zone were

considered an agreement revealed this pattern. The increase from 65% to 76% indicates that 9% of the disagreements were between above the cutoff and monitoring zone, reflecting disagreements on children who broadly demonstrate higher developmental abilities or a greater number of emerging or mastered skills (Haring et al., 1978). The remaining 24% of disagreements were between eligible and typical, reflecting children who may broadly demonstrate lower developmental ability or a greater number of developing or emerging skills.

Relationships between Parent and ECE Raw Scores on Developmental Domains

The summary scores of parents and ECEs were compared for each developmental domain. On the ASQ-3, summary scores for each developmental domain are derived by adding scores for ratings on each of the six items, so that total scores range from 0 to 60 (Squires et al., 2009). Intraclass correlation coefficients (ICCs) were used to compare summary score ratings of each parent-ECE pair. ICCs ranged from .46 (fine motor) to .73 (gross motor) and F-tests showed statistically significant differences in informants' mean scores for all five developmental domains ($p < .001$). As two groups of informants, the ICCs suggest fair to moderate relationships between raw scores although the mean scores differed across type of informant.

The ICC findings in the present study are comparable to the range of ICCs reported in the *ASQ-3 User's Guide* (Squires et al., 2009). The ASQ-3 authors compared parent and examiner ratings on the ASQ-3 for 107 children, ages unspecified. Examiners completed the ASQ-3 after administering the BDI-2 (Newborg et al., 2004). Squires et al. reported ICCs ranging from .43 (communication) to .69 (personal-social), but did not specify the ICCs for gross motor, fine motor, and problem solving. Although the range of ICCs is similar to the range reported in the present study, the developmental domains associated with the upper and lower limits differ: fine motor had the lowest ICC in the present study compared to communication in Squires et al., while gross motor had the highest ICC in the present study compared to personal-social in

Squires et al. As described in previous sections, fine motor may have a lower congruence between parents and ECEs due to differing demands across contexts, with more opportunities to observe and support fine motor skills occurring in early education and care centers. Additionally, the higher ICCs for communication domain compared to the other ASQ-3 domains might have been obtained in the present study given the study sample included ECEs who observe and interact with the children daily, compared to an examiner who is only interacting with the child at a single point in time.

Examination of Informant Response Patterns

Overall average item-level congruence between parents and ECEs across all age cohorts and ASQ-3 domains was 78%. Early research in score congruence typically interpreted disagreements between parents and professionals as being attributed to parents' tendency to "overestimate" their child's ability in relation to skills reflected on developmental assessment instruments (Johnson, Poteat, & Kushnick, 1986; Knobloch, Stevens, Malone, Ellison, & Risemberg, 1979; Miller, Manhal, & Mee, 1991). In the present study, comparisons of response patterns across informant types were conducted to provide additional information about the correspondence between parent and ECE perspectives.

The frequency and percentage of the selection of a response choice (yes, sometimes, not yet, and omitted) was calculated for each informant type. Results of these analyses showed differences in the pattern of response selection across parents and ECEs. Compared to parents, ECEs selected yes and not yet more often, while parents selected sometimes more often when compared to ECEs. To support the premise that parents were "overestimating" children's ability, response patterns should correspond to parents frequently indicating response choices suggesting mastery of developmental skills (i.e., yes or sometimes) with ECEs selecting response choices suggesting developing or emerging skills (i.e., sometimes or not yet). Instead, the present study

showed that ECEs selected response choices at the extremes while parents frequently selected the response choice (sometimes) that reflects emerging skills. Parents might be more likely to select sometimes because they have more opportunities to observe children's activities and interactions across multiple contexts or people while ECEs select a response choice (either yes or not yet) that is representative of only one context. Alternatively, parents might select sometimes more often because they are unsure if there child demonstrates the skill across contexts and people or they might not understand completely the intent of the ASQ-3 item in relation to what differentiates sometimes from yes. Regardless of which interpretation is considered, it is important to note that these findings represent global response patterns across the 126 informant pairs, not the differences in response patterns across informants for individual children.

Item-level agreement between parents and ECEs for the fine motor domain ranged from 73% to 78% across the four age cohorts. Comparisons of item-level agreement for three versus two response choices showed that when agreement did not occur between parents and ECEs, it was primarily related to emerging fine motor skills. This finding is further supported by examination of response patterns for this developmental domain, which differed from the overall response pattern reported above. Consistent with the item-level agreement findings, response patterns showed ECEs selected yes more often than parents for fine motor items while parents selected sometimes more often than ECEs for the fine motor items. In addition, for fine motor domain items, response patterns showed that parents selected not yet more often than ECEs and omitted responses most often in the fine motor domain relative to the other four domains (i.e., 12 of the 22 omissions occurring in the fine motor domain). Additionally, differences in response patterns across informant types for the fine motor domain were slightly larger than for other developmental domains, which had differences across informant types ranging from 0% - 4%

(Table 4-4). In the fine motor domain, ECEs selected yes for 84% of responses compared to 79% of responses for parents, which is a difference of 5%. This fine motor response pattern was due to differences between parents and ECEs at the 0 – 12 month and 27 – 36 month age cohort (+9% each, respectively). Thus, parents were consistently rating fine motor skills using response choices that reflected developing skills (Haring et al., 1978), as shown by selecting not yet and sometimes more often than ECEs.

Child Age and Length of ECE-Child Relationship as Predictors of Congruence

As described in Chapter 4, using beta regression, child age in months and length of relationship with the ECE were statistically significant predictors of the proportional means. Only length of child-ECE relationship was a statistically significant predictor of the proportional variance. The magnitude of the relationship between the predictors and the dependent variables, however, was small. McFadden's pseudo R^2 (Younger, 1979) provided additional information about the magnitude of the relationship between the predictors (child age and length of ECE-child relationship) and the dependent variable (proportion of overall parent-ECE agreement) and was .09. The finding that the magnitude of the relationship is relatively small may be due, in part, to the study inclusion criteria. ECEs had to know the child for at least 1 month to be enrolled in the present study. Had the sample included professionals who completed an ASQ-3 at one point in time and who did not know the child for at least 1 month, the magnitude of the relationship between this predictor variable and the dependent variables might have differed. None of the studies reviewed in Chapter 2 that focused on congruence investigated predictors of parent-professional congruence.

Limitations of the Present Study

Several limitations of the present study should be considered in relation to reported results and interpretations. First, the results reported represent a single point in time completion of the ASQ-3 by two informants to examine congruence. It is possible that congruence between two informants might change over time. In a study where the ASQ was implemented with 36 at-risk mothers, serial completions of the ASQ were associated with higher correlations between mothers' and professionals' ratings over time (Squires & Bricker, 1991). Thus, it is possible that congruence might have been different had data collection occurred at a different time of year, which would have changed the length of the child-ECE relationship, or if serial administration of the ASQ-3 had occurred throughout the year.

Second, the informants did not have the same training and experiences related to the ASQ-3 before completing the instrument for children in the present study. For the majority of parents (67.5%) in the present study, it was the first time they completed the ASQ-3 for their child, while all of the ECEs in the study had training related to the ASQ-3 and experienced completing the ASQ-3 before completing it for a child in the present study. Congruence between two informants who have extended time with a child might differ when sampling informants who have similar experiences with a multi-dimensional screening instrument.

