Commentary on the review of measures of early childhood social and emotional development: Conceptualization, critique, and recommendations

Susan B. Campbell [a], Susanne A. Denham [b], Grace Z. Howarth [b], Stephanie M. Jones [c], Jessica Vick Whittaker [d], Amanda P. Williford [d], Michael T. Willoughby [e], Monica Yudron [c], Kristen Darling-Churchill [f,*]

[a] University of Pittsburgh, Department of Psychology, Pittsburgh, PA 15260 USA
[b] George Mason University, Department of Psychology, Fairfax, VA 22030 USA
[c] Harvard Graduate School of Education, Cambridge, MA 02138 USA
[d] University of Virginia Curry School of Education, Charlottesville, VA 22904 USA
[e] RTI International, Research Triangle Park, NC 27709 USA
[f] Child Trends, 7315 Wisconsin Avenue, Suite 1200W, Bethesda, MD 20814 USA

ARTICLE INFO

ABSTRACT

This compilation of five papers provides commentary from researchers devoted to the study of a variety of components that contribute to the broader domain of social and emotional development in early childhood. These components include social competence, emotional competence, behavior problems, self-regulation, and executive function. Each section provides a general definition of the construct, highlighting how it fits in a broader model of social and emotional development, and summarizing its relationship with a range of developmental outcomes. The papers then address developmental and contextual issues that are essential to consider when selecting a measurement tool for social and emotional development in early childhood, and discuss the field of extant measures available for each area of development. Presented intentionally as a part of a single paper, these contributions together provide a comprehensive response to the review, methods, and recommendations presented by Halle and Darling-Churchill (in this issue).

Social competence in early childhood: Challenges in measuring an emergent skill

Stephanie M. Jones and Monica Yudron, Harvard University

Social competence becomes increasingly important as infants and toddlers transition into the early childhood years and their direct interactions and engagement with peers expands. In this paper, we focus on the development and measurement of social competence in young children. However, before beginning it is important to characterize the manner in which social competence fits into the broader category of social emotional skills. In general, social and emotional competence in early childhood provides a critical foundation for the mastery of a range of skills important to successful academic behaviors and achievement (e.g., Denham, Caverly, Schmidt, & Blair, 2002; Jones & Bouffard, 2012). In consequence, social-emotional skills have been included as part of the school readiness indicators commonly used to represent, and drive improvement in, young children's ability to succeed in kindergarten and early elementary school (National School Readiness Indicators Initiative, 2005). But, what is social-emotional competence? Conceptualizing and measuring these skills is not straightforward because social-emotional competence is typically considered a broad category comprising a set of more specifically delineated skills. In Fig. 1, we present one conceptual model for the inter-relationship of three primary domains of social emotional skills: cognitive, emotional, and social. As the model suggests these skills are distinct in some ways, but are also fundamentally interwoven and reciprocally influence one another over development (Jones & Bouffard, 2012).

Defining and measuring social-emotional competence in early childhood is particularly challenging as many of these emergent skills may be indistinct from antecedent competencies and because social-emotional

---

* Corresponding editor. Tel.: +1 240 223 9236; fax: +1 240 200 1239.
E-mail address: kchurchill@childtrends.org (K. Darling-Churchill).
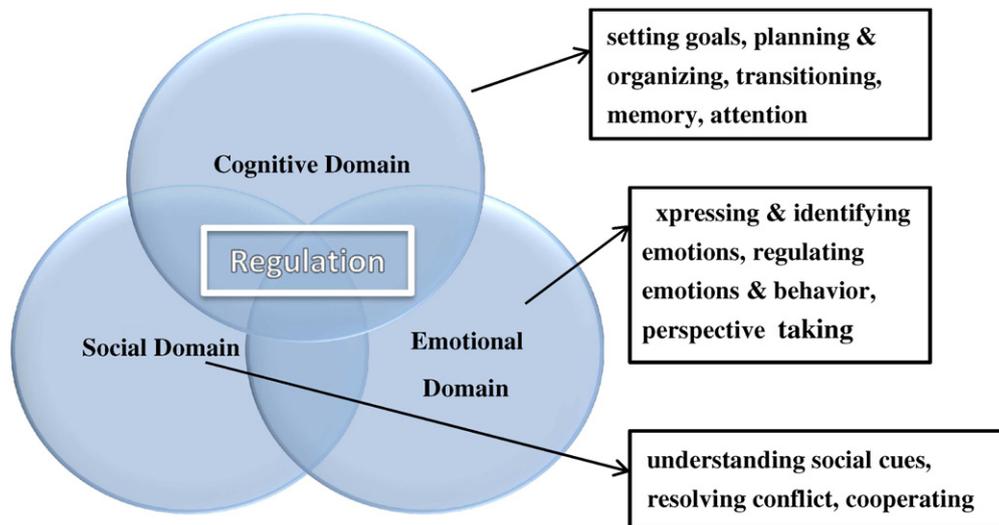
Fig. 1. Domains of social-emotional learning (SEL) and example component skills.

development in early childhood is both rapid and non-linear. In this period of life, we believe that understanding a child's social-emotional competence overall is best done by focusing on discrete skills within each of the domains captured in Fig. 1. In this commentary, we focus on conceptualization and measurement primarily within the social domain of the broader construct of social-emotional competence.

Social competence in early childhood

Early life antecedents of social competence include a child's temperament, self-regulatory skills, emotional understanding, social information processing, and communication skills (Fabes, Gaertner, & Popp, 2006). For example, a child's emotional knowledge is both antecedent to and concurrently developing and expanding in tandem with their social competence. A child's ability to identify and appropriately respond to the emotional cues of a social partner influences the success of their interaction (Trentacosta & Fine, 2009). Yet, it is the behavior that manifests in the context of the interaction, in part resulting from each child's emotional knowledge of themselves and others that represents social competence.

As such, social competence is generally characterized as the effectiveness of a child in social interactions with peers and adults (Fabes et al., 2006). Social competence is distinct from emotional or regulatory competence in that social competence is often conceptualized as the *enactment*, or behavioral manifestation, of these other competencies. In order to be socially competent, a child has the skills to (1) develop positive relationships with others, (2) coordinate and communicate her actions and feelings with social partners, and (3) recognize and regulate her emotions and actions in social settings and interactions. For example, a 4-year-old child entering a prekindergarten classroom must engage peers in order to create play experiences that are mutually satisfactory and fulfilling. In the process, this child will need to negotiate instances when the peer or peers' preferences diverge or conflict with their own. Play can only continue when these instances of conflict are navigated successfully. Self-regulation is an underlying or foundational competency which supports – enabling or impeding – children's success in specific social tasks and in the complex social milieu of early education environments.

These three basic skills are considered general skills within the social competence domain – skills that are important regardless of context. However, mastery of these general skills may be demonstrated

differently depending upon the context. Obedience and nonaggression in the face of provocation – behaviors that are adaptive in a school setting – may be a non-adaptive strategy with peers in settings where assertiveness is valued. In addition, there is some research to suggest that adults conceptualize socially competent behaviors in young children differently depending on the sex and/or race/ethnicity of the child (e.g., Nelson et al., 2013). As social competence is apparent in children's interactions with social partners, its measurement should not be divorced from the social and physical contexts in which social interactions take place.

Measuring early childhood social competence

Impetus for measuring social competence

A large body of evidence links social competence in early childhood to a range of outcomes of interest to practitioners and policymakers. For example, Downer and Pianta (2006) found that more socially competent preschool children tended to outperform their less socially competent peers in academic achievement measures administered in first grade. Social competence in early childhood has also been associated with decreased probability of problem behaviors in middle childhood and adolescence (Bornstein, Hahn, & Haynes, 2010). Social competence is likely an important lever for changing child outcomes in a number of developmental domains (behavioral and academic). Second, social competence is malleable, particularly in early childhood. In 2011, the *Journal of Applied Developmental Psychology* dedicated an entire issue to studies investigating how teachers influenced child outcomes through changes to classroom-level characteristics (Bierman, 2011). Together, these articles illustrate the powerful influence that teachers have in shaping social contexts and interactions which, in turn, enhance children's existing social skills, and support those that are emerging.

Social competence in early childhood and the preschool years is a consequence of children's history of relationships and their experiences in multiple contexts including, of course, the home environment (e.g., Dykas & Cassidy, 2011). For example, parents of young children model with their own behavior, and in interactions with their children, methods for experiencing and expressing negative emotions (Nelson et al., 2013), engaging in exchanges with one or more social partners (Feldman, Bamberger, & Kanat-Maymon, 2013), and navigating conflict (Rispoli, McGoey, Koziol, & Schreiber, 2013). The salience of a child's

home context and relationships for the development of social competence does not diminish with a child's entry into formal care or educational settings. However, non-parental caregivers in informal and formal child care and education settings can support the expansion of children's social competence skills particularly since children often engage in more frequent interactions with larger groups of peers in these settings. Furthermore, non-parental caregivers can provide important insights into a child's social competence.

Considerations for measuring social competence

There are several issues to consider when measuring social competence in young children. Generally, instruments should be appropriate to the age and developmental stage of the children involved, both in the content assessed as well as in the method used to collect the information. Some antecedents to social competence are not fully developed in very young children. For example, communication skills are thought to be important to the development of social competence. While the majority of children begin to gesture and point by fifteen months of age, few are speaking with fluidity before age 3.

Additionally, in order to reflect on the emotions of others and coordinate social interactions, children must understand that others have emotions and motivations that are distinct from their own. This ability, called theory of mind, is thought to emerge after a child's eighteenth month of life (Knudsen & Liszkowski, 2012). As children's theory of mind emerges, they are challenged to exert more effort in regulating their own emotions in social situations where peers might present evidence of different goals in shared play. For these reasons, social competence in children younger than three cannot be measured in the same way as it might in older children. Two important precursor skills that can be measured in this young age group, and which are widely believed to be critical for the development of theory of mind (and subsequent social competence), are a child's understanding of attention and of intentionality (Fabes et al., 2006). For these reasons, the majority of instruments measuring early childhood social competence are surveys administered to teachers and parents.

As children age, a different set of measurement issues arise. First, social competence is a skill best captured in dyadic or group settings. Because of this, a child's representations and perceptions of her own social competence are useful complements to teacher, peer, and parent reports of social competency. Reports that do not align across reporters can be triangulated by contextualized measures using observation protocols such as the Individualized Classroom Assessment Scoring System (inCLASS; Downer, Booren, Lima, Luckner, & Pianta, 2010). This, and related measures, capture features of social interactions and the settings in which they occur that have a direct relationship to each child's current level of social competence and the potential of the setting to enhance that social competence.

Few studies of social competence in preschool-aged children (from 3 years to 5 years) involve instruments that ask children to describe or rate their own skills or the skills of their peers. However, as Song and Wang (2013) describe in a recent paper, it is possible to adapt measures to the language and reflective capacities of young children. In this study, the *Perceptions of Peer and Self Questionnaire* (Rudolph, Hammen, & Burge, 1995) was adapted by asking children to place faces (where a smiling face indicated that a child agrees with a statement and a frowning face indicated disagreement) in buckets after the assessor read a statement such as "I make friends easily" (Song & Wang, 2013). Santos, Vaughn, Peceguina, Danial, and Shin (2014), in a study of developmental trajectories of social competence in preschool children, used photographs of classmates during interviews with individual children to elicit information about children's social networks.

Since the measure of social competence hinges on the *effectiveness* or *appropriateness* of interactions a child has with others (Fabes et al., 2006), any instruments selected ought to be sensitive to differences in social behavioral norms that may influence child interactional styles.

For example, in rating child social competence with adult social partners (i.e., teachers) the degree to which children are expected to defer to authority will influence the nature of the social exchange. In a study of preschool children from eight different countries, investigators found that conceptions of anger and anxiety in young children varied across cultural contexts (Lafreniere et al., 2002). Measurement instruments must be carefully validated for important sub-groups to which children may belong.

As suggested in recent research, characteristics of context are important in shaping social interactions among children in the setting. These characteristics include but are not limited to the: (1) emotional climate of the classroom or home environment, (2) relational style of the caregiver including parents and teachers, (3) nature of the pedagogic approach underlying classroom strategies or parent-child interaction style, and (4) behavior management norms in school and classroom or the approach to managing conflict or negativity in the home (Bierman, 2011; Rispoli et al., 2013). Direct observations of parent-child interactions in the laboratory or home are typically used to characterize the home context. There are several setting-level observational measures which attend to the relevant classroom or school features including the Classroom Assessment Scoring System (CLASS; La Paro, Pianta, & Stuhlman, 2004).

Similarly, children's social competence is likely to be influenced by compositional features of the setting in which it is observed. For example, recent research has examined how aggregated peer interactions (above and beyond individual peer relationships) establish a classroom climate that influences social-emotional and academic adjustment (Yudron, Jones, & Raver, 2014). These studies suggest that interventions can and do influence children's outcomes more broadly via "spillover" effects from one child to another. For example, Neidell and Waldfogel (2008) find that the saturation of children in Kindergarten classes with preschool experience positively influenced individual children's reading and math achievement through 3rd grade – even for children who did not attend preschool. Importantly, these authors also found that classroom-level peer externalizing behaviors negatively influence individual children's achievement. This is consistent with earlier work indicating the powerful role of classroom levels of aggressive behavior in exacerbating or mitigating individual trajectories toward aggression (e.g., Kellam, Ling, Merisca, Brown, & Ialongo, 1998).

The mechanisms underlying such effects are less clear; however there is some work suggesting that the composition of the classroom (e.g., the saturation or fraction of children with normative beliefs about aggression and prosocial behavior, or with behavioral or academic challenges) generates a set of norms and attitudes about behavior and achievement (e.g., Henry, 2008) that are linked to children's developmental outcomes. These findings have important implications for interventions because they suggest multiple mechanisms of impact: as preschool- and school-based interventions shift individual children's skills and behaviors, they also shift the regulatory and behavioral composition of classrooms, potentially enhancing direct individual effects on children both within and over time. Such findings have equal implications for measurement as it is clear from prior research that children's social interactions in aggregate are thought to establish patterns through which children's experiences and competencies influence their peers, both concurrently and in future classrooms or groupings.

Current instruments: Strengths and areas for growth

All of the instruments identified as measuring social competence in the review by Halle and Darling-Churchill (2016-in this issue) rely on teacher or parent reports of child behaviors. Instruments that rely on teacher and parent reports of child behaviors, such as the Social Skills Rating System (SSRS) or Devereux Early Childhood Assessment Clinical Form (DECA-C; Crane, Mincic, & Winsler, 2011), are simple to administer and have been used widely. Another advantage of measures like

these is that their reliability and validity has been established in a wide range of child populations.

Instruments not included in the review but contained in a longer inventory of measures (Federal Interagency Forum on Child & Family Statistics, 2015) include an assessment soliciting child feedback on a series of hypothetical and provocative interactions with peers (The Challenging Situations Task; Denham, Bouril, & Belouad, 1994) and another with child behavior ratings made by raters observing an interaction between a parent and child (Two Bags Test; Andreassen & Fletcher, 2007). These instruments were not included in the review by Halle and Darling-Churchill (2016-in this issue) because they must be administered by trained assessors, and thus are not well suited for the purpose of the original study – to identify measures for use at the federal level to develop indicators or benchmarks of early childhood social and emotional development. However, these two tools are straightforward, have been used in multiple studies, and are effective regardless of a child's verbal ability – even the Challenging Situations Task in which a child is asked to point at an image in response to a prompt.

Current measurement instruments that elicit responses from indirect observers such as parents and teachers are not well-suited to understanding situated social competence. That is, they do not anchor child social skills and competencies in the contexts in which they matter most or are likely to vary. They also provide limited information about the biases of the raters. As the literature we have reviewed in this paper suggests, it is quite important to understand how a rater conceptualizes social competence. Rater bias constrains the degree to which the social competence skills being measured are general versus specific to particular contexts. As noted above, children may have developed skills that are adaptive in some social contexts but not in others. One way to understand bias is through the use of vignettes that anchor rater perceptions of social competence.