Third, the informants sampled in the present study are not necessarily representative of the general population of parents and ECEs involved in early education and care programs. The early education and care centers were based at a university, and therefore, were only available to faculty, staff, and students affiliated with the university. All parents in the present study had some college experience, which is a contrast to parents represented in other score congruence studies reviewed in Chapter 2. When specified, approximately 40% or less of the parent sample had at least some college experience (Bricker & Squires, 1989; Bricker, et al., 1988; Burgess, et

al., 1984; Chiu & DiMarco, 2010; Coghlan, et al., 2003; Frankenburg, et al., 1987; Frankenburg, von Doornick, Liddell, & Dick, 1976; Glascoe, 2002; Squires & Bricker, 1991; Squires, et al., 1997). In addition, the majority of ECEs in the present study had a bachelor's degree, primarily in the areas of early childhood/early childhood special education or child development. The sample of ECEs included in this study generally have more training than the general population of ECEs considering many states have a minimum requirement of a high school diploma to work in a child care program (Bureau of Labor Statistics, 2013; Children's Defense Fund, 2005). It is estimated that approximately 20% of teachers in child care centers have a high school diploma or less (Child Care Aware of America, 2012). ECEs in the present study were employed by NAEYC accredited centers, while only 10% of child care centers in the United States are nationally accredited (Child Care Aware of America, 2012). Results from the present study might not generalize to the general population of informants or centers given the characteristics of the centers involved in the present study and the providers and parents affiliated with them.

Fourth, while the total number of children and informant pairs in the present study (n = 126) exceeded the minimum number specified in the power analysis (n = 120), there were a disproportionate numbers of children in the preschool age cohorts relative to the other age cohorts. The power analysis suggested 30 children were needed in the 0 – 12 month age cohort and in the 14 – 24 month age cohort. Although thorough recruiting methods were used, only 17 children in the 0-12 month age cohort and 27 children in the 14-24 month age cohort had complete data to be included in the analyses. As a result, the proportion of questionnaires for each age cohort was inversely disproportionate to the normative sample for the ASQ-3 (Squires & Bricker, 2009) as shown in Table 3-1. Research aims examining congruence by age cohorts or age cohorts and developmental domains should be interpreted with caution and results should be

considered preliminary. Questions concerning variability in congruence by child age cohort or age in months should be further explored in future studies with larger samples of infants (0 -12 months) and toddlers (14-24 months).

Implications of Findings

The present study contributes to the developmental screening score congruence literature by using procedures and analyses that permitted the exploration of similarities and differences in perspectives about children's developmental skills between parents and early childhood educators. Findings from the present study suggest that parents and ECEs generally share similar perspectives about children's development when completing the same multi-dimensional developmental screening instrument in the same way. Although congruence varied to some extent across developmental domains, age cohorts, and by age cohorts and developmental domains, when two response categories rather than three were examined for both items and classification, average percent agreement approximated or exceeded 80%. ICCs indicated moderate relationships between informants' scores for most of the ASQ-3 domains. Taken together, these data suggest that parents and ECEs shared relatively similar perspectives about this sample of young children. When differences in perspectives existed, they often were associated with one informant scoring sometimes and the other informant scoring yes, which suggests that differences are primarily occurring when children have emerging developmental skills, meaning that they are beginning to generalize an acquired skill across contexts and people (Haring et al., 1978). Thus, disagreements between parents and ECEs are not necessarily an indication that one informant is "overestimating" or the other informant is "underestimating," instead, comparing the similarities and differences across informants can provide information as to the child's current phase of learning (i.e., acquisition, fluency, generalization, adaptation;

Haring et al., 1978) or about the importance of sharing and verifying perspectives in relation to contextual expectations and skills observed within and across contexts.

Although there has been a misconception that a goal of research focused on congruence is to provide evidence for parental estimates of child ability being similar to professionals (see reviews by Dinnebeil & Rule, 1994; Snyder et al., 1992), proponents for multi-informant approaches to assessment, including screening, recognize the need to incorporate information from both practitioners and parents. Striving for exact agreement in judgments or ratings would negate the benefits of including multiple perspectives in assessment (Suen et al., 1995). Parents and professionals bring different sets of observations and expertise to the assessment process and each perspective is essential to making informed decisions as a team. The interest in investigating congruence was to understand the relationship between the perspectives of two informants who spend extended time with the child across different contexts. Understanding the similarities and differences in the perspectives of parents and ECEs might help inform strategies that facilitate the formation of a multiply linked mesosystem in which there is a compatibility of expectations and activities for children (Bronfenbrenner, 1979). Continuity across settings may contribute to the development of mutual goals, including mutual goals related to development, learning, and behavior, thus increasing the frequency of opportunities for a child to experience developmentally enhancing learning opportunities across different contexts and people.

Given the typically brief interactions between parents and practitioners in early care and education settings, it often is difficult for parents and ECEs to regularly exchange information about children's development as well as their abilities, needs, and preferences (Bailey & Wolery, 1992). One possibility for enhancing information exchange in early care and education settings is to take advantage of existing or anticipated practices and incorporate strategies that actively

involve families in sharing perspectives about their child's development, learning, and behavior. Developmental screening is increasingly becoming a common practice in early education and care programs (National Resource Council, 2008). The inclusion of parents in this process fulfills recommendations of conducting multi-informant assessment (DEC, 2007; NAEYC 2003, 2012) while also presenting a potentially effective and efficient strategy for establishing shared understandings of children's development, learning, and behavior to enhance children's experiences provided across multiple settings (Bronfenbrenner, 1974, 1979). A review of the historical background of developmental screening showed that although parent involvement has been long regarded as a recommended practice in screening, it has not been consistently implemented (AOA, 1978; Paget & Nagle, 1976). Difficulties with incorporating parent input have been partially due to vague policies and guidelines for practitioners as well as, until recently, the limited availability of instruments, such as the ASQ-3, that meet standards for educational and psychological measurement yet permit judgment based scoring by informants rather than direct elicitation of children's skills (Glascoe et al., 1990; Katoff & Reuter, 1980; Macy, 2012).

In the past decade, prominent organizations in the medical, psychological, and educational fields have unanimously supported the gathering of multiple sources of evidence to inform assessment, specifically, the inclusion of parent input in the decision-making process (DEC, 2007; Glascoe, 1991; NASP, 2009; Neisworth & Bagnato, 2005). Using parent-completed developmental screening instruments could serve dual benefits of improving the ecological validity of screening results and enhancing information exchange between parents and ECEs (Bagnato 2007).

The results of the present study can also inform developmental screening practices in early education and care centers, especially given recent federal-level recognition of the need for universal developmental screening of young children (AAP, 2001; Council on Children with Disabilities et al., 2006; NASP, 2009; Obama for America, 2012). To highlight the potential utility of sharing perspectives across contexts using developmental screening instruments, a case study of the parents and ECEs of "Child A" will be reviewed.

At the time of data collection, Child A was 34 months of age and had known the two ECEs in her classroom for 10 months. Child A was a Caucasian female with a significant visual impairment. Overall percent agreement between this child's mother and one of the ECEs in her classroom was 83% for item agreement and 80% for classification agreement. These percentages are generally considered to be high congruence between parents and ECEs. Examining item-level response patterns, however, provides a more nuanced view of perspectives about Child A's development from each informant's perspective. Across the 30 items, Child A's mother only endorsed one item, indicating that Child A could turn pages in a book. Child A's ECE also endorsed this item, in addition to five other items which her mother had indicated she was not yet demonstrating. The ECE indicated that Child A was able to push a toy on wheels (e.g., wagon) and could sometimes point to seven body parts when asked, demonstrate receptive understanding of "up" and "down," string small beads, and use a spoon for feeding. Differences between not yet and sometimes may indicate that Child A is beginning to develop these skills and her mother may not provide opportunities or activities in which to observe these developing skills or may need some support in recognizing these discrete skills or how to teach these skills to her young child with a visual impairment. Although her mother and ECE have similar perspectives about her overall development (i.e., agreed that she was not yet exhibiting skills reflected on 24 of 30

ASQ-3 items), regular use of developmental screening instruments by both informants continually monitors Child A's developmental skill progression and can provide structure or guidance for shared conversations.

Systematic developmental screening programs afford early education and care centers the opportunity to use program data as a needs assessment for professional development of staff on the observation, documentation, or support of developmental skills. This need became evident when comparing the ASQ-3 results from one infant classroom to the other infant classrooms involved in the present study. Of the four infant classrooms in the present study, one of the infant classrooms had 6 of 7 children identified by ECEs as being below the cutoff on the ASQ-3 personal-social domain and therefore eligible for further evaluation. Of these children, the parents of 2 children also identified their child as eligible for further evaluation, while the remaining 4 children were only identified by ECEs. The frequency of below the cutoff classifications was particularly high in this one classroom compared to the other infant classrooms, signifying a possible need for professional development on the observation, documentation, or support of personal-social skills for the ECEs assigned to that classroom.

Future research might replicate the present study in different early childhood settings or with other populations to explore generalizability of results, particularly, with a larger sample of infant-toddler parent-ECE pairs. A larger sample of parent-ECE pairs might permit the application of other exploratory analyses that would provide additional or different information about predictors of parent-ECE congruence. Additional studies might be conducted in which multiple and periodic ASQ-3 forms are collected from parent-ECE pairs to explore whether there are changes in congruence over time, considering the promising results of Squires and Bricker

(1991) in which serial completions of the ASQ-3 was associated with higher correlations between mothers' and professionals' scores.

Based on findings of the present study, researchers might want to investigate further the ecological validity of several ASQ-3 (Squires & Bricker, 2009) items. According to the *ASQ-3 User's Guide*, items were developed for the ASQ after a review of standardized developmental tests available at the time the first version of the instrument (IMQ; Bricker & Squires, 1989a) was published, nonstandardized tests focused on early development, textbooks, and literature about early developmental milestones (Squires et al., 2009). ASQ items were developed based on the following criteria: (a) skills could be easily observed or elicited by parents, (b) skills were highly likely to occur within homes and child care settings, and (c) skills reflected important developmental milestones. In the present study, half of the omitted items occurred in the fine motor domain, which raises questions about the ecological validity of some fine motor items The item omitted most often by parents related to a child's ability to string beads or macaroni noodles. With respect to another ASQ-3 fine motor item, a parent noted she did not rate an item involving her child's use of scissors because child-safe scissors were not available in the home. These examples raise questions about the extent to which some materials associated with fine motor items on the ASQ-3 are present in young children's home environments. Present study findings suggest additional studies about the ecological validity of some ASQ-3 items might be warranted, particularly given the consequential validity of parents either omitting or scoring these items incorrectly (e.g., false positive or false negative screening decisions). Rewriting several fine motor items to reflect skills more likely to be observed in the home setting that sample similar developmental skills and conducting studies to validate these items might also be warranted.

Results of the present study provides information that may be beneficial for enacting strategies related to supporting the formation of multiply linked mesosystems (Bronfenbrenner, 1979) by showing the potential of adjusting existing program practices to serve multiple goals of early identification, parent involvement in the assessment process, and most importantly, shared discussion about children's developmental skill progression across the multiple contexts in which children learn and grow. While parents and ECEs typically had high congruence in rating discrete developmental skills and recommending whether to refer children for further evaluation, examination of the variations in congruence by developmental domain and age cohort show that the nuances of these perspectives can be complex. Rather than viewing differences in perspectives as a concern or an indication of inaccuracies in judgment or scoring, findings of the present study suggest explanations for and discussions about disagreements between informants should be explored. In the present study, disagreements between parents and ECEs occurred primarily when rating emerging skills, and thus, variations in congruence might largely be a reflection of the child's current phase of learning. The reported variations in congruence by age cohort and developmental domain suggest that differences might be due to complexities in the sequence of development, with the timing of important developmental milestones heightening the awareness of informants about particular developmental skills. When exploring explanations for congruence, it is also important to consider the differing expectations or opportunities provided to young children across contexts. Most notably, results from the present study show that instead of striving for exact agreement between parents and professionals about children's developmental skills, research and practice should continue to preserve the richness of incorporating multiple perspectives and uphold parents as essential informants and team members.

APPENDIX: INSTRUCTIONS FOR PARTICIPANTS

Instructions for Early Childhood Educators

Thank you for taking the time to complete this form. Please complete all of the items on this form as best as you can. When you complete this form, please do not discuss your ratings with **the child's parent or other caregiver**.

You will be asked about the child's skills during different activities. For each item, you will indicate how often the child performs the skill by marking one of three items:

Select **Yes** if the child uses the skill consistently at different times and across different people and settings.

Select **Sometimes** if you have occasionally observed the child use this skill, the child uses this skill with some settings and some people, or the child uses this skill with some help from an adult.

Select **Not Yet** if you have not observed the child use this skill or if the child can only do this skill with complete adult assistance.

Let's look at an example!

Item: *Does your child eat with a fork?*

When the child is eating food that requires the use of a fork, you have observed the child use a fork all or most of the time, regardless of who he/she is eating with, then you would mark "yes."

When the child is eating food that requires the use of a fork, he/she switches between using a fork or hands to eat, needs a reminder to use a fork, or only uses a fork at school, then you would mark "sometimes."

When the child is eating food that requires the use of a fork, he/she only uses hands to eat, or can only use a fork if you place your hand over his/her hand to guide them, then you would mark "not yet."

Item: *Does your baby pick up a small toy with only one hand?*

When the child is playing with toys slightly larger than the palm of the hand, he/she can hold it in one hand without dropping it, regardless of whether he/she is sitting, standing, and lying down, then you would mark “yes.”

When the child is playing with toys slightly larger than the palm of the hand, he/she can only hold it for a moment before dropping it, can only do it when in certain positions, or can only do it with some small toys, then you would mark “sometimes.”

When the child is playing with toys slightly larger than the palm of the hand, he/she can only hold onto it if supported by both hands, then you would mark “not yet.”

If you need any clarification on how to respond to a particular item, please feel free to contact:

Cathleen Pasia
Cell: (XXX) XXX – XXXX
Email: c.pasia@ufl.edu

Instructions for Parents

Thank you for taking the time to complete this form. Please complete all of the items on this form as best as you can. When you complete this form, please do not discuss your ratings with **any staff from [CENTER NAME]**.