Final thoughts on measuring social and emotional development in early childhood

Given the material presented above, we encourage those working toward developing an indicator, or set of indicators, to represent children's social competence to consider very carefully: (1) the component skills, and their developmental, stage-relevant instantiations; (2) moving beyond adult report to consider (i) vignette-based or some other performance-based assessment and (ii) direct observations using existing tools (e.g., inCLASS or Penn Interactive Preschool Play Scales [PIPPS]); and (3) the context in which, and perspective from which, such assessments or observations are generated. One promising approach to understanding child social competence development is to triangulate ratings of social competence across different instruments and raters. In a recent paper investigating trajectories of social competence across the preschool years, observers rated children's social strategies within the classroom context and their engagement in social activities. Additionally, children were interviewed about the number of friends within their classroom (Santos et al., 2014). In general, it is important to anchor the behaviors expected of young children (in this case, the behaviors understood to represent social competence) in the relevant contexts in which they learn and play, and to then closely align those with relevant, and multiple, measurement tools.

State and federal education policy can be written to increase the weight placed on school and classroom climates that support social competence development rather than hinder it. Strategies for achieving this include, but are not limited to, incorporating social competence language into early learning standards and requiring relevant training for early childhood educator certifications. In an era of increasing public interest in center-based early childhood care and education, efforts must be made to reinforce the salience of the social aspects of these contexts in the face of pressure to require mathematics- and literacy-focused curricula in programs that receive local, state, or federal funding.

Social policies have the potential to impact parents' opportunities to appropriately care for and bond with children in the first five years of life. These policies range from those that provide paid maternal and paternal leave from work in the months following a child's birth to child care subsidies that make high quality child care a reality for more families. A large and compelling body of research, briefly touched upon in this paper, highlights the importance of mother-child, father-child, and whole family interactions in the development of skills that are precursors to and aspects of social competence. Paid leave policies can act to create the opportunity for these interactions. As to equalizing the quality of these interactions at home and in formal settings of child care and education, much still needs to be done in understanding what this might look like across cultural and geographic locations.

Emotional competence in early childhood: Construct and measurement considerations

Susanne A. Denham and Grace Z. Howarth, George Mason University

Although early childhood educators and parents have long recognized the importance of emotional competence in supporting school readiness, positive relationships, and overall wellbeing and adjustment, federal and state education initiatives have focused almost exclusively on academic skills achievement (Jones & Bouffard, 2012; Raver & Knitzer, 2002). In recent policy reports, however, scholars have encouraged policymakers to consider a more holistic approach to early childhood education, one that balances positive social-emotional outcomes with more traditional academic outcomes. Specific suggestions include funding high quality early childhood education where teachers and administrators are trained and supported in social-emotional learning (SEL), improving access to early childhood education, early identification and intervention for socially at-risk children, and teaching SEL skills to children and families (Jones & Bouffard, 2012; Raver & Knitzer, 2002). Furthermore, as the focus of this Special Issue suggests, creation of SEL-inclusive early childhood programming goes hand-in-hand with accurate and complete reporting on the nation's children's social-emotional development.

Thus, for policymakers considering proposed SEL educational initiatives (and thus the allocation of considerable resources), as well as those who seek strong evidence-based data on children's SEL, a strong empirical base is important. In this commentary, we focus on one aspect of this base: the measurement of emotional competence in young children, and the relation of emotional competence to subsequent developmental outcomes. We recognize that measuring emotional competence is valuable outside the context of early childhood education, but focus on the early childhood education context due to the power of policy in shaping it.

Why is emotional competence important? Relation with subsequent outcomes

We define emotional competence as the ability to purposefully and fully express a variety of emotions, understand the emotions of self and others, and regulate emotional expressiveness and experiences when necessary. Emotional competence undergoes dramatic changes within the first five years of life, with the expression of more sophisticated, nuanced, and regulated emotions and beginning to understand, identify, and empathize with others' emotions. Hence, preschoolers express a variety of vivid, but not incapacitating, emotions and are becoming able to discern their own and others' emotional states, and to talk about them rather fluently. They are also beginning to "up-"or "down"-regulate emotions, depending on their goals (Denham, 1998; Hyson, 1994; Saarni, 1999).

Children's enduring patterns of emotional expressiveness are indicative of their temperament, a relatively stable trait; as such, it is easy to envision why emotional expressiveness is such a potent intrapersonal support for, or roadblock to, learning and interacting with age mates.

If a child is often emotionally negative, it is no wonder when his peers flatly assert, "He hits. He bites. He kicked me this morning. I *don't like* him." But the happier child's smile and body language are like beacons signaling "Come join me" to adults and age mates alike. In fact, more emotionally positive children also show better learning outcomes. At the same time, the sad, angry, or fearful child may not be well accepted by peers and teachers, and, be less able to allocate resources to learning.

Emotion regulation is the second vital aspect of emotional competence. When intensity, duration, or other parameters of the experience and expression of emotion are "too much" or "too little" to meet goals and expectations of the child and/or social partners, emotion regulation is needed. During preschool, emotion regulation becomes both necessary, due to children's increasingly complex emotionality and the demands of their social world, and possible, because of their increased comprehension and control of their emotionality. Older preschoolers begin to collaborate with caregivers' efforts to regulate their emotions. But, failures of emotion regulation can still be seen throughout the preschool period, in outbursts of temper and distress that impede social interaction. Such emotion regulation is associated with social competence (i.e., effectiveness in social interaction; Rose-Krasnor, 1997) and learning.

Emotion knowledge is the third key component of young children's emotional competence, and it is crucial for getting along with peers and adults. For example, if a preschooler sees peers bickering, and correctly deduces that one suddenly experiences sadness or fear, rather than intensified anger, she may comfort her friend rather than retreat or enter the fray. Interactions with such an emotionally knowledgeable age mate would likely render that playmate better liked. Similarly, teachers are likely attuned to the behavioral evidence of such emotion knowledge – the use of emotion language, the sympathetic reaction – and to evaluate it positively. Emotion knowledge allows a preschooler to react appropriately, thus bolstering her relationships, and contributing to a positive environment in which to concentrate on learning.

Understanding emotions that are experienced and expressed during interaction, as well as being able to regulate them, are abilities that undergird human contact. Thus, emotional competence is crucial to children's interactions and relationships with others. Furthermore, these aspects of emotional competence are key components of school readiness. For example, emotion knowledge uniquely predicts both concurrent and future social competence, classroom adjustment, and academic success in early childhood – above and beyond the influence of gender, age, and various risk factors (e.g., poverty, maternal depression) (Denham, Bassett, Thayer, et al., 2012; Denham et al., 2011; Garner & Waajid, 2008; Izard et al., 2001; Leerkes, Paradise, O'Brien, Calkins, & Lange, 2008; Shields et al., 2001). Regulated emotional expressiveness and lack of dysregulation similarly predicts such outcomes (Denham, Bassett, Thayer, et al., 2012; Graziano, Reavis, Keane, & Calkins, 2007; Howes, Calkins, Anastopoulos, Keane, & Shelton, 2003). For these reasons, emotional competence is important in early childhood classrooms and in policy-making decisions (Denham, 2006).

In fact, all aspects of emotional competence work together to promote school success (Denham, Bassett, Mincic, et al., 2012. Children who enter school with more positive profiles of emotional competence (i.e., are emotionally positive and regulated, and understand emotions) are more likely to develop positive and supportive relationships with peers and teachers, participate more in the classroom, and achieve more through the early school years (Garner & Waajid, 2008; Howes et al., 2003; Izard et al., 2001; Leerkes et al., 2008; Shields et al., 2001). Conversely, children who enter school with less positive profiles are more often rejected by peers, develop less supportive relationships with teachers, participate and enjoy school less, achieve at lower levels, and are possibly at risk for later school failure (Denham, Bassett, Thayer, et al., 2012; Herndon, Bailey, Shewark, Denham, & Bassett, 2013; Ladd, Birch, & Buhs, 1999; Raver & Knitzer, 2002).

Unfortunately, kindergarten teachers report that a significant proportion of students enter their classrooms without critical emotional competence skills (Rimm-Kaufman, Pianta, & Cox, 2000). Further, emotional competence appears to be particularly important for at-risk children. For example, high levels of emotion regulation appear to buffer family- and SES-related risk (Denham, Bassett, Thayer, et al., 2012). Thus, given the contributions of early childhood emotional competence to outcomes across developmental domains, support for the development of emotional competencies is a promising area for research, policy, and educational action. To support this goal, explication of the very *need for* and *use of* emotional competence assessments is necessary. We now discuss such issues.

The importance of measuring emotional competence

Emotional competence is not only a foundational skill for socially responsible behavior (Denham et al., 2014), but it also promotes safe and supportive learning environments, and, as we have seen, facilitates relationship building, wellbeing, and success in school. Such outcomes are desired by all stakeholders. From a policy perspective, then, an evidence base that guides curricula, as well as families, educators, and society as a whole is also important.

Research on interventions to foster emotional competence is optimistic, showing that emotional competence is a teachable skill and that early interventions are effective long term (CASEL, 2013; Committee for Children, 2011). To guide and evaluate these interventions, and to report on the progress of the nation's children, appropriate emotional competence assessment tools are needed. Measures assessing this crucial domain also inform policymakers and educators about developmentally appropriate emotions and behaviors across early childhood, and provide insight into when and how to target interventions for children who lag behind. Such knowledge can inform educational initiatives and broader policy changes aimed at maximizing positive development. Criteria for such assessment tools, corroborating and extending those presented in Halle and Darling-Churchill (2016-in this issue), are now reviewed.

Criteria for selecting and developing measures of early childhood emotional competence

As already described, emotional competence skills form integral facets of developmental tasks during the infant/toddler/preschool period. As previously noted, research suggests that, to maximize a child's performance across domains of social, academic, and intrapersonal wellbeing, educators, parents, and policymakers should ask the question: *What are the emotional competence skills of this child? How is the development of emotional competence fostered in this environment?* Means to measure these skills in both research and educational arenas are crucially needed. Adequate–and hopefully excellent–assessment tools must be identified. Although assessment tools are also needed for policy-related and research purposes, we focus on the need to use emotional competence assessments in applied settings and for evaluation (potentially serving functions of formative, and perhaps even summative assessment).

Emotional competence has historically been measured via a wide variety of mechanisms, including informant ratings, direct assessment, and observation. Rating systems, when carefully crafted, may be, in some ways, easiest to use in applied settings (as well as useful for research purposes). Our position is nonetheless that observational and direct assessments are avenues to best understand the young child. The time required for such measures, as well as resources required for training and observer/coder reliability, are hurdles to be overcome if these tools are to be practical for applied usage. To this end, recent work focuses on shortening and ultimately computerizing such observational and direct assessment tools, which we will review.

To bridge the information and understanding gaps regarding assessment of early emotional competence that exist among researchers, educators, and policymakers, we need a model of how such assessment

can be useful, a plan for using assessment, and criteria for selecting amongst extant assessment tools. Applied assessment of any domain of development is an integral, indispensable part of systems that include: (a) clear goals and benchmarks (i.e., standards); (b) evidence-based curricula and instruction, along with support for teachers to implement such programming (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011); and (c) universal and targeted screening and progress monitoring (formative and summative). Emotional competence assessment within early childhood education settings needs to be viewed within this integrated framework.

Model of the role of assessment and plan for using assessment

Fig. 2 shows our thinking on the relations among the elements in this system: (a) age-appropriate developmental tasks are the background substrate upon which these skills are demonstrated and developed; (b) standards must emerge for these important skills; (c) standards inform assessment, and vice versa; (d) both standards and assessment need to lead to instruction (often leading to further, regular assessment and revised standards, and supported by both professional development and curriculum); and (e) finally, change in SEL skill is the endpoint to which we strive. Given this hypothesized system, it is important to align standards and assessment, a crucial topic beyond the scope of this article.

We also follow Brassard and Boehm's (2007) sequence of assessment use in selecting specific tools. Ideally, screening of emotional competence strengths *and* weaknesses by both parents and teachers would be first steps; although the review by Halle and Darling-Churchill (2016-in this issue) does not include screeners, they are important in full usage of assessment tools. Screening is followed by formative assessment of specific skills and milestones. Formative assessment ("assessment *for* learning") could be more a criterion-referenced, rather than norm-referenced, endeavor. Given that some writers assert that

formative assessment need not be limited by certain requirements of summative assessments (e.g., psychometric reliability), it could be that teachers could simply use standards for formative assessment. Our position, however, is that psychometrically adequate measures should be utilized; this view guides our choice of measures. Formative assessment is followed by programming for both classrooms and parents (Domitrovich, Cortes, & Greenberg, 2007; Havighurst, Wilson, Harley, & Prior, 2009) and finally by diagnostic workups for children showing deficits in the areas. This sequence allows for implementation of a three-tiered model of instruction – from universally providing SEL instruction to all children, to targeted interventions for those at risk, and individualized work for those presenting the most persistent challenges (Hemmeter, Ostrosky, & Fox, 2006); for example, screening cut-offs can demarcate targeted and at-risk children whose educational needs may differ from those not at risk.

Finally, summative assessment ("assessment *of* learning") is often associated with high-stakes accountability assessment related to the No Child Left Behind Act. Summative assessments are often given one time at the end of the semester or school year to evaluate students' performance against a defined set of content standards. In many domains of development these are typically administered statewide, and are usually used as part of an accountability program or to otherwise inform policy; alternatively, summative assessments could also be teacher-administered tools used solely for student evaluation, or used as measures of pre-/post-change in response to programming. They are the least flexible assessments. It is not our view that any of the former uses of summative assessment are appropriate for young children. However, understanding growth and change in emotional competence at the *local level* (e.g., school, classroom) and for *research purposes* can be quite important.

The terrain of inserting tools for assessing emotional competence into the system depicted in Fig. 2 is very new and largely uncharted, especially for competence-based measures of emotional development, as
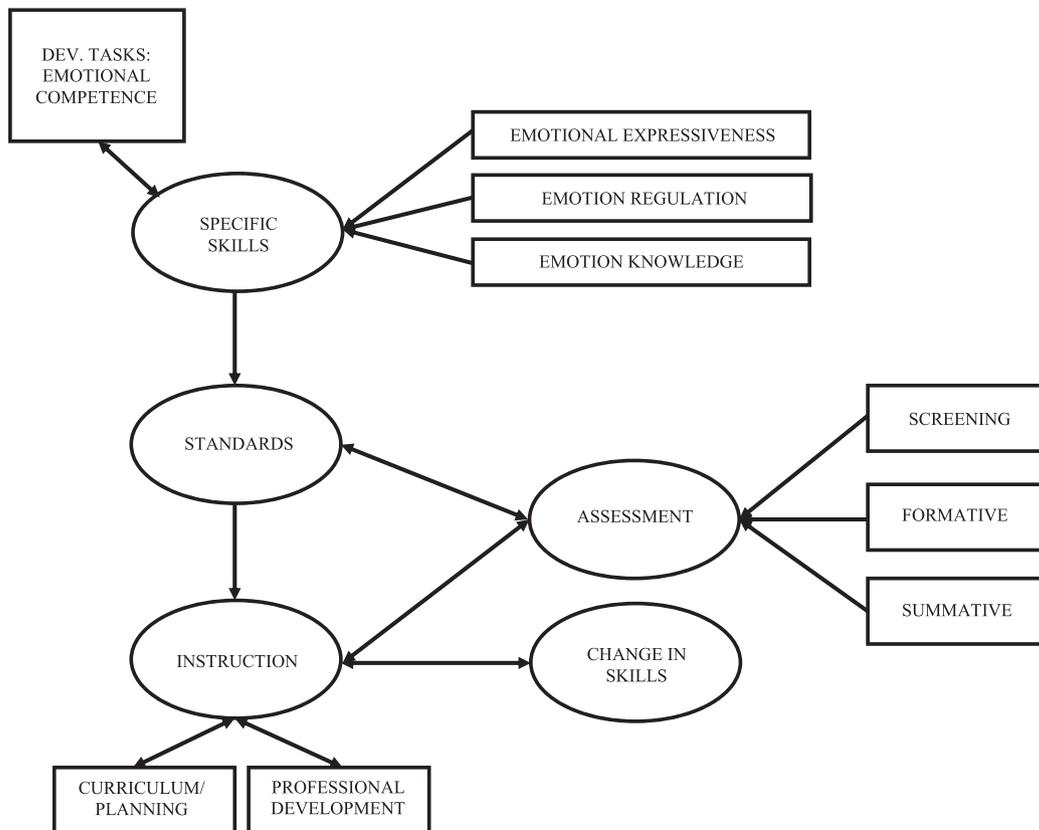
Fig. 2. Interrelated system of emotional competence development in educational settings.

opposed to measures of social behavior and behavior *problems*. To begin moving toward such a model and assessment plan, however, criteria for "best-bet" measures are outlined next, followed by choices of assessment tools and commentary.