You will be asked about your child's skills during different activities. For each item, you will indicate how often your child performs the skill by marking one of three items:

Select **Yes** if your child uses the skill consistently at different times and across different people and settings.

Select **Sometimes** if you have occasionally observed your child use this skill, your child uses this skill with some settings and some people, or your child uses this skill with some help from an adult.

Select **Not Yet** if you have not observed your child use this skill or if your child can only do this skill with complete adult assistance.

Let's look at some examples!

Item: *Does your child eat with a fork?*

When your child is eating food that requires the use of a fork, you have observed your child use a fork all or most of the time, regardless of who he/she is eating with, then you would mark "yes."

When your child is eating food that requires the use of a fork, he/she switches between using a fork or hands to eat, needs a reminder to use a fork, or only uses a fork at school, then you would mark "sometimes."

When your child is eating food that requires the use of a fork, he/she only uses hands to eat, or can only use a fork if you place your hand over his/her hand to guide them, then you would mark "not yet."

Item: *Does your baby pick up a small toy with only one hand?*

When your child is playing with toys slightly larger than the palm of the hand, he/she can hold it in one hand without dropping it, regardless of whether he/she is sitting, standing, and lying down, then you would mark “yes.”

When your child is playing with toys slightly larger than the palm of the hand, he/she can only hold it for a moment before dropping it, can only do it when in certain positions, or can only do it with some small toys, then you would mark “sometimes.”

When your child is playing with toys slightly larger than the palm of the hand, he/she can only hold onto it if supported by both hands, then you would mark “not yet.”

If you need any clarification on how to respond to a particular item, please feel free to contact:

Cathleen Pasia
Cell: (XXX) XXX – XXXX
Email: c.pasia@ufl.edu

LIST OF REFERENCES

Alpern, G. D., Boll, T. J., & Shearer, M. S. (1980). *Alpern-Boll Developmental Profile II Manual.* Aspen, CO: Psychological Development Publications.

American Academy of Pediatrics, Committee on Children with Disabilities. (2001). Developmental surveillance and screening of infants and young children. *Pediatrics, 108,* 192-196.

American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington DC: American Educational Research Association.

American Orthopsychiatric Association. (1978). Developmental assessment in EPSDT. *American Journal of Orthopsychiatry, 48*(1), 7-21.

Bailey, D. B., & Wolery, M. (1992). *Teaching infants and preschoolers with disabilities* (2nd ed). New York, NY: Macmillan Publishing Company.

Bagnato, S. J. (2007). *Authentic assessment for early childhood intervention: Best practices.* New York, NY: Guilford Press.

Barnett, D. W. (1984). An organizational approach to preschool services: Psychological screening, assessment, and intervention. In C. Maher, R. Ilbac, &J. Zins (Eds.), *Organizational psychology in schools: A handbook for practitioners* (pp. 53-58) Springfield, IL: Charles C Thomas.

Bayley, N. (1969). *Bayley Scales of Infant Development.* San Antonio, TX; Psychological Corporation.

Bayley, N. (1993). *Bayley II Scales of Infant Development*. San Antonio, TX: Psychological Corporation.

Boan, C. H., Aydlett, L., & Multunas, N. (2007). Early childhood screening and readiness assessment. In R. J. Nagle (Ed.), *Psychoeducational assessment of preschool children* (4th ed., pp. 49-67). Mahwah, NJ: Lawrence Erlbaum.

Bradley-Johnson, S., & Johnson, C. M. (2007). Infant and toddler cognitive assessment. In B. A. Bracken (Ed.), *The psychoeducational assessment of preschool children* (4th ed., pp. 325-357). Needham Heights, MA: Allyn & Bacon.

Brassard, M. R., & Boehm, A. E. (2007). *Preschool assessment: Principles and practices.* New York, NY: Guilford Press.

Bricker, D., & Squires, J. (1989a). *Infant Monitoring System.* Eugene, OR: Center for Human Development, College of Oregon

Bricker, D., & Squires, J. (1989b). The effectiveness of parental screening of at-risk infants: The Infant Monitoring Questionnaires. *Topics in Early Childhood Special Education, 9*(3), 67-85. doi:10.1177/027112148900900306

Bricker, D., & Squires, J. (1999). *Ages & Stages Questionnaires (ASQ): A parent-completed, child-monitoring system (*2nd ed.*).* Baltimore, MD: Paul H. Brooks

Bricker, D., Squires, J., Kaminski, R., & Mounts, L. (1988). The validity, reliability, and cost of a parent-completed questionnaire system to evaluate at-risk infants. *Journal of Pediatric Psychology, 13*(1), 55-68.

Brigance, A. H., & Glascoe, F. P. (2002). *Brigance Infant & Toddler Screen*. North Billerica, MA: Curriculum Associates.

Bronfenbrenner, U. (1974). Developmental research, public policy, and the ecology of childhood. *Child Development, 45,* 1-5.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design.* Cambridge, MA: Harvard University Press.

Brooks-Gunn, J., & Lewis, M. (1983). Screening and diagnosing handicapped infants. *Topics in Early Childhood Special Education, 3*(1), 14-28. doi:10.1177/027112148300300107

Bureau of Labor Statistics, U.S. Department of Labor. (2013). *Occupational outlook handout, 2012-2013 edition, childcare workers.* Retrieved from http://www.bls.gov/ooh/personal-care-and-service/childcare-workers.htm

Burgess, D. B., Asher, K. N., Doucet, H. J., Reardon, K., & Daste, M. R. (1984). Parent report as a means of administering the prescreening developmental questionnaire: An evaluation study. *Journal of Developmental and Behavioral Pediatrics, 5,* 195-200.

Campos, J. A., Squires, J., & Ponte, J. (2011). Universal developmental screening: Preliminary studies in Galicia, Spain. *Early Child Development and Care, 181,* 475-485.

Child Care Aware of America. (2012). *Child care in America: 2012 state fact sheets.* Retrieved from www.naccrra.org

Children's Defense Fund. (1978). EPSDT in practice: What's happening in the field? *American Journal of Orthopsychiatry, 48*(1), 77-95.

Children’s Defense Fund. (2005). *Child care basics.* Retrieved from http://www. childrensdefense.org/child-research-data-publications/data/child-care-basics.pdf

Chiu, S., & DiMarco, M. A. (2010). A pilot study comparing two developmental screening tools for use with homeless children. *Journal of Pediatric Health Care, 24*(2), 73-80. doi:10.1016/j.pedhc.2009.01.003

Coghlan, D., Kiing, J., & Wake, M. (2003). Parents' evaluation of developmental status in the Australian day-care setting: Developmental concerns of parents and carers. *Journal of Paediatric Child Health, 39,* 49-54.

Colligan, R. C. (1981). Prediction of reading difficulty from parental preschool report: A 3-year follow-up. *Learning Disability Quarterly, 4*(1), 31-37. doi:10.2307/1510710

Council on Children with Disabilities, Section on Developmental Behavioral Pediatrics, Bright Futures Steering Committee, & Medical Home initiatives for Children with Special Needs Project Advisory Committee. (2006). Identifying infants and young children with developmental disorders in the medical home: An algorithm for developmental surveillance and screening. *Pediatrics, 118,* 405-420. doi:10.1542/peds.2006-1231

Diamond, K. E. (1993). The role of parents' observations and concerns in screening for developmental delays in young children. *Topics in Early Childhood Special Education, 13*(1), 68-81. doi:10.1177/027112149301300108

Diamond, K. E., & Le Furgy, W. G. (1988). Screening for developmental handicaps: Outcomes from an early childhood screening program. *Physical & Occupational Therapy in Pediatrics, 8*(1), 43-59. doi:10.1300/J006v08n01_04

Diamond, K. E., & Squires, J. (1993). The role of parental report in the screening and assessment of young children. *Journal of Early Intervention, 17*(2), 107-115. doi:10.1177/105381519301700203

Dinnebeil, L. A., & Rule, S. (1994). Congruence between parents' and professionals' judgments about the development of young children with disabilities: A review of the literature. *Topics in Early Childhood Special Education, 14*(1), 1-25. doi: 10.1177/027112149401400105

Dionne, C., Squires, J., Leclerc, D., Peloquin, J., & McKinnon, S. (2006). Cross-cultural comparison of a French Canadian and U.S. developmental screening test. *Developmental Disabilities Bulletin, 34*(1-2), 43-56.

Division for Early Childhood. (1993). *DEC recommended practices: Indicators of quality in programs for infants and young children with special needs and their families.* Pittsburgh, PA: Author.

Division for Early Childhood. (2007). *Promoting positive outcomes for children with disabilities: Recommendations for curriculum, assessment and program evaluation.* Retrieved from *www.dec-sped.org/.../Prmtg_Pos_Outcomes_Companion_Paper.pdf*

Dunbar, S. B., & Reed, C. N. (1999). A developmental screening program in primary health care: Meeting the challenges of children in low-income families. *Infant-Toddler Intervention, 9*(2), 195-202.

Dworkin, P. H. (2006). Promoting development through child health services introduction to the help me grow roundtable. *Journal of Developmental and Behavioral Pediatrics, 27,* S2-S4. doi:10.1097/00004703-200602001-00002

Earls, M. F., Andrews, J. E., & Hay, S. S. (2009). A longitudinal study of developmental and behavioral screening and referral in North Carolina's assuring better child health and development participating practices. *Clinical Pediatrics, 48*(8), 824-833.

Eddey, G. E., Robey, K. L., Zumoff, P., & Malik, Z. Q. (1995). Developmental delay in children from birth to six years of age. *Infant-Toddler Intervention: The Transdiscplinary Journal, 5*(3), 233-241.

Elbers, J., Macnab, A., McLeod, E., & Gagnon, F. (2008). The Ages and Stages Questionnaires: Feasibility of use as a screening tool for children in Canada. *Canadian Journal of Rural Medicine, 13*(1), 9-12.

First, L. R., & Palfrey, J. S. (1994). The infant or young child with developmental delay. *The New England Journal of Medicine, 330*(7), 478-483. doi:10.1056/NEJM199402173300708

Flamant, C., Branger, B., Tich, S. N., Rochebrochard, E., Savagner, C., Berlie, I., & Roze, J. (2011). *Public Library of Science, 6*(5), 1 – 6.

Frankenburg, W. K. (1987). *Revised Denver Prescreening Developmental Questionnaire (PDQ II).* Denver, CO: DDM, Inc.

Frankenburg, W. K., & Dodd, J. (1975). *Denver Developmental Screening Test-Revised (DDST).* Denver, CO: DDM, Inc.

Frankenburg, W. K., Dodd, J., Archer, P., Bresnick, B., Maschka, P., Edelman, N., & Shapiro, H. (1992). *The Denver-II: Training manual.* Denver, CO: Denver Developmental Materials.

Frankenburg, W. K., Fandal, A. W., & Thornton, S. M. (1987). Revision of the Denver prescreening developmental questionnaire. *The Journal of Pediatrics, 110*, 653-657.

Frankenburg, W. K., & North, A. F. (1974). *A guide to screening for the EPSDT program under Medicaid.* Washington, DC: Social and Rehabilitation Service, HEW.

Frankenburg, W. K., van Doorninck, W. I., & Liddell, T. (1976). *Denver Prescreening Developmental Questionnaire (PDQ).* Denver, CO: DDM, Inc.

Frankenburg, W. K., van Doorninck, W. J., Liddell, T. N., & Dick, N. P. (1976). The Denver Prescreening Developmental Questionnaire (PDQ). *Pediatrics, 57*(5), 744.

Frey, A., & Walker, H. (2006). Education policy for children, youth, and families. In J. M. Jenson & M. W. Fraser (Eds.), *Social policy for children and families: A risk and resilience perspective* (pp. 67-92). New York, NY: Chatham House/ Seven Bridges Press.

Frisk, V., Montgomery, L., Boychyn, E., Young, R., vanRyn, E., McLachlan, D., & Neufeld, J. (2009). Why screening Canadian preschoolers for language delays is more difficult than it should be. *Infants and Young Children, 22*(4), 290-308.

Garbarino, J. (1990). The human ecology of early risk. In S. J. Meisels, & J. P. Shonkoff (Eds.), *Handbook of early intervention* (pp. 78-96). New York, NY: Cambridge University Press.

Gasson, N., & Piek, J. P. (2003). A longitudinal investigation of the development of fullterm, preterm and 'at risk' infants from birth to four years using a parent-report measure. *Australian Journal of Psychology,* 78-82.

Glascoe, F. P. (1991). Developmental screening: Rationale, methods, and application. *Infants & Young Children, 4*(1), 1-10.

Glascoe, F. P. (2001). Are overreferrals on developmental screening tests really a problem? *Archives on Pediatric and Adolescent Medicine, 155,* 54-59.

Glascoe, F. P. (2002). The Brigance Infant and Toddler Screen: Standardization and validation. *Journal of Developmental and Behavioral Pediatrics, 23,* 145-150.

Glascoe, F. P. (2005). Screening for developmental and behavioral problems. *Mental Retardation and Developmental Disabilities Research Reviews, 11*(3), 173-179. doi:10.1002/mrdd.20068

Glascoe, F. P. (2007a). Developmental and behavioral screening. In J. Rojahn (Ed.), *Handbook of intellectual and developmental disabilities.* (pp. 353-371). New York, NY: Springer Publishing. doi:10.1007/0-387-32931-5_19

Glascoe, F. P. (2007b). *Parents' Evaluation of Developmental Status*. Nolensville, TN: Ellsworth & Vandermeer Press.

Glascoe, F. P. (2010). Developmental and behavioral considerations. In S. Tanski, L. Garfunkel, P. Duncan, & M. Weitzman (Eds.), *Performing preventative services: A Bright Futures handbook.* Retrieved from http://brightfutures.aap.org/continuing_education.html

Glascoe, F. P., Byrne, K. E., Chang, B., Stickland, B., Ashford, L. G., & Johnson, K. L. (1992). Accuracy of the Denver-II in developmental screening. *Pediatrics, 89,* 1221 – 1224.

Glascoe, F. P., & Dworkin, P. H. (1993). Obstacles to effective developmental surveillance: Errors in clinical reasoning. *Journal of Developmental and Behavioral Pediatrics, 14,* 344-349.