Specific criteria for emotional competence assessment tools

We know that some states are acknowledging the importance of emotional competence skills, and their assessment. To move the field forward, we need to determine whether there are adequate extant assessment tools to make better decisions about how to facilitate children's emotional competence. Denham, Wyatt, Bassett, Echeverria, and Knox (2009) and Kendziora, Weissberg, and Dusenbury (2011) have enumerated criteria for social-emotional assessment tools specifically for applied usage (most, if not all, of these recommendations would hold for research usage). Several are paramount. First, any assessment measuring these constructs should have some documentation (e.g., a manual) that contains a description of the measure, the constructs assessed, and assignment of items to scales; it is helpful, furthermore, if descriptor text is given for each item in rating scales. The manual should make it clear whether and how the measure is useful for multiple purposes (e.g., screening, summative evaluation, evaluation of programming). This area of concern is far from systematized and is often overlooked in selecting appropriate measures for widespread use (as discussed in Halle and Darling-Churchill, 2016-in this issue). Specifically for the skills focused upon here, documentation should make clear that the assessment is appropriate for an age period from infancy through kindergarten, as well as the local population. Families of measures cutting across age periods is desirable, but age-appropriateness is more important. Moreover, because of our clear focus on emotional competence, we do not necessarily require an assessment tool to include more than one domain, although this is acknowledged as useful in a larger sense.

Second, qualities of actual assessment tools must be considered. As noted in Halle and Darling-Churchill (2016-in this issue), psychometric foundations must be excellent; assessment tools should have at least adequate reliability and validity in all their forms, and should be fair, unbiased, and generalizable across ages specified, as well as demographic groups. All measures cited herein meet psychometric requirements, but issues of fairness and generalizability require more study. Norms and psychometric data for measures must be obtained for diverse samples representing U.S. demographics, with cultural sensitivity regarding the norms for various SEL behaviors in different cultures. Native language and dialect must be considered when selecting, using, or developing and norming parent-reports.

Third, we must think about utility; each assessment should have norms or benchmarks available for interpreting results and change over time. The criteria set forth by Halle and Darling-Churchill (2016-in this issue) fail to take into account the importance of norms and benchmarks, which are fundamental to the utility of assessment tools in tracking the results of instruction and programming. Further, all such tools should be administrable within a reasonable time frame (e.g., 10–20 min or less). The acceptability of administration time in part depends on whether all children in a school or classroom, or only select children, are assessed. Collecting ratings for all students will produce as complete a profile as possible of the child's, classroom's, and school's competencies (for possible use in formative and summative assessment), but time constraints may restrict assessment to a random sample of students in each classroom. Although obtaining data on all students is ideal, restricted usage could still be valuable for formative and summative functions. Particularly in cases when teachers are unable to complete ratings on all the children in their classrooms, parent ratings could help fill in the gaps and, moreover, provide a more detailed portrait of a child's emotional competencies in settings outside of school.

Training and certification of assessors, where necessary, must be standardized and potentially repeated at intervals, to maintain quality control; we would argue that raters also must understand the constructs and methodology involved, for any assessment to be valid. Finally, where possible and for most uses, electronic administration and scoring is desired because it is both faster and less expensive than paper-based administration and hand-scoring. All of these criteria regarding utility are reflected in cost: costs of assessment tools in terms of training, completion time, skill and equipment required, test forms, and/or scoring, must be reasonable.

Selection of measures for emotional competence

No single, brief measure stands a chance of meeting all these criteria, and especially impossible is finding measures that are brief but comprehensive of all aspects of emotional competence across the pertinent age range; we must find tools that measure individual components. Crucially, we do not think it wise to pluck items on a theoretically-driven but ad hoc basis to form short indicators of unknown and probably dubious psychometric reliability and validity, despite the pressing need for such indicators. There are several extant measures, including parent- and teacher-report as well as direct assessment, which meet many criteria, capture developmentally appropriate aspects of emotional competence and could be profitably utilized.

We review such measures as they apply to the specific areas of emotional competence: emotional expressiveness, emotion regulation, and emotion knowledge. In general, the inventory of measures presented by Halle and Darling-Churchill (2016-in this issue) is quite useful; however, gaps in measurement and issues with several of the measures as pertains to fine-grained assessment of emotional competence will be noted.

Measures of emotional expressiveness

Here we discuss a number of promising measures not discussed in detail in the review by Halle and Darling-Churchill (2016-in this issue) due to not meeting stringent criteria related to the strength of their psychometrics. Overall, there are several useful measures of emotional expressiveness. These include various temperament measures (outlined below), as well as the empathy scale on the Social Skills Improvement System (SSIS; Gresham & Elliott, 2008). Further, the Minnesota Preschool Affect Checklist-Revised/Shortened (MPAC-R/S; Denham, Bassett, Thayer, et al., 2012) is considered a promising observational tool, and two short scales on the Adaptive Social Behavior Inventory (ASBI) are useful in this regard. Other somewhat less promising possibilities are also noted.

Temperament measures

We consider temperament measures, including the Rothbart family of measures (i.e., the Child Behavior Questionnaire [CBQ] and its Very Short Form [CBQ-VSF], Early Childhood Behavior Questionnaire [ECBQ], and Infant Behavior Questionnaire [IBQ]), as useful in describing an infant's or preschooler's typical patterns of emotional expressiveness across many everyday contexts (Gartstein & Rothbart, 2003; Putnam, Gartstein, & Rothbart, 2006; Putnam & Rothbart, 2006; Rothbart, Ahadi, Hershey, & Fisher, 2001). It is important to note that these measures include both *negative* and *positive* emotionality. The questionnaires are generally parent report, although some modifications for teacher report have been undertaken (e.g., Schussler, 2012).

In general, and for the CBQ in particular, two higher-order temperament factors are pertinent to the assessment of emotional expressiveness: (a) negative affectivity, and (b) surgency. Negative affectivity items involve discomfort experienced in over-stimulating situations, frustration, anger, and inability to soothe oneself, fearfulness, and sadness. It is easy to see how this potent combination could make interacting with both peers and adults problematic.

Surgency is an aspect of temperament measures with scales on approach to novel stimuli, smiling, activity, and high level pleasure. Hence, a child high on this dimension of temperament might be a lot of fun to be around — eagerly initiating contact with others, finding interesting things to do, sharing positive affect. On the other hand, there could be "too much of a good thing," with surgent children possibly seen as overly active and boisterous, risk-taking, and impulsive.

Note that separate scales are not available on the CBQ-VSF, although negative affectivity is an isolated factor (Teglasi et al., 2015). Thus, the cost of specificity is longer measures; it is therefore recommended that separate scales, which target emotional expressiveness specifically, be chosen from the very lengthy measures. Use of emotional expressiveness subscales (e.g., Sadness, Anger, Soothability, High Intensity Pleasure) could allow a focus on emotional expressiveness across both positive and negative valences. The CBQ-VSF could be used as a screener.

Social skills improvement system (SSIS)

This assessment tool covers a wide age range (3–18 years). Its stated purposes fit well within our model of assessment: (a) screening for problem behaviors; (b) identifying strengths and weakness, and identifying individuals functioning below normative expectations; (c) providing a baseline for post-intervention progress evaluation and tracking progress; and (d) gathering longitudinal research data. Thus, this may indeed be a strong choice overall. It includes the following sub-constructs: Communication, Cooperation, Assertion, Responsibility, Empathy, Engagement, and Self-Control. For our purposes, there is really only one scale, Empathy, that focuses on emotional competence; early research suggests that empathy items primarily reflect the act of being perceptive to the feelings of others (i.e., making efforts to both understand and respond to others' feelings), and stand somewhat apart from the structure of the rest of the measure. This aspect of emotional expressiveness, although circumscribed, can be very important to understand for individual students and classrooms. The SSIS is an updated version of the Social Skills Rating System, which is recommended in the review of measures presented by Halle and Darling-Churchill (2016-in this issue).

Minnesota Preschool Affect Checklist (MPAC-R/S)

The MPAC-R/S is noted as promising in the review of measures presented by Halle and Darling-Churchill (2016-in this issue) and includes 18 items. Regarding emotional expressiveness and regulation, these are scales for positive and negative emotion (e.g., "the child displays positive/negative emotion in any manner (i.e., facial, vocal, or bodily), as well as reactions to frustration (e.g., "the child promptly verbally expresses feelings arising from a problem situation, then moves on to the same or a new activity"). Other scales on productive/unproductive involvement (e.g., "the child is engrossed…emotionally invested in activity that has a positive emotional function), peer skills (e.g., "the child smoothly approaches an already ongoing activity and gets actively involved"), and prosocial behaviors (e.g., taking turns, sharing) could be useful to address social competence and in components, below. Each child is observed for four 5-min epochs across the data collection period of approximately eight weeks. Scores for items equal the sum of occurrences across all four epochs.

Structure of the MPAC-R/S shows emotionally negative/aggressive, emotionally regulated/prosocial, and emotionally positive/productive components (Denham, Bassett, Thayer, et al., 2012). In terms of psychometrics, internal consistency, test–retest reliability, and construct stability are good. Age, gender, and risk differences have been found, as well as relations with other aspects of emotional competence (i.e., children's emotion knowledge contributed to later emotionally regulated/prosocial behavior) and achievement (i.e., preschool emotionally negative and aggressive behaviors contributed to concurrent and kindergarten school success) (Herndon et al., 2013). The measure is now computerized.

The MPAC-R/S is a good example of an observational measure to glean information on emotional competence. In many cases, important information about children is acquired by watching them, and given that the MPAC-R/S requires just 20 min, it could provide information relatively economically. Moreover, adaptation of the measure for use by teachers is ongoing. Empirically supported observation measures of social-emotional behaviors are important for educational and other applications, and their development should be encouraged.

Other scales

The Adaptive Social Behavior Inventory (ASBI; Hogan, Scott, & Bauer, 1992) consists of 30 items addressing Express, Comply, and Disrupt factors. Although it has many positive attributes (and could likely be used in other domains), the only emotional competence items referring to emotional expressiveness include "Confident" and "Proud". There is one item on emotion knowledge, and one on emotion regulation.

The Infant-Toddler Social-Emotional Assessment (ITSEA; Carter, Briggs-Gowan, Jones, & Little, 2003) is one of the six measures recommended by Halle and Darling-Churchill (2016-in this issue) and does have psychometrically adequate scales on empathy and emotional positivity, for parent report of children 12 to 36 months-old. Empathy items (e.g., "tries to help when someone is hurt", "is worried or upset when children cry", "tries to make you feel better when you are upset") form an internally consistent scale, Three items on emotional positivity include "laughs easily or a lot", "is affectionate with loved ones", and "smiles a lot" (marginally internally consistent).

Gaps

It is clear that one potential gap to be addressed is *short* expressiveness-focused scales that include positivity for teacher-report. The Positive and Negative Affect Scale (PANAS; Watson, Clark, & Tellegen, 1988) could be adapted for such use. It assesses how the child/ "feels on average," for 12 negative emotions (e.g., sad, angry) and 3 positive emotions (e.g., excited, enthusiastic). The PANAS has been modified for parent report for preschoolers and early elementary grades, and examines overall emotional expressiveness. It is brief, and has excellent psychometric properties.

Summary of recommended measures of emotion expressiveness

The Rothbart family of temperament scales, the SSIS, ASBI, and ITSEA for restricted use, and the MPAC-R/S observational tool, as well as the PANAS appear promising. All the measures have adequate psychometrics. In order to meet the need of brevity, selected scales of the Rothbart measures or the CBQ-VSF screener would need to be used, and the MPAC-R/S could potentially capture emotional expressiveness in a briefer time, although additional research to verify reliability and validity of using fewer observation periods would be necessary. Some elements of necessary documentation may be somewhat sparse for the ASBI and ITSEA. Formative and summative assessment could be accomplished with all measures.

Measures of emotion regulation

Overall, there are several useful measures of emotion regulation. These include various temperament measures (although there are caveats), as well as the Emotion Regulation Checklist (ERC). Further, the Minnesota Preschool Affect Checklist-Revised/Shortened (MPAC-R/S) is considered a promising observational tool (already described). Other scales in the inventory reviewed by Halle and Darling-Churchill (2016-in this issue) are seen as less useful because they do not as directly or psychometrically adequately assess specifically *emotion regulation*, but several are mentioned here.

Temperament questionnaires

The Rothbart family of temperament questionnaires also addresses regulation as a key factor. The effortful control factor encompasses

scales measuring inhibitory control, maintenance of attentional focus, low intensity pleasure, and perceptual sensitivity. Effortful control is often considered a close proxy for emotion regulation, and is associated with sensitivity to the emotional experiences of peers. Nonetheless, in our view this factor is not a direct enough measure of emotion regulation to be as useful as some consider it. Thus, the CBQ's scales related to emotion regulation, or internally consistent abbreviations thereof, could be useful, but caveats would need to be stated.

Emotion Regulation Checklist (ERC)

The Emotion Regulation Checklist (ERC) (Shields & Cicchetti, 1997; Shields et al., 2001) is a 24-item teacher report including items on lability, intensity, valence, flexibility, and appropriateness of emotions expressed. The measure consists of two subscales, "Lability/Negativity (e.g. "is prone to negative outbursts") and "Emotion Regulation" (e.g. "is empathetic to others"). This measure demonstrates high internal consistency for each subscale, as well as concurrent and predictive validity (Shields & Cicchetti, 1997; Shields et al., 2001). More specifically, the ERC distinguishes regulated from dysregulated children. Further, emotion regulation scores predict later school adjustment, whereas emotional lability/negativity predicts poorer outcomes. Despite these strengths, it should be noted that the measure conflates regulation with expressiveness within its subscales; if an overall summary of these two aspects of emotional competence, focusing on negative expressiveness and its control, is warranted, this is a good choice for teacher report. It could likely be adapted for parent report.

Other scales

The Devereux Early Childhood Assessment-Clinical Form (DECA-C) is another of the six measures recommended by Halle and Darling-Churchill (2016-in this issue) and a nationally normed assessment that evaluates within-child protective factors of preschool children (LeBuffe & Naglieri, 1999) – attachment, self-control, and initiative. The DECA-C can be completed by both parents and teachers. Its results provide the examiner/researcher with competence-based information that may be useful for educational planning. Nonetheless, only a few self-control items refer to emotion regulation (e.g., "can calm herself/himself down when upset"), suggesting that this measure might be more useful for other domains.

Summary of recommended measures of emotion regulation

The ERC is the only emotion regulation measure that appears promising. It has moderate to adequate reported psychometrics, and is brief. Documentation as put forward here is sparse, but formative and summative assessment (and potentially screening) could be accomplished using this measure. Clearly there are gaps necessitating research on both parent-/teacher-report ratings systems, as well as any sort of brief but ecologically valid observation for assessing early childhood emotion regulation.

Measures of emotion knowledge

This area of emotional competence is relatively sparsely measured. It is likely that it is difficult to report on exactly what a child understands about emotions, because some inference would be required. There is one teacher- or parent-report measure, but direct assessment seems more appropriate for this reason.

Teacher- or parent-report

Only the ITSEA's earliest version (Briggs-Gowan & Carter, 1998) has a small subscale (3 items) that examines emotion knowledge, or as labeled, Emotional Awareness. Although the ITSEA as a whole is an excellent measure, extracting three items (e.g., asking teachers or parents if the child: "talks about own feelings", "talks about feelings of others", "aware of others' feelings") is less than ideal. On the other hand, the

scale is internally consistent for two-year-olds, but due its subsequent exclusion from the ITSEA, validity evidence is lacking.

Direct Assessment: The Affect Knowledge Test, Short Version (AKT-S)

The AKT-S (Denham, Bassett, Brown, Way & Steed, 2013; there are two parallel versions) utilizes puppets to measure preschoolers' developmentally appropriate understanding of emotional expressions and situations, and has recently been computerized. Changes (e.g., shortening) not captured in one inventory of measures (Federal Interagency Forum on Child & Family Statistics, 2015) render it more useful for both research, policy-related, and educational purposes.