Glascoe, F. P., Martin, E. D., & Humphrey, S. (1990). A comparative review of developmental screening tests. *Pediatrics, 86*(4), 547-554.

Gollenberg, A. L., Lynch, C. D., Jackson, L. W., McGuinness, B. M., & Msall, M. E. (2010). Concurrent validity of the parent-completed Ages and Stages Questionnaires, 2nd ed. with the Bayley Scales of Infant Development II in a low-risk sample. *Child Care, Health & Development, 36*(4), 485-490. doi:10.1111/j.1365-2214.2009.01041.x

Gracey, C. A., Azzara, C. V., & Reinherz, H. (1984). Screening revisited: A survey of U.S. requirements. *The Journal of Special Education, 18*(2), 101-107. doi:10.1177/002246698401800203

Green, H., & Scriven, G. (1980). Current practices in the identification of children with handicaps. *Early Child Development and Care, 6*(3-4), 125-134. doi:10.1080/0300443800060304

Guiberson, M., & Rodríguez, B. L. (2010). Measurement properties and classification accuracy of two Spanish parent surveys of language development for preschool-age children. *American Journal of Speech-Language Pathology, 19*(3), 225-237.

Hall, J. D., Bramlett, R. K., Barnett, D. W., & Cox, F. W. (1994). Classification of risk status in kindergarten screening: A comparison of alternative measures. *Journal of Psychoeducational Assessment, 12*(2), 154-164. doi:10.1177/073428299401200205

Handal, A. J., Lozoff, B., Breilh, J., & Harlow, S. D. (2007a). Effect of community of residence on neurobehavioral development in infants and young children in a flower-growing region of Ecuador. *Environmental Health Perspectives, 115*(1), 128-133.

Handal, A. J., Lozoff, B., Breilh, J., & Harlow, S. D. (2007b). Sociodemographic and nutritional correlates of neurobehavioral development: A study of young children in a rural region of Ecuador. *Revista Panamericana De Salud Pública, 21*(5), 292-300.

Haring, N.G., Lovitt, T.C., Eaton, M.D., & Hansen, C.L. (1978). *The fourth R: Research in the classroom.* Columbus, OH: Charles E. Merrill Publishing Co.

Harrington, R. G. (1984). Preschool screening: The school psychologist's perspective. *School Psychology Review, 13*, 363-374.

Harris, P. A., Thielke, R., Taylor, R., Payne, J. Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap): A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics, 42*, 377 – 381. doi:10.1016/j.jbi.2008.08.010.

Heo, K. H., Squires, J., & Yovanoff, P. (2008). Cross-cultural adaptation of a pre-school screening instrument: Comparison of Korean and US populations. *Journal of Intellectual Disability Research, 52*(3), 195-206.

Hix-Small, H., Marks, K., Squires, J., & Nickel, R. (2007). Impact of implementing developmental screening at 12 and 24 months in a pediatric practice. *Pediatrics, 120*(2), 381-389. doi:10.1542/peds.2006-3583

Ikeda, M. J., Neessen, E., & Witt, J. (2008). Best practices in universal screening. In A. Thomas, and J. Grimes (Eds.), *Best practices in school psychology V.* (pp. 103 – 114). Bethesda, MD: National Association of School Psychologists.

Ireton, H. (1990). Developmental screening measures. In J. Goldman (Ed.), *Developmental assessment in clinical child psychology: A handbook.* (pp. 78-99). Elmsford, NY: Pergamon Press.

Improving Head Start for School Readiness Act, 42 U.S.C. § 1304.20, 2007

Individuals with Disabilities Education Act, 20 U.S.C. 1400 § 602(23), 2004

Individuals with Disabilities Education Act, 20 U.S.C. 1419 § 300, 1986

Jackson, B. J., & Needelman, H. (2007). Building a system of child find through a 3-tiered model of follow-up. *Infants & Young Children, 20*(3), 255-265. doi:10.1097/01.IYC.0000277756.00938.e2

Janson, H., & Squires, J. (2004). Parent-completed developmental screening in a Norwegian population sample: A comparison with US normative data. *Acta Paediatrica, 93,* 1525-1529.

Jee, S. H., Szilagyi, M., Ovenshire, C., Norton, A., Conn, A., Blumkin, A., & Szilagyi, P. G. (2010). Improved detection of developmental delays among young children in foster care. *Pediatrics, 125*(2), 282-289. doi:10.1542/peds.2009-0229

Johnson, D., Poteat, G. M., & Kushnick, T. (1986). Comparison of mental age estimates made by pediatricians and mothers of preschool children. *Journal of Pediatric Psychology, 11,* 385-392.

Kapci, E. G., Kucuker, S., & Uslu, R. I. (2010). How applicable are Ages and Stages Questionnaires for use with Turkish children? *Topics in Early Childhood Special Education, 30*(3), 176-188. doi:10.1177/0271121410373149

Katoff, L., & Reuter, J. (1980). Review of developmental screening tests for infants. *Journal of Clinical Child Psychology, 9*(1), 30-34. doi:10.1080/15374418009532939

Katz, K. S. (1989). Strategies for infant assessment: Implications of P.L. 99-457. *Topics in Early Childhood Special Education, 9*(3), 99 – 109.

Kenny, T. J., & Culbertson, J. L. (1993). Developmental screening for preschoolers. In D. J. Willis (Ed.), *Testing young children: A reference guide for developmental, psychoeducational, and psychosocial assessments.* (pp. 73-100). Austin, TX: PRO-ED.

Kerstjens, J. M., Bos, A. F., ten Vergert, E. M. J., de Meer, G., Butcher, P. R., & Reijneveld, S. A. (2009). Support for the global feasibility of the Ages and Stages Questionnaire as developmental screener. *Early Human Development, 85*(7), 443-447. doi:10.1016/j.earlhumdev.2009.03.001

Kim, M. M., & O'Connor, K. S. (1996). Do parents and professionals agree on the developmental status of high-risk infants? *Pediatrics, 97*(5), 676-681.

Klamer, A., Lando, A., Pinborg, A., & Greisen, G. (2005). Ages and Stages Questionnaire used to measure cognitive deficit in children born extremely preterm. *Acta Paediatrica, 94*(9), 1327-1329. doi:10.1080/08035250510042825

Knobloch, H., Stevens, F., & Malone, A. (1980). *Manual of development diagnosis: The administration and interpretation of the revised Gesell and Armatruda developmental and neurologic examination.* Hagerstown, MD: Harper & Row.

Knobloch, H., Stevens, F, Malone, A., Ellison, P., & Risemberg, H. (1979). The validity of parental reporting of infant development. *Pediatrics, 63,* 872-878.

Kochanek, T. T. (1993). Enhancing screening procedures for infants and toddlers: The application of knowledge to public policy and program initiatives. In M. A. Graham (Ed.), *Implementing early intervention: From research to effective practice* (pp. 46-66). New York, NY: Guilford Press.

Lichtenstein, R., & Ireton, H. (1984). *Preschool screening: Identifying young children with developmental and educational problems.* Orlando, FL: Grune & Stratton.

Lindsay, N. M., Healy, G. N., Colditz, P. B., & Lingwood, B. E. (2008). Use of the Ages and Stages Questionnaire to predict outcome after hypoxic-ischaemic encephalopathy in the neonate. *Journal of Paediatrics & Child Health, 44*(10), 590-595. doi:10.1111/j.1440-1754.2008.01388.x

Macy, M. (2012). The evidence behind developmental screening instruments. *Infants & Young Children, 25,* 19-61. doi: 10.1097/IYC.0b013e31823d37dd

Marks, K., Hix-Small, H., Clark, K., & Newman, J. (2009). Lowering developmental screening thresholds and raising quality improvement for preterm children. *Pediatrics, 123,* 1516-1523.

McCarthy, D. (1972). *McCarthy scales of children's abilities.* New York, NY: Psychological Corp.

McCrae, J. S., Cahalane, H., & Fusco, R. A. (2011). Directions for developmental screening in child welfare based on the Ages and Stages Questionnaires. *Child & Youth Services Review, 33*(8), 1412-1418.

McLean, M., & McCormick, K. (1993). Assessment and evaluation in early intervention. In L. F. Pearl (Ed.), *Family-centered early intervention with infants and toddlers: Innovative cross-disciplinary approaches* (pp. 43-79). Baltimore, MD: Paul H Brookes Publishing.

Meisels, S. J., & Wasik, B. A. (1990). Who should be served? Identifying children in need of early intervention. In J. P. Shonkoff (Ed.), *Handbook of early childhood intervention.* (pp. 605-632). New York, NY: Cambridge University Press.

Miller, S. A., Manhal, M., & Mee, L. L. (1991). Parental beliefs, parental accuracy, and children's cognitive performance: A search for causal relations. *Developmental Psychology*, *27*, 267 - 276.

Moore, B. D. (1978). Implementing the developmental assessment component of the EPSDT program. *American Journal of Orthopsychiatry*, *48*(1), 22-32.

Nagle, R. J. (2007). Issues in preschool assessment. In B. A. Bracken (Ed.), *The psychoeducational assessment of preschool children* (4th ed., pp. 29-48). Needham Heights, MA: Allyn & Bacon.

National Association of School Psychologists. (2009). *Early childhood assessment* (Position statement). Bethesda, MD: Author.

National Association for the Education of Young Children. (2003, November). *Early childhood curriculum, assessment, and program evaluation: Building an effective, accountable system in programs for children birth through age 8.* Retrieved from www.naeyc.org/positionstatements/cape

National Association for the Education of Young Children. (2008). *Overview of NAEYC early childhood program standards.* Retrieved from http://www.naeyc.org/files/academy/file/OverviewStandards.pdf

National Association for the Education of Young Children. (2012, October). *All criteria document.* Retrieved from http://www.naeyc.org/academy/primary/viewstandards

National Research Council. (2008). *Early childhood assessment: Why, what, and how.* Committee on Developmental Outcomes and Assessments for Young Children, C. E. Snow, & S. B. Van Hemel (Eds.). Washington, DC: The National Academies Press.

Neisworth, J. T., & Bagnato, S. J. (2005). DEC recommended practices: Assessment. In S. Sandall, M. L. Hemmeter, B. J. Smith, & M. E. McLean (Eds.), *DEC recommended practices: A comprehensive guide for practical application in early intervention/early childhood special education* (pp. 45-61). Longmont, CO: Sopris West.

Neisworth, J. T., & Fewell, R. R. (1990). Forward. *Topics in Early Childhood Special Education, 10*(3), ix-x. doi:10.1177/027112149001000301

Newborg, J., Stock, J. R., Wnek, L., Guidubaldi, J., & Svinicki, J. (1984). *Battelle Developmental Inventory.* Itasca, IL: Riverside Publishing.

Newborg, J., Stock, J. R., Wnek, L., Guidubaldi, J., & Svinicki, J. (2004). *Battelle Developmental Inventory* (2nd ed.). Chicago, IL: Riverside Publishing.

Obama for America (2012). *Barack Obama and Joe Biden's plan to empower Americans with disabilities.* Retrieved from www.barackobama.com

Owen, M. T., Ware, A. M., & Barfoot, B. (2000). Caregiver-mother partnership behavior and the quality of caregiver-child and mother-child interactions. *Early Childhood Research Quarterly, 15,* 413-428.

Paget, K. D., & Nagle, R. J. (1986). A conceptual model of preschool assessment. *School Psychology Review, 15*(2), 154-165.

Plomgaard, A. M., Hansen, B. M., & Greisen, G. (2006). Measuring developmental deficit in children born at gestational age less than 26 weeks using a parent-completed developmental questionnaire. *Acta Paediatrica, 95,* 1488-1494. doi:10.1080/08035250600684438

Portney, L. G., & Watkins, M. P. (2000). *Foundations of clinical research: Applications to practice* (2nd ed.). Upper Saddle River, NJ: Prentice Hall Health.

Richter, J., & Janson, H. (2007). A validation study of the Norwegian version of Ages and Stages Questionnaires. *Acta Paediatrica, 96,* 748-752. doi:10.1111/j.1651.2227.2007.00246.x

Ryan-Krause, P., Meadows-Oliver, M., Sadler, L., & Swartz, M. K. (2009). Developmental status of children of teen mothers: Contrasting objective assessments with maternal reports. *Journal of Pediatric Health Care, 23*(5), 303-309. doi:10.1016/j.pedhc.2008.09.001

Rydz, D., Srour, M., Oskoui, M., Marget, N., Shiller, M., Birnbaum, R., Majenmer, A., & Shevell, M. I. (2006). Screening for developmental delay in the setting of a community pediatric clinic: A prospective assessment of parent-report questionnaires. *Pediatrics, 118,* e1178-e1186. doi:10.1542/peds.2006-0466

Salvia, J., & Ysseldyke, J. E. (1985). *Assessment in special and remedial education* (3rd ed.). Boston, MA: Houghton Mifflin Company.

Sand, N., Silverstein, M., Glascoe, F. P., Gupta, V. B., Tonniges, T. P., & O'Connor, K. G. (2005). Pediatricians' reported practices regarding developmental screening: Do guidelines work? Do they help? *Pediatrics, 116,* 174-179.

Scott, K. G., & Carran, D. T. (1989). Identification and referral of handicapped infants. In H. J. Walberg (Ed.), *Handbook of special education: Research and practice: Vol. 3. Low incidence conditions.* (pp. 227-241). Elmsford, NY: Pergamon Press.

Scott, F. G., Lingaraju, S., Kilgo, J. L., Kregel, J., & Lazzari, A. (1993). A survey of pediatricians on early identification and early intervention services. *Journal of Early Intervention, 17,* 129-138. doi: 10.1177/105381519301700205

Sheehan, R. (1988). Involvement of parents in early childhood assessment. In T. D. Wachs & R. Sheehan (Eds.), *Assessment of young developmentally disabled children.* (pp. 75 – 90). New York, NY: Plenum Press.

Sheldrick, R. C., & Perrin, E. C. (2009). Surveillance of children's behavior and development: Practical solutions for primary care. *Journal of Developmental and Behavioral Pediatrics, 30*(2), 151-153. doi:10.1097/DBP.0b013e31819f1bfb

Shelton, T. L. (1989). The assessment of cognition/intelligence in infancy. *Infants and Young Children, 1,* 10-25.