Children's understanding of emotion is assessed using puppets with detachable faces that depict happy, sad, angry, and afraid expressions. First, children are asked to both verbally name the emotions depicted on these faces, and then to nonverbally identify them by pointing. This procedure taps their ability to recognize expressions of emotion. Then, in two subtests of emotion situation knowledge, the puppeteer makes standard facial and vocal expressions of emotions while enacting emotion-laden stories, such as fear during a nightmare or happiness at getting some ice cream. Children place on the puppet the face that depicts the puppet's feeling in each situation. In the first subtest, the puppet feels emotions that would be common to most people, such as those mentioned above. In the second subtest, children are asked to make inferences of emotions in equivocal situations. This individualized subtest measures how well children identify others' feelings in situations where the "other" feels differently than the child, and which could easily elicit one of two different emotions in different people, as in feeling happy or afraid to get into a swimming pool.

The AKT-S and its computerized version require about 10 min to administer and have been studied with three diverse samples totaling over 900 children. Reliability and validity are good to excellent (Denham, Bassett, et al., 2013; Denham, Way, Kalb, Warren-Khot & Bassett, 2013; Denham, unpublished data). The longer version upon which the AKT-S was adapted shows similar good-to-excellent reliability (internal consistency and construct stability), with validity from a wide variety of sources, both in terms of early school success and social competence (see also Denham, Ji, & Hamre, 2010).

Gaps and summary of recommended measures of emotion knowledge

It is clear that there is a dearth of useful measures of emotion knowledge. It is unclear whether parent- and teacher-reports could ever garner a level of detail on this aspect of emotional competence that exceeds that of the ITSEA, due to the amount of inference required; thus, direct assessment is seen as more appropriate. The AKT-S shows good properties, with appropriate documentation, and ability to be used for both formative and summative assessment, especially given that it includes parallel versions that could be used for multiple assessment points. Research supporting development of other means of assessing preschool emotion knowledge would be valuable.

Considerations for emotional competence measurement in diverse populations: Developing new measures

Given the paucity of good measures focusing clearly on the dimensions of the emotional competence domain, and the need to have assessment tools that meet our model of assessment and concomitant plan for early childhood education assessment, it is clear that the area is ripe for measurement development. In designing and selecting measures for emotional competence, there are several issues that should be considered. First, measures must be developmentally appropriate, with knowledge of what to expect of younger children during this period. Across early childhood, direct assessments need to be unambiguous and allow children the option of responding non-verbally (Denham, Way, et al., 2013). Additionally, gender often affects children's emotional competence, with boys often expressing more negative emotion during play or providing more dysregulated responses (Denham, Bassett,

Thayer, et al., 2012). The physicality of young boys' play and other developmentally normal gender differences are worth considering when judging classroom behavior and responses to direct assessments. Further, children's differing temperaments must be taken into account when judging their emotional competence, as their enduring behavioral propensities may impact their emotional competence. Children with varying developmental disabilities may have difficulty remembering emotional experiences and associating them with events, as well as detecting and understanding signals of emotion, may send incongruent or confusing emotional signals, and often have cognitive impairments that slow the development of their emotional competence.

Last, it is important to consider the social and cultural backgrounds of children. There are often differences in the way individuals from different cultures express, understand, and socialize emotions. For example, children from Asian cultures have been found to express emotions and understand facial expressions differently, and receive less discussion of emotions and more criticism of the expression of emotions from parents, than children from the U.S. (Denham, Mason, Kochanoff, Neal, & Hamada, 2003; Wang, 2003; Watanabe, Kobayashi, Bassett, & Denham, 2012). Further, children at-risk due to poverty often demonstrate compromised emotional competence during this age range (Denham, Bassett, Mincic, et al., 2012; Denham et al., 2011; Garner & Waajid, 2012; Shields et al., 2001). All of these noted considerations necessitate care in assessment, with ecologically valid, and where necessary, adaptive methods, as well as norming sensitive to these differences.

Conclusions

The state of assessment for emotional competence, within the specific parameters of emotional expressiveness, emotion regulation, and emotion knowledge components, is promising but far from completely adequate to meet early childhood education, policy, and research purposes. In this article we have defined our domain, explicated the importance of early emotional competence with regard to developmental outcomes of social competence and early school success. We then put forward a comprehensive model for using such assessment specifically in early childhood contexts, expounded on criteria for adequacy of measures, and reviewed measures for each emotional competence component, noting gaps and needs for future research. Finally, we noted specific needs for developing new measures for use with diverse populations, very important in our changing nation. It is our hope that these arguments will be useful in future research and applications.

Measuring social and emotional development in early childhood: Should problem behaviors be included?

Susan B. Campbell, University of Pittsburgh

The goal of this special issue is to delineate measurement strategies for those aspects of social and emotional development in early childhood – i.e. from toddlerhood to the transition to kindergarten – that predict positive adjustment and success in school. This inquiry will provide information about children's social and emotional development that can inform policy makers and educators concerned with children's well-being at the federal, state and local level as they consider programs meant to support development in early childhood and inform school readiness initiatives. In addition, the larger field of researchers will also learn more about considerations and gaps when developing and using measures of social and emotional development.

The emphasis on measuring strengths and competencies is a welcome counterweight to the more usual focus on behavior problems and psychopathology. There is wide agreement that social and emotional competencies are considered key components of both positive adjustment and school-readiness over and above language and other cognitive skills (e.g., Ladd, Herald, & Kochel, 2006; Raver & Zigler, 1997), whereas behavior problems are among the major impediments

to academic success in young children (e.g., Rimm-Kaufman et al., 2000). Thus, a comprehensive approach to describing and understanding young children's development for the purposes of informing policy and program decision-makers by necessity should include an assessment of both social and emotional strengths and of problem behaviors that reflect poor regulation of behavior and emotion. Although my focus for this commentary is the measurement of behavior problems in young children, I first discuss theoretical and conceptual issues, thereby placing early behavior problems in context and then consider how competence and behavior problems are related to one another. This is especially important given the goal of assessing children's functioning from the perspective of "well-being" and school readiness, with an emphasis on normative development, rather than psychopathology. The question of how these various subdomains overlap and fit together as well as how they change with development is another important issue to consider.

A developmental and conceptual framework

As noted above, discussion of social and emotional development in young children may seem incomplete without some attention to behavior problems, but problems can be meaningfully assessed only within the broader context of other child characteristics (e.g., language and cognitive development, social and emotional competence), parenting quality and the family environment, community resources, and cultural expectations. That is, behavior problems in young children must be conceptualized in terms of the transactional (Sameroff, 2009) and ecological (Bronfenbrenner, 1977) models that define developmental psychopathology (Cicchetti & Cohen, 1995; Cummings, Davies, & Campbell, 2000; Shonkoff & Phillips, 2000; Sroufe, 1990). A transactional model underscores the constantly shifting bidirectional influences between the developing child, whose needs and competencies are changing rapidly during infancy and early childhood, and the caregiving environment that also changes and is changed by the developing child. Within this framework attachment theory specifically addresses the quality of the relationship between the young child and primary caregivers (Ainsworth, Blehar, Waters, & Wall, 1978; Bowlby, 1969). The child's use of the caregiver as a secure base for exploration, the sensitivity of the caregiver to the child's changing needs, and the child's internal representations of and expectations for relationships will help to shape the child's emerging sense of self, social competence, and ability to regulate emotion (Laible & Thompson, 2007). More recent conceptualizations of young children's social-emotional development emphasize the interplay between children's biological and personality characteristics (i.e., temperament) and the caregiving environment, suggesting that some children are more sensitive than others to childrearing practices and social context (Belsky & Pleuss, 2009; Boyce & Ellis, 2005). The ecological model, in turn, places these dynamic developmental, parent-child, and family processes in the broader context of the neighborhood, community, and culture in which the child and family reside.

A developmental perspective on emerging social competence and behavior problems also must take into account the profound changes in cognitive, social, emotional, and communicative development that occur over the first five years of life. Thus, emerging social competence reflects the mastery of stage-salient developmental tasks and challenging transition points; in contrast, behavior problems often reflect difficulties negotiating the tasks and transitions that characterize typical development during infancy and early childhood (Cicchetti, 1990; Sroufe, 1990). During infancy these include establishing routines, regulating impulses, communicating wants and needs, and establishing attachment relationships with primary and secondary caregivers. During toddlerhood and the preschool period these include the development of language, social cognition, and emotional expression, negotiating separation and reunion, establishing autonomy and a sense of self, and forming relationships with peers and preschool teachers (Campbell, 2002). Normative events like the birth of a sibling or entry into childcare

may facilitate emerging social competence or tax a child's resources, partly as a function of the child's temperament and the support and scaffolding available from caregivers. Thus, both behavior problems and social competence must be considered within this broader developmental, family, and ecological context. This is especially important during infancy and early childhood when young children are dependent on parents and other caregivers to meet their rapidly changing developmental needs and to help them adapt to changes, and when signs of problems are most often unstable.

Subdomains of social and emotional development

The subdomains identified for measurement in this special issue are important for understanding young children's development and have clear implications for school readiness and children's academic functioning. At the same time, it is important to recognize their interrelatedness and overlap and to ask how they fit together in predicting positive adjustment and school success. Presumably children high on social competence as defined in the review by Halle and Darling-Churchill (2016-in this issue) will also be relatively high on indicators of emotional competence and self-regulation. For example, children who easily make friends, share toys, and take turns and are also cooperative with teachers (social competence) are likely to be able to understand the feelings of playmates (emotional competence) and to appropriately regulate their own positive and negative emotions (self-regulation). Thus, although these subdomains are considered helpful ways of organizing our concepts of social and emotional development, they are highly inter-related and measures of these subdomains are likely to be moderately to highly inter-correlated.

Do these behaviors indicative of social and emotional competence also reflect the absence of behavior problems? Several examples highlight the complexities inherent in trying to tease apart the constructs of social competence, emotional competence, and self-regulation from behavior problems. To what degree are the specific behavior problems that are often identified in young children, such as poor impulse control (can't wait, grabs toys), poor attentional control (can't pay attention for more than a few seconds), aggressive behavior (hits other children), and non-compliance (doesn't listen, defiant), reflections of poor self-regulation and limited social competence? Does the inclusion of a measure of behavior problems add to the prediction of either adjustment or school readiness over and above measures of social and emotional competence and self-regulation? This is a critical question and suggests the importance of examining profiles of functioning across these subdomains. Might some children who are low on measures of problem behavior also be low on measures of competence because they are quiet and withdrawn? Might some children who are boisterous and not well regulated receive high ratings on behavior problems, but also high ratings on some indicators of social engagement and positive affect? How might these children differ from those who are low on behavior problems and high in competence in terms of school readiness? Are there implications of these different profiles for school readiness and targeted interventions? These are questions that seem worthy of exploration.

Some recent research demonstrates not only that children with better social skills do better on measures of pre-reading and math skills, but that social competence explains unique variance in academic competence, in some cases mediating effects of problem behaviors partially or entirely. For example, McWayne and Cheung (2009) found that behavior problems assessed at the end of the Head Start year did not predict academic or social outcomes in first grade once competencies, including positive interaction with peers, attentional focus, and motivation for learning were considered. Similarly, Arnold, Kupersmidt, Voegler-Lee, and Marshall (2012) studied a diverse sample of preschool children and examined the links between attention problems, aggression, and social skills as well as how these constructs predicted pre-academic skills including literacy and number concepts. There

were moderate associations among behavior problems and social skills, with attention problems the strongest predictor of poorer pre-academic functioning. However, better social skills contributed independent variance to the prediction of pre-academic functioning even with both aggression and attention problems controlled. Thus, both social skills and attentional difficulties were associated with academic performance, whereas teacher ratings of aggressive behavior were not, possibly reflecting the less frequent and more time-limited nature of aggression in preschool classrooms.

Given findings such as these, Bulotsky-Shearer, Manz, et al., (2012) have argued that positive peer interactions in preschool can act as a protective mechanism fostering a range of skills that support school readiness during naturally occurring social exchanges. In the context of positive peer play, children may learn to regulate their own behavior, attend to the behavior of others, cooperate and share, explore toys that facilitate learning, engage in reciprocal conversations, listen to a peer's ideas and make up play scenarios that foster creativity. Positive interactions with peers and teachers may also support motivation for learning. Indeed, Bulotsky-Shearer, Bell, Romero and Carter (2012) reported that the associations between Head Start children's problem behaviors and learning outcomes were partially explained by their interactive play skills. Consistent with this argument, Nix, Bierman, Domitrovich, and Gill (2013) tested the effects of an intervention with Head Start children meant to improve social competence and emotion understanding as well as emergent literacy; they reported that gains in social and emotional competence were associated with better pre-academic skills over and above gains in language measures. Taken together, these studies highlight the importance of a strength-based focus on early development, with problem behaviors considered within a broader framework of social and emotional competence.

Why measure behavior problems in young children?

Although it is possible that measures of behavior problems will not add much variance to the prediction of outcomes in a large normative sample of young children once social and emotional competence and self-regulation are taken into account, measures of problems may be important in identifying children at risk for serious problems. That is, when problems do predict outcomes, the results may be driven by the few children at the extremes of the distribution. Thus, some measure of behavior problems in toddlers and preschoolers is worth including in any larger-scale screening because early-emerging problems, evident in late toddlerhood or the early preschool period *may* persist and become entrenched over the course of development. Problems that are *stable* across the preschool period *may* cascade into more serious and debilitating adjustment difficulties in childhood and adolescence that take their toll not only on the child and family, but also prove costly to society by taxing the resources of a number of social systems including the health and education systems, the child welfare system, and the juvenile justice system. These links between *serious and stable* problems in early childhood, especially in the context of high levels of family stress, and later outcomes have been substantiated in numerous studies over the past several decades (e.g., Campbell, Pierce, Moore, Marakovitz, & Newby, 1996; Greenberg et al., 1999; Moffitt, Caspi, Dickson, Silva, & Stanton, 1996; Shaw, Gilliam, Ingoldsby, & Nagin, 2003; Shaw, Hyde, & Brennan, 2012). Thus, early identification and early intervention are important goals to pursue.

At the same time it is important to remember that many problems evident early are transient, reflecting a difficult developmental transition, a reaction to a stressful life event such as the birth of a sibling, entry into child care, or family turmoil as noted above. The recent interest in identifying and diagnosing behavior problems in very young children is, therefore, a double-edged sword. On the one hand, it is important to provide help and support for young children and families dealing with stressful life events and transitions. On the other hand, and especially during a time of shrinking resources and limited services,

it is particularly important to identify children who are truly at risk for serious and debilitating problems that are likely to continue well beyond the preschool years. Concerns can be raised about both over-identification of problem behaviors that are time-limited and under-identification of serious problems. Studies using only cross-sectional designs run the risk of providing incomplete and potentially inaccurate information on the prevalence of serious behavior problems in young children. Longitudinal studies are necessary to differentiate between transient adjustment difficulties and more persistent and potentially severe problems. Furthermore, in line with the arguments in the review presented by Halle and Darling-Churchill (2016-in this issue), problems and competencies are best assessed together to provide a more complete picture of children's strengths and difficulties in early childhood and during the transition to school.

Problems in young children are more likely to persist beyond preschool age when they are more severe and chronic in early childhood, and evident across situations and relationships (Campbell, 2002). In addition, decades of research indicate that the family environment is the crucial ingredient in helping young children overcome early problems or in exacerbating early difficulties (e.g., Belsky, Hsieh, & Crnic, 1998; Kochanska, Philibert, & Barry, 2009). Behavior problems in young children do not occur in a vacuum and they do not reside in the child as such, but they reflect the wider family and social environment. A behavior problem is not something the child "has" like an infection, but a way of adapting to a challenging environment (Sroufe, 1997). Thus counting symptoms or problem behaviors will not provide much useful information in the absence of complementary information on the child's family and social context, and the balance of risk and protective factors that will determine long-term outcomes. Emphasis on assessing social and emotional strengths, therefore, provides a different lens through which to view problem behaviors, given the focus on competence, resilience, and child well-being.

That said, we know unequivocally that poor prenatal care and pediatric care, chronic poverty, poor nutrition, high levels of family stress, parent depression, lack of warm, responsive, and stimulating caregiving in infancy and early childhood, and the presence of either harsh or disengaged parenting are among the most robust predictors of persistent behavior problems in young children (e.g., Campbell, Shaw, & Gilliom, 2000; Deater-Deckard, Dodge, Bates, & Pettit, 1998; NICHD Early Child Care Research Network [ECCRN], 2004, 2005; Shaw et al., 2012; Shonkoff & Phillips, 2000). Thus, we need to focus on how to help families who are struggling because of limited material and other resources so that they can support young children's social, cognitive, and emotional development. This means building on child and parenting competence and helping parents to set reasonable and age-appropriate limits as children enter the "terrible two's" (e.g., Love, Chazen-Cohen, Raikes, & Brooks-Gunn, 2013; Shaw, Dishion, Supplee, Gardner, & Arnds, 2006). The assessment of problem behaviors in tandem with measurement of social competence and emotion regulation will provide important information on children's adjustment more broadly conceived and have implications for preschool intervention and family support programs.

It is also important to recognize that many children living in stressful circumstances overcome adversity. Moreover, the studies cited above demonstrate that social skills are associated with resilience and academic success in children at risk and that interventions meant to scaffold social and emotional competence along with engagement in learning (Nix et al., 2013) appear to promote school readiness as children transition to kindergarten.

Conceptualization and measurement of behavior problems

The measurement of behavior problems in children using parent and teacher ratings has a long history (Achenbach & Edelbrock, 1978). With the advent of computers and advanced statistical techniques, most notably factor analysis, scores of studies using somewhat different instruments have identified the broad dimensions of internalizing (anxiety, depression, social withdrawal) and externalizing problems (oppositional and aggressive behavior, conduct problems, attention deficit disorder) (see Achenbach & Edelbrock, 1978; 1981; Achenbach & Rescorla, 2000). This work in the field of child psychopathology was initially designed to identify children with serious problems or those at risk for disorder, as well as to contribute to the assessment and diagnosis of children already referred to mental health clinics. In addition, these measures were initially developed for and normed on school-aged children and adolescents. More recent interest in the mental health and adjustment of younger children has led to the development of measures of problem behavior in toddlers and preschoolers (Achenbach & Rescorla, 2000; Carter & Briggs-Gowan, 2006). Much of this work is also geared either to the early identification of problems or to clinical assessment. Given this focus, scoring algorithms for measures such as the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2000) provide little information at the typical end of the distribution, with scores becoming more differentiated once elevated rates of problem behaviors are reported. While the CBCL is included in the review by Halle and Darling-Churchill (2016-in this issue), they also note its solely negative focus and urge caution in relying upon it for use as an indicator of typical child development in the social and emotional arena.

In contrast, the field of developmental psychopathology has emerged over the last several decades (e.g., Cicchetti & Cohen, 1995; Cummings et al., 2000; Sroufe & Rutter, 1984) with a focus on identifying risk factors for adjustment problems in young children, and also protective factors that counteract risk. From this perspective, assessment instruments need to measure the range of competencies and problem behaviors that characterize early development and also allow for the assessment of developmental change.

Conceptualizing adjustment and maladjustment in young children

Adjustment indicators in infancy (0–24 months)

How early can one meaningfully assess behavior problems? I would argue that the concept of behavior problems in children prior to the second birthday, except in very extreme circumstances, is not meaningful. As noted above, during infancy and toddlerhood (0–2), major developmental tasks include establishing routines, regulating emotions and behavior with adult support, and forming attachment relationships with primary caregivers (Kopp, 1989; Sroufe, 1997). Thus, infant behavior is best considered within the context of the relationship between the child and primary caregivers. Problems in these areas are often reflected in irritability, extreme fussiness and difficulty either self-soothing or being soothed, lack of social engagement, and high levels of fearfulness. These dimensions of behavior are best captured under the rubric of *temperament* rather than behavior problems. Measures of infant and toddler temperament assess individual differences in reactivity and regulation (Rothbart, 2007); they are seen as aspects of behavior that underlie basic personality (e.g., extraversion) with implications for later adjustment (e.g., negative affect), executive functioning (e.g., effortful control), and social competence (e.g., sociability).

However, it cannot be overemphasized that dimensions of temperament are only moderately stable in early childhood, are transformed by the caregiving environment, and predict behavior problems only at the extremes and in the context of a poor fit between the child's needs and the caregiving environment (Bates & Pettit, 2007). Thus, infant temperament, measured during the second half of the first year, predicts later behavior problems in some children, but high levels of negative affect and poor self-regulation are likely to predict later difficulties primarily in the context of insensitive or overly harsh parenting; in contrast, temperamental difficultness can be transformed into positive adjustment when parenting is a good fit with children's needs (e.g., Belsky et al., 1998; Kochanska et al., 2009). In the second year, problems with regulation of emotion and behavior may become more challenging to

parents and other caregivers as children become more autonomous, mobile, and verbal, and as they struggle to balance their desire for independence with their developmental limitations. Non-compliance and limit-testing are frequent in the second year, but they too predict serious and continuing problems only at the extremes and in the context of high levels of parent-child conflict and inconsistent parenting (Belsky, Woodworth, & Crnic, 1996; NICHD ECCRN, 2004).

Measures of infant and toddler temperament are readily available and they capture many of the dimensions of behavior that are pertinent to early adjustment (a subtle distinction between adjustment and problems seems pertinent). The most widely used measures are variations on the Infant Behavior Questionnaire (IBQ, Gartstein & Rothbart, 2003). The dimensions captured in the IBQ include Extraversion or surgency (activity level, pleasure, approach, and sociability), Negative Affect (frustration, fear, discomfort, sadness, and distress to limitations), and Effortful Control (attentional focus, soothability, and cuddliness), all of which are appropriate to measure in a general survey of infants and toddlers, at least between 8-9 and 18 months; although the measure can also be used with somewhat younger infants, it is less likely to be stable the earlier it is measured. The IBQ has moderately rated psychometric properties, covers multiple aspects of social and emotional development above and beyond temperament, and relies on parent report, though one negative feature is its length, with 191 items. Of the measures reviewed by Halle and Darling-Churchill (2016-in this issue), the Infant Toddler Social Emotional Assessment (ITSEA; Carter & Briggs-Gowan, 2006) has very good psychometric properties, offers both a parent and teacher report form, and assesses broad dimensions of competence and problem behavior, although it is also long, with 166 items. Its companion instrument, the 42-item Brief-ITSEA (BITSEA; Briggs-Gowan & Carter, 2002),offers potential advantages in this regard, although it is a *screener* for social-emotional/behavioral problems and delays in competence, and not an assessment. As discussed by Halle and Darling-Churchill (2016-in this issue), screeners are not designed to collect nationally representative information for use as indicators or benchmarks.

Adjustment indicators in toddlerhood and preschool

By age 2, as children begin to use language to communicate and to become more autonomous beings, it is possible to assess problem behaviors as well as emotion regulation and social competence. However, at age 2, struggles over autonomy versus dependence, the need for more consistent limit-setting by caregivers, and children's increasing mobility, verbal skills, and interest in peers may be reflected in transient aggression toward playmates, difficulty sharing toys or taking turns, poor regulation of anger and frustration, defiance of adult requests, and temper tantrums. When do these behaviors reflect the "terrible twos" and when are they signs of more persistent problems?

Only a small number of children showing these behaviors at age 2 will continue to evidence serious problems by ages 3 or 4. However, persistent problems are more likely when children are showing quite severe and pervasive aggression, tantrums, and outright defiance that co-occur (Campbell et al., 2000; NICHD ECCRN, 2004). Moreover, problems are more likely to become chronic in the context of harsh parenting on the one hand, and lower levels of positive, engaged and proactive parenting on the other; other indicators of family risk are also important to consider, including parental depression, antisocial behavior, and substance use (e.g., Campbell et al., 2000; Deater-Deckard et al., 1998; NICHD ECCRN, 2004; Shaw et al., 2012). As children develop better language skills and as parents learn to redirect behavior and appropriately scaffold self-regulation, tantrums and aggression decline. Thus, although behavior problems can be identified in young children, it is important to take severity, timing, and social context into account in deciding whether a problem is likely to be transient or a sign of more long-term and serious difficulties that may continue into middle childhood.

With the goal of measuring children's adjustment rather than psychopathology, instruments that reflect both strengths and problem behaviors seem more appropriate than those meant to contribute to clinical assessment. Thus, measures of child temperament, such as the Children's Behavior Questionnaire (CBQ, Putnam & Rothbart, 2006; Rothbart et al., 2001) will provide data on dimensions of child behavior reflecting emerging competence (e.g., effortful control, extraversion) and adjustment difficulties (i.e., negative affect as reflected in shyness, sadness, fear, anger/frustration) during the preschool period and the transition to school. Two measures highlighted in the review by Halle and Darling-Churchill (2016-in this issue) include scales assessing social and emotional competence as well as adjustment problems, thereby providing a more nuanced and complete picture of children's functioning across these subdomains. These measures include the Behavior Assessment System for Children-2 (BASC-2, Reynolds & Kamphaus, 2004) and the Devereux Early Childhood Assessment Form (DECA, LeBuffe & Naglieri, 2013). While the DECA is slightly advantaged with only 62 items, compared with over 100 for the BASC-2, they both meet strong standards of reliability and validity, as well as representativeness of their standardization samples, and offer both parent and teacher forms. Thus, given the availability of measures with good to excellent psychometric properties that also cover several subdomains identified as important to assess, it seems worthwhile to rely on existing measures in fielding any large-scale assessment of the state of young children in the near future. In addition, the extant measures of social-emotional development documented in one inventory of measures (Federal Interagency Forum on Child & Family Statistics, 2015) already have some corpus of empirical research supporting their construct and/or predictive validity. For these reasons it seems unnecessary to develop yet another measure of children's social functioning. Rather time and effort could be more fruitfully spent on collecting high quality data on children's social and emotional competence during the preschool years and the transition to kindergarten. Ideally the assessment would be longitudinal, if feasible.

Implications for social policy

The assessment of emerging social and emotional competence, especially in 3 to 5 year olds, has profound implications for educational policy and for social policy more generally. First, the emphasis on strengths and child well-being is timely and a correction to the more usual emphasis on problem behaviors. Second, a wealth of data indicate that social skills with peers, emotion understanding, and the regulation of behavior develop in the context of supportive relationships with caring adults and that in the absence of such support young children are less able to cope with developmental challenges. One crucially important developmental challenge is the adjustment to child care and/or school settings, requiring young children to adapt to routines, establish new relationships with peers and caregivers or teachers, and meet expectations for academic learning and acceptable behavior. Thus, children are more likely to function well in school when they have a warm relationship with teachers and teachers understand children's developmental needs (NICHD ECCRN, 2003).

In addition, understanding how social-emotional competence and adjustment indicators co-vary in young children at specific ages and then how they change over time would appear to have important implications for educational policy and practice, especially in regard to how best to support competence building in children at risk. A growing body of literature suggests that peer competence and positive teacher-child relationships can help children overcome adjustment difficulties and foster motivation for learning in children living in poverty (Bulotsky-Shearer, Bell, et al., 2012; Bulotsky-Shearer, Manz, et al., 2012; McWayne, Fantuzzo, & McDermott, 2004; Nix et al., 2013). These recent findings argue for a broad-based assessment of emerging social and emotional competence in young children as one avenue for understanding the links among competence, problem behaviors, and

early academic achievement. Finally, data obtained across the subdomains identified in the review summarized in Halle and Darling-Churchill (2016-in this issue) may identify children in need of prevention or intervention, potentially based on profiles of functioning, as well as those who are able to adjust adequately as they transition to preschool or school.

The need to assess self-regulation both objectively and within context

Amanda P. Williford and Jessica Vick Whittaker, University of Virginia

Conceptualization of self-regulation

Self-regulation can be defined as one's own ability to manage emotions, control behaviors, and focus attention to cope effectively with environmental demands (Baumeister & Vohs, 2004; Blair & Razza, 2007; Calkins & Williford, 2009; Rimm-Kaufman, Curby, Grimm, Nathanson, & Brock, 2009). These skills play a vital role in preparing children for academic and social success later in life (Blair, 2002; McClelland, Morrison, & Holmes, 2000). During the preschool years, children dramatically increase their skills to manage emotions, comply with adult demands and directives, delay engagement in specific activities, engage in goal-directed behavior, and attend to activities. Young children's self-regulation skills are core competencies for developing social-emotional skills (CASEL, 2013). Children must manage their own emotions, behaviors, and thoughts in order to interact successfully and form positive relationships with peers and adults.

It is conceptually useful to separate self-regulation into the components of emotion, behavior, and cognition even though these processes are heavily intertwined and dependent upon each other (Calkins & Williford, 2009). Emotion regulation refers to skills and strategies that allow an individual to manage, inhibit, and enhance emotional arousal to promote adaptive behavior (Calkins, 1997). Children's emotion regulation strategies include both external aids (e.g., comforting gestures from another) and internal processes (e.g. taking deep breaths) to manage emotional stimulation (Thompson & Lagattuta, 2006). Behavior regulation refers to one's ability to monitor and manage his/her own behavior, including compliance to adult demands and directives, the ability to control impulsive responses, and the delay of engagement in specific activities (Kuczynski & Kochanska, 1995). Cognitive regulation includes the ability to focus and attend to an activity or task and often includes skills that fall within the domain of executive functioning. There is disagreement within and across the fields of psychology and education as to whether executive function skills are part of children's self-regulation or are a distinct set of skills that contribute to children's self-regulation. Halle and Darling-Churchill (2016-in this issue) provide a useful description of the theoretical distinctions between these constructs while also acknowledging the overlap. And so, we focus on the aspects of self-regulation that focus on management of emotion and behavior in order to be consistent with these authors.

Development of self-regulation measures

Being able to accurately assess young children's self-regulation skills is important for our understanding of how children develop school readiness skills as self-regulation has been found to be a significant mediator of children's academic and social-emotional outcomes (Raver et al., 2011). Additionally, accurate measurement of self-regulation is necessary to identify children who would benefit from prevention and early intervention efforts because they display low self-regulation skills. Measurement of self-regulation typically falls into three categories: 1) Adult report, 2) Direct assessment, and 3) Observation of children's behavior. Progress has been made in each of these areas as far as the

development of valid and reliable assessments, but there is still work to be done.

Halle and Darling-Churchill (2016-in this issue) have identified a variety of parent and teacher report measures of children's self-regulation skills. Parents and teachers spend more time with children than any other adult and so are valuable informants who have summative knowledge about children, and may have intuitive knowledge of what is normative. However, adult report measures have been criticized as lacking validity due in part to bias toward or against individual children and because significant variability in teachers' rating scale report of children's skills can be explained by teacher characteristics (e.g., Waterman, McDermott, Fantuzzo, & Gadsden, 2012).

Direct assessments are often viewed as the "gold standard" when assessing children's skills because they are seen as the most objective estimate of a child's skill. However, historically, these types of measures have been administered by trained data collectors in quiet laboratory settings. Even when direct assessments are administered in a school-based setting, typical procedures involve removing the child from the classroom context and administering assessments in a quiet space with an attentive data collector or teacher. For children who are at particular risk for early school failure due to socio-demographic risk, self-regulatory risk, or developmental delay, these assessments may not be accurate indicators of skills they are able to demonstrate within the actual classroom, particularly skills in the social-emotional domain. The early childhood classroom context (e.g., completing tasks independently, taking turns, stopping a desired activity such as playing with blocks to start a non-desired activity such as cleaning up) places demands on children that do not exist outside of the classroom or during a one-on-one assessment. These demands make it harder for some children to engage in self-regulation in the moment.

Natural observations have been identified as a preferred method to assess child behavior in the classroom because they provide information about behavior within the context where he/she is actually performing (Meisels, 1999). Given how self-regulation can help children access important learning opportunities or content (Raver et al., 2011), it is important to assess the self-regulation skills children demonstrate within the early learning context. However, again, these observations typically require a trained unbiased observer.

In the sections that follow, we comment further on the importance of measuring self-regulation, respond to the measures identified by Halle and Darling-Churchill (2016-in this issue) as showing promise, and provide suggestions for alternative measurement options, and future measurement work in the area of children's self-regulation. It is important to note that our lens for review and recommendation is based on the view that a measure of self-regulation must assess children's skills objectively as they express them within the context of interest (i.e., home or school) in order to be valid. Our research focuses on young children's school readiness, and therefore, when considering assessments of young children's self-regulation in the classroom, one needs to be able to measure a child's regulatory skills *as they are expressed in the classroom* – the context where skill demonstration matters for school success or failure.