Shonkoff, J. P., & Phillips, D. A. (Eds.). (2000). *From neurons to neighborhoods: The science of early childhood development.* Washington DC: National Academy Press.

Simas, A. B., Barreto-Souza, W., & Rocha, A. V. (2010). Improved estimators for a general class of beta regression models. *Computational Statistics and Data Analysis, 54,* 348 – 366.

Skellern, C. Y., Rogers, Y., & O'Callaghan, M. J. (2001). A parent-completed developmental questionnaire: Follow-up of ex-premature infants. *Journal of Paediatric Child Health, 37,* 125-129.

Smith, C., Wallen, M., Walker, K., Bundy, A., Rolinson, R., & Badawi, N. (2012). Validity of the fine motor area of the 12-month Ages and Stages Questionnaire in infants following major surgery. *Physical and Occupational Therapy in Pediatrics, 32*(3), 260-271.

Smith, R. D. (1978). The use of developmental screening tests by primary-care pediatricians. *Journal of Pediatrics, 93,* 524 – 527.

Smithson, M. & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods, 11,* 54-71.

Snyder, P., McLaughlin, T., & McLean, M. (in press). Recommended practices in assessment. In M. McLean, M.L. Hemmeter, & P. Snyder (Eds.), *Essential elements for assessing infants and preschoolers with special needs*. Boston, MA: Pearson.

Snyder, P., Thompson, B., & Sexton, D. (1992, January). *Congruence in maternal and professional early intervention assessments of young children with disabilities*. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX.

Squires, J., & Bricker, D. (1991). Impact on completing infant developmental questionnaires on at-risk mothers. *Journal of Early Intervention, 15*(2), 162-72.

Squires, J., & Bricker, D. (2009). *Ages and Stages Questionnaires* (3rd ed.)*: A parent-completed child-monitoring system.* Baltimore, MD: Brookes.

Squires, J., Bricker, D., & Potter, L. (1997). Revision of a parent-completed developmental screening tool: Ages and Stages Questionnaires. *Journal of Pediatric Psychology, 22*(3), 313-328. doi:10.1093/jpepsy/22.3.313

Squires, J., Carter, A., & Kaplan, P. (2003). Developmental monitoring of children conceived by intracytoplasmic sperm injection and in vitro fertilization. *Fertility & Sterility, 79*(2), 453. doi:10.1016/S0015-0282(02)04685-X

Squires, J., Katzev, A., & Jenkins, F. (2002). Early screening for developmental delays: Use of parent-completed questionnaires in Oregon's healthy start program. *Early Child Development and Care, 172*(3), 275-82.

Squires, J., Nickel, R. E., & Eisert, D. (1996). Early detection of developmental problems: Strategies for monitoring young children in the practice setting. *Journal of Developmental and Behavioral Pediatrics, 17*(6), 420-427. doi:10.1097/00004703-199612000-00008

Squires, J., Potter, L., Bricker, D. D., & Lamorey, S. (1998). Parent-completed developmental questionnaires: Effectiveness with low and middle-income parents. *Early Childhood Research Quarterly, 13*(2), 345-54.

Squires, J., Twombly, E., Bricker, D., & Potter, L. (2009). *ASQ-3 user's guide.* Baltimore, MA: Paul H. Brookes Publishing.

Streiner, D. L. (2003). Diagnosing tests: Using and misusing diagnostic and screening tests. *Journal of Personality Assessment,* 81, 209-219.

Suen, H. K., Logan, C. R., Neisworth, J. T., & Bagnato, S. (1995). Parent-professional congruence: Is it necessary? *Journal of Early Intervention, 19,* 243-252.

Taylor, R. L. (1993). Instruments for the screening, evaluation, and assessment of infants and toddlers. In M. A. Graham (Ed.), *Implementing early intervention: From research to effective practice* (pp. 157-182). New York, NY: Guilford Press.

Teska, J. A., & Stoneburner, R. L. (1980). The concept and practice of second-level screening. *Psychology in the Schools, 17*(2), 192-195. doi:10.1002/1520-6807(198004)17:2<192::AID-PITS2310170210>3.0.CO;2-B

Thompson, L. A., Tuli, S. Y., Saliba, H., DiPietro, M., & Nackashi, J. A. (2010). Improving developmental screening in pediatric resident education. *Clinical Pediatrics, 49*(8), 737-742. doi:10.1177/0009922810363818

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1985). *Stanford-Binet Intelligence Scale* (4th ed.). Chicago, IL: Riverside.

Thurlow, M. L., & Gilman, C. J. (1999). Issues and practices in the screening of preschool children. In J. Kalesnik (Ed.), *Assessing and screening preschoolers: Psychological and educational dimensions* (2nd ed., pp. 72-93). Needham Heights, MA: Allyn & Bacon.

Troude, P., Squires, J., L'Helias, L. F., Bouyer, J., & La Rouchebrochard, E. (2012). Ages and Stages Questionnaires: Feasibility of postal surveys for child follow-up. *Early Human Development, 87*(10), 671-676.

Tsai, H. A., McClelland, M. M., Pratt, C., & Squires, J. (2006). Adaptation of the 36-month Ages and Stages Questionnaire in Taiwan: Results from a preliminary study. *Journal of Early Intervention, 28*(3), 213-225.

United States Department of Education, Offices of Special Education and Rehabilitative Services. (2010). *Thirty-five years of progress in educating children with disabilities through IDEA.* Washington DC: Author.

United States Department of Labor, Bureau of Labor Statistics. (2011). *Women in the labor force: A databook (2011 edition).* Retrieved from http://www.bls.gov/cps/wlf-databook2011.htm

Younger, M. S. (1979). *The first course in linear regression*. North Scituate, MA: Duxbury Press.

Ysseldyke, J. E., & O'Sullivan, P. J. (1987). Predicting preschool screening referral rates from district demographic data. *Journal of School Psychology, 25*(2), 119-129. doi:10.1016/0022-4405(87)90021-5

Yu, L., Hey, E., Doyle, L. W., Farrell, B., Spark, P., Altman, D. G., & Duley, L. (2007). Evaluation of the Ages and Stages Questionnaires in identifying children with neurosensory disability in the Magpie trial follow-up study. *Acta Paediatrica, 96,* 1803-1808.

Zehrbach, R. R. (1975). Determining a preschool handicapped population. *Exceptional Children, 42*(2), 76-83.

BIOGRAPHICAL SKETCH

Cathleen Pasia received her Bachelor of Science in psychology from Michigan State University in the summer of 2007. At the University of Florida, she majored in school psychology with a specialization in early childhood studies. Throughout her graduate studies, her practicum placements reflected a focus on working with young children and their families. While obtaining her doctorate degree, she worked on several research projects, including Institute for Education Sciences Goal 2 Development and Innovation grants, under the direction of Dr. Patricia Snyder. Her research interests include professional development for early childhood practitioners and family engagement in assessment, intervention, and early childhood programs. She completed her predoctoral internship at Cypress-Fairbanks Independent School District, which is accredited by the American Psychological Association. Cathleen graduated with her Doctor of Philosophy in August 2013.