Importance of self-regulation skills

Recent attention has been given to the importance and development of self-regulation (sometimes referred to as character strengths, grit, motivation, and persistence). State Early Learning Guidelines (Daily, Burkhauser, & Halle, 2010) and the Head Start Child Development and Early Learning Framework (U.S. DHHS, 2011) now include goals for the social-emotional domain, including self-regulation, on an equal footing with other domains. There is good reason for this shift. Self-regulation has been conceptualized as a critical feature of school readiness – one that underlies the skills of successful school achievement (Blair, 2002). For instance, children's self-regulation skills predict performance of early math and reading as well social-emotional skills

(Blair, Peters, & Granger, 2004; Blair & Razza, 2007). And, self-regulation has a stronger influence on a student's academic performance than his or her IQ (Duckworth & Seligman, 2005). But the benefits of early self-regulation skills are not limited to early childhood. Longitudinal research that followed pre-k children into adulthood demonstrates that children's display of inhibitory control during pre-k is positively linked with a host of important adolescent and adult outcomes including better coping in adolescence, higher SAT scores, higher levels of education, and less substance abuse as adults (Mischel et al., 2011). And, there has been recent theorizing that high quality pre-k that focuses on developing children's executive function, self-regulatory and social-emotional skills, in addition to academic skills, is responsible for positive outcomes into adulthood (Heckman, Moon, Pinto, Savelyev, & Yavitz, 2010; Moffitt et al., 2011).

Criteria for evaluating measures of self-regulation in young children

Assessment can be used for many different purposes, and defining that purpose is critical in guiding assessment decisions (NRC, 2008). As presented by Halle and Darling-Churchill (2016-in this issue), the review team evaluated measures based on ten criteria, prioritizing reliability and validity while weighting the other criteria equally. However, it is important to keep in mind that depending upon the purpose of assessment, it would likely be appropriate to give differential weight to various criteria.

For researchers, and the field at large, an important criteria not included in the current review is whether measures are sensitive to change – do we see impacts of classroom-based interventions theorized to be associated with changes in children's self-regulation skills? For practitioners, the usefulness of any self-regulation assessment lies in its ability to tie data about a child to the use of an effective strategy or scaffold that will allow that child to be a more successful learner in the classroom. A related important criteria is whether any of the measures are accompanied by information for teachers about how to interpret children's scores so that they can identify the learning needs of their students and provide appropriate scaffolds for their skill development.

The measures selected in the inventory of available measures (Halle & Darling Churchill, in this issue) are appropriate choices for assessing self-regulation. However, most of the identified measures have large overlaps across many of the other social-emotional sub-domains and may not measure the sub-domains distinctly enough. As outlined throughout this issue, as well as noted above, there is overlap in the social-emotional sub-domains. Indeed, it is often difficult to tease apart which skills fall under which sub-domain. This is particularly true in the sub-domain of self-regulation where there is considerable diversity and debate among experts in the field about how to define self-regulation due to gaps in our understanding of how emotional, behavioral, and cognitive aspects of self-regulation are inter-related or distinct (see above). However, we must be precise in our definitions of the skills that we are trying to assess in order to accurately measure those skills, understand their relationship with other domains, and foster the development of them. One of the criteria upon which measures were evaluated was whether a measure "covers 2 or more sub-domains". It is only useful to use a measure that assesses multiple sub-domains, if the measures are precise, reliable, and valid at the sub-domain level (Raver, 2012).

Review of measures of self-regulation

Additional measures to consider

We note that the vast majority of measures selected were teacher/parent report measures (Two Bags Task is an exception). We also strongly recommend the inclusion of direct assessments to assess children's self-regulation skills. We know of two such measures that are

good candidates. The first is Cybele Raver's Preschool Self-Regulation Assessment (PSRA; Smith-Donald, Raver, Hayes, & Richardson, 2007). The PSRA includes a series of tasks that were adapted from developmental lab-based tasks in order to be administered within the early learning context. The tasks assess emotion, behavior, and cognitive aspects of self-regulation and overlap with the executive function area of development, as defined by Halle and Darling-Churchill (2016-in this issue). As part of this battery, Raver and colleagues developed an assessor report that allows a person conducting an assessment to provide a global rating of the child's emotion and behavioral regulation displayed during the direct assessment. This assessment demonstrates good reliability and validity (Smith-Donald et al., 2007) and has been used by early education researchers (e.g. Bierman, Nix, Greenberg, Blair, & Domitrovich, 2008; Williford, Whittaker, Vitiello & Downer, 2013). Second, Megan McClelland and colleagues developed the Head Toes Knees Shoulders Task (HTKS; Cameron Ponitz et al., 2008) which is a game-like direct assessment of children's behavioral self-regulation, including inhibitory control, working memory and attention focusing. The HTKS also has good psychometric properties (Cameron Ponitz, McClelland, Matthews, & Morrison, 2009; Cameron Ponitz et al., 2008; McClelland et al., 2014). The research using these measures indicates that data collectors can be reliably trained to administer these assessments to young children within the preschool and childcare context.

Gaps in measures and current developments

Although research-based measures of self-regulation exist (adult report and direct assessments), most have not been adapted for use in schools or at-scale, and norming data often are not available. Both of these limitations contribute to the dearth of measures available to practitioners. In addition, we know of no naturalistic observation measures that are designed to specifically assess children's display of self-regulation. However, there are existing measures that could be used as models to develop valid, reliable, and contextually based observations of children's self-regulation. For example the Individualized Classroom Assessment Scoring System (*inCLASS*; Downer et al., 2010) is an observational assessment of children's classroom engagement in interactions with teachers, peers, and tasks, comprised of 10 dimensions. Children are observed for 10-min cycles (4–8 cycles are recommended) and then each dimension is rated on a seven-point scale (guided by detailed descriptors of behaviors that indicate low, medium, and high quality). This measure demonstrates adequate reliability and validity (Downer et al., 2010; Williford, Maier, Downer, Pianta, & Howes, 2013; Williford, Whittaker et al., 2013). A similar measure assessing children's expressed self-regulation as it occurs within the classroom context would be useful for both researchers and practitioners because direct assessments administered outside of the classroom may overestimate a child's regulatory capacity in the classroom context.

Implications for the field

As stated throughout this response memo, it is critically important to consider context when understanding a young child's self-regulation. The field is fond of stating how crucial it is to understand a child's behavior and capacities in context; however, we rarely integrate this contextual approach into our assessments, thus rendering them less helpful in practical applications. Recent evidence from the field of cognitive development demonstrates the importance that context plays in children's self-control/impulse control skills. In a laboratory-based, experimental study, researchers found that children's self-control skills varied widely depending upon whether they had just previously interacted with an adult that was dependable versus undependable (Kidd, Palmeri, & Aslin, 2013). Specifically, researchers varied the context in which they conducted a classic experiment that has been used to assess children's impulse control – the marshmallow test. When preschool children

interacted with an adult who they could trust and was dependable, they waited on average 12 min before eating a marshmallow compared to children who interacted with an unfamiliar data collector, who waited on average only 3 min. The results from this study suggest that children's display of critical school readiness skills such as self-control may not be stable and instead, might depend largely upon the context or environment where children are expected to perform the skills; this includes whether the child is interacting with a familiar adult. Additional aspects of context that might be important to consider are: the size of the class, environmental features of the room, adult-to-child ratio, schedule, teacher expectations, peer influences, structure of the activity, engagement level of the activity, and noise level. We need to include measures of these contextual factors if we are going to leverage assessment data to influence a teacher's practice for either a specific child or a classroom of children. In sum, there is a need for an accurate measure of self-regulation that is situationally rather than dispositionally based.

We also need to develop measures of self-regulation that are comparable over time; that recognize the changing nature of children's self-regulation competence and also capture developmental milestones and detect individual differences among children. We also must do more than "control for" age and ethnic group in the measurement of self-regulation. Researchers measure and control for these in their analyses, but preschool teachers have classrooms of children who are from diverse cultural backgrounds and who range in age from 3 to 5 years. For example, teachers may struggle more with the younger children (even if they have a cognitive understanding of the age-related differences in self-regulation) because, in the moment, they expect the same self-regulation behaviors of all children in the classroom. This may increase the likelihood that children from certain groups (e.g., younger children) will be at a persistent disadvantage. Also, the question of construct equivalency across sociolinguistic and sociocultural groups needs to be taken into account in the design of new measures. Culturally-specific socialization practices may lead to different "norms" of self-regulatory behaviors among bi-cultural or bilingual groups (see Halle et al., 2014).

Implications for social policy

As highlighted in the research presented above, children's early self-regulation skills are critical to their later social-emotional well-being and academic achievement (Blair & Razza, 2007). We see several benefits to the identification and further development of valid and reliable measures of children's self-regulation, particularly, as expressed within the classroom context. First, there is potential to develop classroom-based assessments of children's self-regulation that could be directly and specifically tied to what a teacher does in the classroom. Assessments could be used to: 1) Demonstrate that self-regulation skills are critical to other aspects of a child's learning and development (e.g., if a teacher maximizes a child's self-regulatory capacity during circle time the child can learn concepts of print during a book reading); 2) Scaffold a child so that he/she is better able to self-regulate during challenging situations (e.g., by altering the learning environment; implementing a visual cue); and, 3) Improve a child's regulation skills through direct training (e.g., playing games like Simon Says, Mother May I and Red Light/Green Light). Effective strategies and curricula exist to improve children's self-regulation and it would be ideal if data collected through self-regulation assessments increased a teacher's uptake of those strategies.

An assessment of children's self-regulation in the classroom also presents an opportunity to raise awareness of teachers' own behaviors; encourage teachers to view children's behaviors as discrete indicators of the child-context fit (state view); and eliminate a trait-based, "all or nothing" view of particular children. Teachers are often told that they should use assessments to plan instruction and individualize learning opportunities but they are rarely provided any help on how to do this.

This often results in teachers spending a lot of time generating assessment data because it is required, but not using the results of the data to inform their practice. Thus, the development of classroom-expressed, objective measures of children's self-regulation has the potential to be used to help teachers understand the developmental progression of children's self-regulatory skills, and to guide efforts to improve supports for children's development of these skills.

Finally, teachers need data that they can use to create a more productive learning environment for a child. If you ask a teacher which children have trouble regulating their behavior and emotions in the classroom they can tell you who they are – they do not need a test for that. The usefulness of any self-regulation assessment lies in its ability to tie data about a child to the use of an effective strategy or scaffold that will allow that child to be a more successful learner in the classroom. For example, if a self-regulation assessment indicated to a teacher how much longer that child would be able to sustain his or her attention or resist impulsive behavior by using a particular strategy (e.g., a visual cue), that teacher may be more likely to change what she is doing in the classroom and to use that strategy to improve the child's learning. Or, if the data helped the teacher pinpoint specific obstacles to a child's learning that occur during the day, linked to a suggestion she could use to adapt her lessons or how she delivers them, she might be likely to follow that suggestion and try to remove the barriers to that child's learning.

Improving the assessment of children's self-regulation

This special issue is a very good starting point for researchers and policy makers as they think about assessing children's self-regulation as a critical developmental skill needed for success in school and life. It provides useful information about how both federal and non-federal researchers should consider measuring children's social-emotional skills when designing large scale studies. However, many of these measures are not what researchers in the sub-domain of self-regulation would recommend using because they are not precise enough measures of self-regulation and because of the presence of rater bias when using teacher or parent report. Thus, this review allows us to see where the gaps in measurement of self-regulation still exist and should serve as a call to researchers in the fields of early education and developmental psychology to develop objective measures (particularly direct assessments and naturalistic observations) that can be used at scale *and* help inform parent and teacher practice.

In addition, information in this issue should serve to encourage both researchers and federal agencies to think carefully about the design of large scale studies. In particular, we should think carefully when deciding between breadth and depth. For example, there is currently a lack of understanding as to how children's engagement in preschool operates in tandem to their self-regulation and executive functioning to build academic and social competence and how self-regulation can be influenced by the quality of the opportunities provided through the child's proximal environment (i.e. home or school). This gap in our understanding of how self-regulation operates to support or constrain children's school readiness skills and how the quality of children's experiences moderates this link can only be closed by conducting observational and experimental studies that include depth and precision in measurement.

Measurement of the executive function domain

Michael T. Willoughby, RTI International

Definition and controversies in the measurement of executive functions

Executive Functions (EF) refer to a constellation of cognitive processes that, in broad terms, facilitate goal-directed problem solving

pursuits. However, as was noted by Barkley (2012), this working definition is nebulous and a consensus definition has proven elusive. In much of the current literature, EF has been defined more narrowly as consisting of three partially dissociable processes – inhibitory control (IC), working memory (WM), and cognitive flexibility/attention shifting (AS), which provide a foundation for problem solving and future-oriented goal directed behaviors (Miyake & Friedman, 2012). Prior to providing a general commentary on Halle and Darling-Churchill's target article, I consider two issues related to the measurement of EF in general.

Despite theoretical interest in IC, WM and AS as dissociable constructs in early childhood (Garon, Bryson, & Smith, 2008), the factor analytic literature is mixed. Whereas some studies have indicated that EF represents an undifferentiated set of abilities in early childhood (Hughes, Ensor, Wilson, & Graham, 2010; Wiebe, Espy, & Charak, 2008; Wiebe et al., 2011; Willoughby, Blair, Wirth, Greenberg, & Investigators, 2010, 2012), the more recent literature has indicated that IC and WM may be dissociable constructs during this age period (Gandolfi, Viterbori, Traverso, & Usai, 2014; Miller, Muller, Giesbrecht, Carpendale, & Kerns, 2013; Schoemaker et al., 2012; Usai, Viterbori, Traverso, & De Franchis, 2014). There is also evidence that EF may begin as an undifferentiated set of abilities in early childhood and subsequently "fractionate" into more specific dimensions of IC, WM, and AS; however, whether and at what age this occurs is unclear, with evidence spanning early and middle childhood, as well as adolescence (Brydges, Reid, Fox, & Anderson, 2012; Lee, Bull, & Ho, 2013; Lerner & Lonigan, 2014; Shing, Lindenberger, Diamond, Li, & Davidson, 2010; van der Ven, Kroesbergen, Boom, & Leseman, 2013; Xu et al., 2013). The important point is that although it is reasonable to subdivide the broader construct of EF into specific sub-dimensions of inhibitory control, working memory, attention shifting among adults, it is not yet clear whether and when these distinctions become meaningful in childhood samples.

In addition to uncertainties regarding the factor structure of EF across childhood, a more pragmatic question is how best to resolve discrepancies in the assessment of EF using performance-based and questionnaire methods. Children's EFs were initially assessed using performance-based measures that represented downward extensions of tasks that had been used with adults (Passler, Isaac, & Hynd, 1985; Welsh, Pennington, & Groisser, 1991). These measures were subsequently replaced by tasks that were developed specifically for use with young children (e.g., Espy, 1997; Espy, Kaufmann, Glisky, & McDiarmid, 2001; Espy, Kaufmann, McDiarmid, & Glisky, 1999), and efforts to develop performance-based measures of EF that are optimized for use with young children continues to be an active area of research (Carlson, Faja, & Beck, 2015; Garon, Smith, & Bryson, 2014; Ponitz et al., 2008; Weintraub et al., 2013; Willoughby et al., 2010).

Historically, performance-based measures of EF were used for purposes of diagnostic decision making. However, as scientific interest in the construct of EF has grown, there has been increased interest in using EF assessments for purposes of treatment planning and evaluation, as well as in the prediction of functional outcomes more generally (Isquith, Roth, Kenworthy & Gioia, 2014; Silver, 2014). Whereas performance based indicators of EF have long been considered the gold standard method for diagnostic decision making, they have been criticized for the prediction of functional outcomes due to poor ecological validity (Barkley & Murphy, 2010, 2011; Goldberg & Podell, 2000; Sbordone, 2001, 2014; Sbordone & Purisch, 1996). Critics have argued that the highly structured nature of performance-based tasks, including their use in controlled settings and their narrow focus on discrete abilities, do not resemble the "real-world" settings in which an integrated set of EF skills are drawn upon to facilitate contextually relevant problem solving or decision making.

The perceived inadequacies of performance-based measures of EF were a major impetus for the creation of questionnaire-based assessments of EF (Gioia, Espy, & Isquith, 2003; Mooney, Walmsley, & McFarland, 2006; Thorell & Nyberg, 2008). Questionnaire assessments

of EF attended to two key limitations of performance based measures (Isquith, Roth, & Gioia, 2013). First, they represent structured observations of EF behaviors in contextually relevant settings. Second, they represent an assessment of EF behaviors over a much broader window of time (e.g., weeks for questionnaires versus minutes for performance based measures). Despite these strengths, questionnaires assume that individual differences in observed behaviors result from individual differences in underlying cognitive processes of EF. However, there is limited empirical evidence to support this assumption. For example, in a review of 20 studies, caregiver (parent or teacher) ratings correlated poorly with direct assessments of EF, with median $rs = .14–.25$ (Toplak, West, & Stanovich, 2013). Correlations of this magnitude indicate that the two methods only share 1–6% of their variation.

This lack of agreement has led to suggestions that performance-based tasks and behavioral questionnaires measure "different aspects" of the construct of EF, with each making independent contributions to individual functioning (Isquith et al., 2013; McAuley, Chen, Goos, Schachar, & Crosbie, 2010; Toplak et al., 2013). An alternative conclusion has been that whereas performance-based indicators may represent individual differences in underlying EF ability, questionnaires may represent the realization of those abilities in context. Although heuristically useful, the field continues to struggle with practical ways to integrate information across these methods (Silver, 2014). This highlights an essential problem with relying on either method, in isolation, as an indicator of the construct of EF, which is precisely what Halle and Darling-Churchill (2016-in this issue) proposed to do.

In sum, despite widespread interest in the construct of EF, at least two fundamental problems in the conceptualization of this construct exist. First, questions about the dimensionality (structure) of EF abilities remain, including when in development these abilities become differentiated. Second, performance based assessments and behaviorally based questionnaires of EF are weakly associated and appear to measure different phenomena. Clearly, the information that is derived from these different assessment methods should not be considered interchangeable.

General commentary

In the second article of this issue, Halle and Darling-Churchill summarize the characteristics of existing measures of social and emotional development and EF that may be appropriate for use in future federal surveys. Given the diversity of the extant field of measures, the lack of agreement among researchers on how to define the constructs which contribute to social and emotional competencies, and the unique needs of federal agencies, this represented a substantial undertaking. While I was invited to comment specifically on their recommendations regarding the measurement of EF, initially, I raise a few concerns about the broader effort.

Defining the boundaries of social and emotional development

Halle and Darling-Churchill's review (2016-in this issue) relied on the Center on the Social Emotional Foundations for Early Learning's definition of social and emotional development which was characterized as

> the developing capacity of the child from birth through five years of age to form close and secure adult and peer relationships; experience, regulate, and express emotions in socially and culturally appropriate ways; and explore the environment and learn – all in the context of family, community, and culture (Center on the Social Emotional Foundations for Early Learning, 2008).

In an effort to underscore the importance of this construct, the authors highlighted a number of federal agencies (Office of Head Start, Office of Special Education, Administration for Children and Families) that have prioritized social and emotional development as an important

milestone of early childhood. Finally, given the breadth of social and emotional development, the authors relied on their review of 15 overlapping resources in order to define the specific subdomains of social and emotional development that would be the focus of their report (see Halle & Darling-Churchill, 2016-in this issue).

My first concern is whether it is appropriate to highlight EF as uniquely relevant to social and emotional development. This is not to suggest that EF is unrelated to social and emotional development. Rather, it is to question whether we should privilege EF over other aspects of child functioning that may play an equally or more important role in social and emotional development (e.g., children's expressive language development).

A related concern is whether it is appropriate to treat EF and self-regulation as distinct aspects of social and emotional development. The subdomains determined through the literature review performed by Halle and Darling-Churchill (2016-in this issue), used a variety of different terms (self-regulation, cognitive regulation, emotion regulation, effortful control, self-management) to describe a conceptually similar set of behaviors. The problems associated with the proliferation of terms for a similar set of behaviors has been summarized elsewhere (Morrison & Grammer, in press; Zhou, Chen, & Main, 2012). Although the review team made an attempt to distinguish EF from self-regulation on conceptual grounds, I was unable to comprehend their distinction and did not find it to be compelling.

Blair and Ursache (2011; Ursache, Blair, & Raver, 2012) proposed a bidirectional model to describe the association between self-regulation and EF. In their model, self-regulation was contingent on but also broader in scope than EF. An essential idea was that self-regulation involves "the management of attention, emotion, and stress response systems in ways that facilitate the use of EF abilities in the service of goal-directed actions" (Ursache, Blair, & Raver, 2012; p. 123). In Blair and Ursache's model, self-regulation involves the integration of cognitive and emotional resources that are necessary for goal-directed behaviors. It is not clear that the exclusive use of questionnaire methods is sufficient for preserving this distinction. This conceptual confusion represents the current status of the field and derives, in part, from multidisciplinary efforts to measure (and label) individual differences in domain general (including cognitive, motivational, and temperamental) processes that are related to functionally relevant outcomes.

Strategy for comparison of measures

The objective of Halle and Darling-Churchill's review (2016-in this issue) was to identify and critically evaluate existing measures (separately for each subdomain of social and emotional development) that had the strongest potential to be included in future federal studies or surveys. This represented an enormously difficult task in part because of the large number of candidate measures but also due to the lack of uniform information across measures to facilitate direct comparisons. Here, I raise some concerns with the overall strategy that was used for measure evaluation.

The authors identified 10 criteria upon which each measure would be evaluated: reliability, validity, size/diversity of sample, availability in language other than English, requirements for training, availability of parent and teacher forms, domain coverage (i.e., does an instrument include two or more subdomains of social and emotional development), administration time, cost, and age coverage covers wide age range. Each criterion was evaluated on a three-point Likert rating scale (strong, moderate, weak) based on standards that were abstracted from previous studies. Finally, a "box counting" approach was used to assign a score to each measure that indicated the number of criteria that were rated as "moderate" or "strong".

Many of the decision rules that were used to inform the Likert rating (i.e., what degree of evidence was necessary to rate measures as 'strong', 'moderate', 'weak') were necessarily subjective (e.g., N = 250 is not large but N = 350 is; test-retest reliability ≥ .70 was acceptable

but < .70 was not). This decision posed the risk of implicitly prioritizing the identification of questionnaires that have been in wide use for a long time. Relative to performance-based assessments, questionnaires are quicker to administer and do not require training. Questionnaires that have been in wide use for long periods are more likely to have parent and teacher forms available, and to have been used in large samples. Questionnaires that are used in large samples have greater opportunities to be utilized by multiple research groups (not just questionnaire developers), thereby increasing the chance that at least some reliability or validity had been reported. Given the large costs of federal surveys, it may be judicious to prioritize these pragmatic issues. However, this approach does not necessarily identify the "best" measures of a given construct. By implicitly prioritizing measures that have been widely used, this has the potential to constrain the ability of researchers to ask innovative questions using large scale datasets.

Two of the rated criteria deserve special comment. With respect to the sample size, a measure was deemed "strong/easy to use" if it had been used in a large (defined as N > 300) sample with racial/ethnic and socio-economic diversity. With respect to language, a measure was deemed "strong/easy to use" if it was available in a language other than English. While these criteria attend to practical issues, the important issue is not that a questionnaire has been used in a large sample or has been translated into a different language. An arguably more important question is whether a questionnaire has been demonstrated to exhibit equivalent psychometric properties across racial/ethnic, socioeconomically, or language subgroups. In the psychometric literature, this is known as tests of measurement invariance (e.g., Reise, Widaman, & Pugh, 1993).

To the extent that there is interest in measuring change in social and emotional development constructs across time, a related question is whether a given measure exhibits equivalent psychometric properties across time for racial, ethnic, or socioeconomically defined subgroups. This question is evaluated using tests of longitudinal invariance (e.g., Widaman, Ferrer, & Conger, 2010). Many of the measures that were rated favorably with respect to sample size and availability in other languages would not have been rated as such had this criterion been reframed as measurement invariance. The point is not to be unnecessarily critical of the decision rules that the measures review team adopted but rather to underscore how those decision rules impact one's evaluation of the "best" measures for inclusion in future federal surveys and other research efforts.

Specific commentary on recommendations for measures of executive function

In Halle and Darling-Churchill (2016-in this issue), the Preschool Learning Behavior Scale [PLBS] and the Behavioral Assessment Systems for Children, Second edition [BASC-2] were identified as possible measures of EF. Notably, the BASC-2 is appropriate for use with children in both of the age ranges (0–3, 4–5 years). I have two criticisms of these recommendation.

Conceptual clutter

In a commentary on the status of EF in early childhood, Morrison and Grammer (in press) recently commented

> The fuzzy distinction between EF and self-regulation (as well as effortful control) – in combination with existing debates about the measurement of both constructs – has contributed greatly to a proliferation of constructs, resulting in our view in kind of a 'conceptual clutter' whose similarities and differences are not readily apparent.

Their appraisal of the literature is widely held. It is with this perspective in mind that I raise concerns about the distinction between EF and self-regulation. In order to underscore that point, it is noteworthy that

the questionnaires that were recommended for self-regulation *both* subsume those that were recommended for EF (i.e., the BASC-2 and the PLBS were both identified as candidates for use in future federal surveys in the domains of self-regulation *and* EF). The high degree of overlap in measurement decisions for self-regulation and EF raises questions about the distinctiveness of these subdomains. This may be particularly true when questionnaires are the exclusive assessment strategy available to make this distinction.

Muddled definitions of executive function

The BASC-2 is currently marketed as a behavioral and emotional screening system with coverage of internalizing and externalizing problems, school problems, and adaptive skills. The original frontal lobe/executive control scale was developed on the basis of expert ratings on items on the BASC that were believed to reflect executive function (Barringer & Reynolds, 1995). Caregiver ratings of the BASC-2 executive function scale correlate strongly ($rs = .63$–$.83$) with their ratings of attention deficit/hyperactivity disorder (ADHD) symptoms (Sullivan & Riccio, 2006). To the extent that EF scales can be constructed from a broad set of behavioral and emotional screening items, which were not initially developed to measure EF, this raises questions about the distinction between EF and other dimensions of social and emotional development (e.g., behavior problems).

The recommended measures for executive function exclude the Behavior Rating Inventory of Executive Function-Preschool version (BRIEF-P; Gioia et al., 2003) and the Childhood Executive Functioning Inventory (CHEXI; Thorell & Nyberg, 2008) which were explicitly developed to measure behaviors characteristic of EF. While these measures were included in the longer inventory of investigated measures, they failed to meet the threshold for inclusion in the review by Halle and Darling-Churchill (2016-in this issue) due to moderate ratings of their psychometric properties. Here I focus on the BRIEF-P because it spans the entire age period of interest and because it has been in use for a much longer period of time than the CHEXI. Despite the strengths of the BRIEF (see Isquith et al., 2013), the scores on this questionnaire appear to be more strongly correlated with ADHD behaviors than they are with performance-based measures of EF (Mahone & Hoffman, 2007; McAuley et al., 2010; Toplak, Bucciarelli, Jain, & Tannock, 2009). This suggests that the BRIEF-P may be better conceptualized as a general indicator of behavioral functioning than of EF. Even the scale developers continue to emphasize a model of assessment that involves the combined assessment of EF using questionnaire and performance-based indicators (Isquith et al., 2013).

Contributions of this research effort

The coordinators of this project have undertaken an important but difficult task. My critical evaluation of their recommendations regarding EF should not be interpreted as criticism of the broader importance of their work. The development of surveillance and reporting systems that focus on social and emotional development in early childhood represents an important endeavor, and the authors' efforts will serve a useful function for federal partners. The work presented in Halle and Darling-Churchill (2016-in this issue) is distinguished from conventional research-oriented papers by its strong consideration of the practical needs of federal surveys. Indeed, many of the criticisms raised here reflect the tensions that exist between the competing demands of selecting measures for research purposes versus meeting the practical constraints of the federal reporting systems.

At a broader level, Halle and Darling-Churchill (2016-in this issue) demonstrate one approach for the systematic review of measures. Whereas systematic methods exist for research synthesis (i.e., meta-analysis), similar methods for measure evaluation are not widely

adopted. Future reviews of measures in the early childhood or developmental fields may benefit from progress being made in other disciplines. For example, in the area of health research (specifically health-related, patient-reported outcomes), the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) initiative represents an impressive effort that includes a standard taxonomy of criteria on which measures can be evaluated and a standardized instrument for measure evaluation and comparison (see http://www.cosmin.nl/).

Conclusion

In light of the time demands and costs that are associated with federal and other survey data collection efforts, questionnaires have been prioritized in the assessment of social and emotional development. However, questionnaire data are not sufficient for unambiguously representing individual differences in EF abilities in young children. Instead, a combination of questionnaire and performance-based indicators are both necessary to meet this objective, despite the fact that the field continues to struggle with how best to integrate across these two sources of data that are often in poor agreement. Given this state of affairs, I would discourage the sole reliance on any of the questionnaires for purposes of measuring EF in future federal surveys or other large scale research efforts that focus on social and emotional development.

Acknowledgments

References

Achenbach, T. M., & Edelbrock, C. (1978). Classification of child psychopathology: A review and analysis of empirical efforts. *Psychological Bulletin, 85,* 1275–1301.
Achenbach, T. M., & Edelbrock, C. (1981). Behavioral problems and competencies reported by parents of normal and disturbed children aged four through sixteen. *Monographs of the Society for Research in Child Development, 46* (Serial No. 188).

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

Ainsworth, M., Blehar, M., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the strange situation.* Hillsdale, NJ: Erlbaum.

Andreassen, C., & Fletcher, P. (2007). *Early Childhood Longitudinal Study, Birth Cohort (ECLS–B) Psychometric Report for the 2-Year Data Collection (NCES 2007–084).* Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Arnold, D. H., Kupersmidt, J. B., Voegler-Lee, M. E., & Marshall, N. A. (2012). The association between preschool children's social functioning and their emergent academic skills. *Early Childhood Research Quarterly, 27,* 376–386.

Barkley, R. A. (2012). *Executive functions: What they are, how they work, and why they evolved.* New York: Guilford Press.

Barkley, R. A., & Murphy, K. R. (2010). Impairment in occupational functioning and adult ADHD: The predictive utility of executive function (EF) ratings versus EF tests. *Archives of Clinical Neuropsychology, 25*(3), 157–173. http://dx.doi.org/10.1093/arclin/acq014.

Barkley, R. A., & Murphy, K. R. (2011). The nature of executive function (EF) deficits in daily life activities in adults with ADHD and their relationship to performance on EF tests. *Journal of Psychopathology and Behavioral Assessment, 33*(2), 137–158. http://dx.doi.org/10.1007/s10862-011-9217-x.

Barringer, M. S., & Reynolds, C. R. (1995). Behavior ratings of frontal lobe dysfunction. *Paper presented at the National Academy of Neuropsychology.*

Bates, J. E., & Pettit, G. S. (2007). Temperament, parenting, and socialization. In J. E. Grusec, & P. D. Hastings (Eds.), *Handbook of socialization: Theory and research* (pp. 153–177). New York, NY: Guilford Press.

Baumeister, R. F., & Vohs, K. D. (2004). *Handbook of self-regulation: Research, theory, and applications.* New York: Guilford Press.

Belsky, J., Hsieh, K., & Crnic, K. (1998). Mothering, fathering, and infant negativity as antecedents of boys' externalizing problems and inhibition at age 3 years: differential susceptibility to rearing experience? *Development and Psychopathology, 10,* 301–319.

Belsky, J., & Pleuss, M. (2009). The nature (and nurture?) of plasticity in early human development. *Perspectives on Psychological Science, 4,* 345–351.

Belsky, J., Woodworth, S., & Crnic, K. (1996). Trouble in the second year: Three questions about family interaction. *Child Development, 67,* 556–578.

Bierman, K. L. (2011). The promise and potential of studying the "invisible hand" of teacher influence on peer relations and student outcomes: A commentary. *Journal of Applied Developmental Psychology, 32*(5), 297–303. http://dx.doi.org/10.1016/j.appdev.2011.04.004.

Bierman, K. L., Nix, R. L., Greenberg, M. T., Blair, C., & Domitrovich, C. E. (2008). Executive functions and school readiness intervention: Impact, moderation, and mediation in the Head Start REDI Program. *Development and Psychopathology, 20,* 821–843.

Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of child functioning at school entry. *American Psychologist, 57,* 111–127.

Blair, C., Peters, R., & Granger, D. (2004). Physiological and neuropsychological correlates of approach/withdrawal tendencies in preschool: further examination of the behavioral inhibition system/behavioral activation system scales for young children. *Developmental Psychobiology, 45*(3), 113–124.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78,* 647–663.

Blair, C. B., & Ursache, A. (2011). A bidirectional model of executive functions and self-regulation. In K. D. Vohs, & R. F. Baumeister (Eds.), *Handbook of self-regulation* (pp. 300–320) (2nd ed.). New York: Guilford Press.

Bornstein, M. H., Hahn, C. S., & Haynes, O. M. (2010). Social competence, externalizing, and internalizing behavioral adjustment from early childhood through early adolescence: developmental cascades. *Development and Psychopathology, 22*(4), 717–735. http://dx.doi.org/10.1017/S0954579410000416.

Bowlby, J. S. (1969). *Attachment and loss. Attachment, Vol. I,* New York: Basic Books.

Boyce, W. T., & Ellis, B. J. (2005). Biological sensitivity to context: I. An evolutionary-developmental theory of the origins and functions of stress reactivity. *Development and Psychopathology, 17,* 271–301.

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

Briggs-Gowan, M. J., & Carter, A. S. (2002). *Brief Infant-Toddler Social and Emotional Assessment (BITSEA) manual, version 2.0.* New Haven, CT: Yale University.

Briggs-Gowan, M. J., & Carter, A. S. (1998). Preliminary acceptability and psychometrics of the infant–toddler social and emotional assessment (ITSEA): A new adult-report questionnaire. *Infant Mental Health Journal, 19,* 422–445.

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist, 32,* 513–531.

Brydges, C. R., Reid, C. L., Fox, A. M., & Anderson, M. (2012). A unitary executive function predicts intelligence in children. *Intelligence, 40*(5), 458–469. http://dx.doi.org/10.1016/j.intell.2012.05.006.

Bulotsky-Shearer, R. J., Bell, E. R., Romero, S. L., & Carter, T. M. (2012). Preschool interactive peer play mediates problem behavior and learning for low-income children. *Journal of Applied Developmental Psychology, 33,* 53–65.

Bulotsky-Shearer, R. J., Manz, P. H., Mendez, J. L., McWayne, C. M., Sekino, Y., & Fantuzzo, J. W. (2012). *Child Development Perspectives, 6,* 225–231.

Calkins, S. D. (1997). Cardiac vagal tone indices of temperamental reactivity and behavioral regulation in young children. *Developmental Psychobiology, 31,* 125–135.

Calkins, S. D., & Williford, A. P. (2009). Taming the terrible twos: Self-regulation and school readiness. In O. A. Barbarin, & B. H. Wasik (Eds.), *Handbook of child development and early education: Research to practice* (pp. 172–198). New York, NY: Guilford Press.

Cameron Ponitz, C. E., McClelland, M. M., Jewkes, A. M., Connor, C. M., Farris, C. L., & Morrison, F. J. (2008). Touch your toes! Developing a direct measure of behavioral regulation in early childhood. *Early Childhood Research Quarterly, 23*(2), 141–158.

Cameron Ponitz, C., McClelland, M. M., Matthews, J. S., & Morrison, F. J. (2009). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. *Developmental Psychology, 45*(3), 605.

Campbell, S. B. (2002). *Behavior problems in preschool children: Clinical and developmental issues* (2nd ed.). New York: Guilford Press.

Campbell, S. B., Pierce, E. W., Moore, G., Marakovitz, S., & Newby, K. (1996). Boys' externalizing problems in early elementary school: Pathways from early behavior problems, maternal control, and family stress. *Development and Psychopathology, 8,* 701–720.

Campbell, S. B., Shaw, D. S., & Gilliom, M. (2000). Early externalizing behavior problems: Toddlers and preschoolers at risk for later maladjustment. *Development and Psychopathology, 12,* 467–488.

Carlson, S. M., Faja, S., & Beck, D. M. (2015). Incorporating early development into measurement approaches: The need for a continuum of measures across development. In J. A. Griffin, L. S. Freund, & P. McCardle (Eds.), *Executive function in preschool age children: Integrating measurement, neurodevelopment, and translational research.* Washington, DC: American Psychological Association.

Carter, A., & Briggs-Gowan, M. J. (2006). *Infant toddler social emotional assessment.* San Antonio, TX: Pearson.

Carter, A. S., Briggs-Gowan, M. J., Jones, S. M., & Little, T. D. (2003). The infant–toddler social and emotional assessment (ITSEA): Factor structure, reliability, and validity. *Journal of Abnormal Child Psychology, 31,* 495–514.

Cicchetti, D. (1990). Perspectives on the interface between normal and atypical development. *Development and Psychopathology, 2,* 329–334.

Cicchetti, D., & Cohen, D. J. (1995). Perspectives on developmental psychopathology. In D. Cicchetti, & D. J. Cohen (Eds.), *Developmental psychopathology. Theory and methods, I.* (pp. 3–20). New York: Wiley.

Collaborative for Academic, Social, and Emotional Learning (2013). *2013 CASEL guide: Effective Social and Emotional Learning Programs Preschool and Elementary School Edition.* Chicago, IL: Collaborative for Academic, Social, and Emotional Learning.

Committee for Children (2011). Second Step: Social-emotional skills for early learning. *Early Learning: Review of Research,* 1–15 (Retrieved from http://www.cfchildren.org/Portals/0/SS_EL/EL_DOC/EL_Review_Research_SS.pdf).

Crane, J., Mincic, M. S., & Winsler, A. (2011). Parent-teacher agreement and reliability on the Devereux Early Childhood Assessment (DECA) in English and Spanish for ethnically diverse children living in poverty. *Early Education and Development, 22,* 520–547.

Cummings, E. M., Davies, P., & Campbell, S. B. (2000). *Developmental psychopathology and family process: Research, theory, and clinical implications.* New York: Guilford Press.

Daily, S., Burkhauser, M., & Halle, T. (2010). *A review of school readiness practices in the states: Early learning guidelines and assessments.* Washington, DC: Child Trends.

Deater-Deckard, K., Dodge, K. A., Bates, J. E., & Pettit, G. S. (1998). Multiple risk factors in the development of externalizing behavior: Group and individual differences. *Development and Psychopathology, 10,* 469–493.

Denham, S. A. (1998). *Emotional development in young children.* New York: Guilford Press.

Denham, S. A. (2006). Social-emotional competence as support for school readiness: What is it and how do we assess it? *Early Education & Development, 17*(1), 57–89. http://dx.doi.org/10.1207/s15566935eed1701_4.

Denham, S. A., Bassett, H. H., Brown, C. A., Way, E., & Steed, J. (2013). "I know how you feel": Preschoolers' emotion knowledge contributes to early school success. *Journal of Early Childhood Research, 13*(3), 252–262 (Advance online publication http://dx.doi.org/1476718x13497354).

Denham, S. A., Bassett, H. H., Mincic, M. M., Kalb, S. C., Way, E., Wyatt, T., et al. (2012). Social-emotional learning profiles of preschoolers' early school success: A person-centered approach. *Learning and Individual Differences, 22,* 178–189.

Denham, S. A., Bassett, H. H., Thayer, S. K., Mincic, M. M., Sirotkin, Y. S., & Zinsser, K. (2012). Observing preschoolers' social-emotional behavior: Structure, foundations, and prediction of early school success. *Journal of Genetic Psychology, 173,* 246–278.

Denham, S. A., Bassett, H. H., Way, E., Kalb, S. C., Warren-Khot, H. K., & Zinsser, K. (2014). How would you feel? What would you do?" Development and underpinnings of preschoolers' social information processing. *Journal of Research in Childhood Education, 28,* 182–202.

Denham, S. A., Bassett, H. H., Way, E., Mincic, M., Zinsser, K., & Graling, K. (2011). Preschoolers' emotion knowledge: Self-regulatory foundations, and predictions of early school success. *Cognition and Emotion, 26,* 667–679.

Denham, S. A., Bouril, B., & Belouad, F. (1994). Challenging situations task. *Child Study Journal, 24*(1), 1–21.

Denham, S. A., Caverly, S., Schmidt, M., & Blair, K. (2002). Preschool understanding of emotions: Contributions to classroom anger and aggression. *Journal of Child Psychology and Psychiatry, 43*(7), 901–916.

Denham, S. A., Ji, P., & Hamre, B. (2010). *Compendium of social-emotional learning and associated assessment measures.* Chicago, IL: Collaborative for Academic, Social, and Emotional Learning.

Denham, S. A., Mason, T., Kochanoff, A., Neal, K., & Hamada, H. (2003). Emotional development. In Dawn Cavalieri (Ed.), *International encyclopedia of marriage and family relationships* (pp. 419–426) (second ed.). New York: Macmillan.

Denham, S. A., Way, E., Kalb, S. C., Warren-Khot, H. K., & Bassett, H. H. (2013). Preschoolers' social information processing and school readiness: The Challenging Situations Task. *British Journal of Developmental Psychology, 31,* 180–197.

Denham, S. A., Wyatt, T. M., Bassett, H. H., Echeverria, D., & Knox, S. S. (2009). Assessing social-emotional development in children from a longitudinal perspective. *Journal of Epidemiology and Community Health, 63,* I37–I52.

Domitrovich, C. E., Cortes, R. C., & Greenberg, M. T. (2007). Improving young children's social and emotional competence: A randomized trial of the preschool "PATHS" curriculum. *The Journal of Primary Prevention, 28*, 67–91.

Downer, J. T., Booren, L. M., Lima, O. K., Luckner, A. E., & Pianta, R. C. (2010). The individualized classroom assessment scoring system (inCLASS): Preliminary reliability and validity of a system for observing preschoolers' competence in classroom interactions. *Early Childhood Research Quarterly, 25*(1), 1–16. http://dx.doi.org/10.1016/j.ecresq.2009.08.004.

Downer, J. T., & Pianta, R. C. (2006). Academic and cognitive functioning in first grade: Associations with earlier home and child care predictors and with concurrent home and classroom experiences. *School Psychology Review, 35*, 11–30.

Duckworth, A. L., & Seligman, M. E. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science, 16*(12), 939–944.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing Students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82*, 405–432.

Dykas, M. J., & Cassidy, J. (2011). Attachment and the processing of social information across the life span: theory and evidence. *Psychological Bulletin, 137*(1), 19–46.

Espy, K. A. (1997). The Shape School: Assessing executive function in preschool children. *Developmental Neuropsychology, 13*(4), 495–499.

Espy, K. A., Kaufmann, P. M., Glisky, M. L., & McDiarmid, M. D. (2001). New procedures to assess executive functions in preschool children. *The Clinical Neuropsychologist, 15*(1), 46–58.

Espy, K. A., Kaufmann, P. M., McDiarmid, M. D., & Glisky, M. L. (1999). Executive functioning in preschool children: Performance on A-not-B and other delayed response format tasks. *Brain and Cognition, 41*(2), 178–199.

Fabes, R. A., Gaertner, B. M., & Popp, T. K. (2006). Getting along with others: Social competence in early childhood. In K. McCartney, & D. Phillips (Eds.), *Handbook of early childhood development* (pp. 297–316). Malden, MA: Blackwell.

Federal Interagency Forum on Child and Family Statistics (2015). *Inventory of existing measures of social and emotional development in early childhood.* Washington, DC: Author (Retrieved from http://www.childstats.gov/forum/deliverables.asp).

Feldman, R., Bamberger, E., & Kanat-Maymon, Y. (2013). Parent-specific reciprocity from infancy to adolescence shapes children' social competence and dialogical skills. *Attachment & Human Development, 4*(15), 407–423.

Gandolfi, E., Viterbori, P., Traverso, L., & Usai, M. C. (2014). Inhibitory processes in toddlers: a latent-variable approach. *Frontiers in Psychology, 5*. http://dx.doi.org/10.3389/fpsyg.2014.00381.

Garner, P., & Waajid, B. I. (2008). The associations of emotion knowledge and teacher–child relationships to preschool children's school-related developmental competence. *Journal of Applied Developmental Psychology, 29*, 89–100.

Garner, P. W., & Waajid, B. (2012). Emotion knowledge and self-regulation as predictors of Preschoolers' cognitive ability, classroom behavior, and social competence. *Journal of Psychoeducational Assessment, 30*(4), 330–343.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*(1), 31–60.

Garon, N., Smith, I. M., & Bryson, S. E. (2014). A novel executive function battery for preschoolers: Sensitivity to age differences. *Child Neuropsychology, 20*(6), 713–736. http://dx.doi.org/10.1080/09297049.2013.857650.

Gartstein, M. A., & Rothbart, M. K. (2003). Studying infant temperament via the Revised Infant Behavior Questionnaire. *Infant Behavior & Development, 26*, 64–86.

Gioia, G. A., Espy, K. A., & Isquith, P. K. (2003). *The Behavior Rating Inventory of Executive Function-Preschool version (BRIEF-P).* Odessa, FL: Psychological Assessment Resources.

Goldberg, E., & Podell, K. (2000). Adaptive decision making, ecological validity, and the frontal lobes. *Journal of Clinical and Experimental Neuropsychology, 22*(1), 56–68. http://dx.doi.org/10.1076/1380-3395(200002)22:1;1-8;FT056.

Graziano, P. A., Reavis, R. D., Keane, S. P., & Calkins, S. D. (2007). The role of emotion regulation in children's early academic success. *Journal of School Psychology, 45*, 3–19.

Greenberg, M. T., Lengua, L. J., Coie, J. D., Pinderhughes, E. F., Bierman, K., Dodge, K. A., et al. (1999). Predicting developmental outcomes at school entry using a multiple risk model: Four American communities. *Developmental Psychology, 35*, 403–417.

Gresham, F. M., & Elliott, S. N. (2008). *Social skills improvement system rating scales manual.* Minneapolis, MN: NCS Pearson.

Halle, T., Vick Whittaker, J., Zepeda, M., Rothenberg, L., Anderson, R., Wessel, J., et al. (2014). The social-emotional development of Dual Language Learners: Looking back at existing research and moving forward with purpose. *Early Childhood Research Quarterly, 29*(4), 734–749.

Halle, T. G., & Darling-Churchill, K. E. (2016). Review of measures of social and emotional development. *Journal of Applied Developmental Psychology, 45*, 8–18 (in this issue).

Havighurst, S. S., Wilson, K. R., Harley, A. E., & Prior, M. R. (2009). Tuning into kids: An emotion-focused parenting program – Initial findings from a community trial. *Journal of Community Psychology, 37*, 1008–1023.

Heckman, J., Moon, S. H., Pinto, R., Savelyev, P., & Yavitz, A. (2010). Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope perry preschool program. *Quantitative Economics, 1*(1), 1–46.

Hemmeter, M. L., Ostrosky, M., & Fox, L. (2006). Social and emotional foundations for early learning: A conceptual model for intervention. *School Psychology Review, 35*, 583–601.

Henry, D. B. (2008). Changing classroom social settings through attention to norms. In M. Shinn, & H. Yoshikawa (Eds.), *Toward positive youth development: Transforming schools and community programs.* London: Oxford University Press.

Herndon, K. J., Bailey, C. S., Shewark, E., Denham, S. A., & Bassett, H. H. (2013). Preschoolers' emotion expression and regulation: Relations with school adjustment. *Journal of Genetic Psychology, 174*, 642–663.

Hogan, A. E., Scott, K. G., & Bauer, C. R. (1992). The Adaptive Social Behavior Inventory (ASBI): A new assessment of social competence in high-risk three-year-olds. *Journal of Psychoeducational Assessment, 10*, 230–239.

Howes, R., Calkins, S., Anastopoulos, A., Keane, S., & Shelton, T. (2003). Regulatory contributors to children's kindergarten achievement. *Early Education and Development, 14*, 101–119.

Hughes, C., Ensor, R., Wilson, A., & Graham, A. (2010). Tracking executive function across the transition to school: A latent variable approach. *Developmental Neuropsychology, 35*(1), 20–36.

Hyson, M. C. (1994). *The emotional development of young children: Building an emotion-centered curriculum.* New York: Teachers College Press.

Isquith, P. K., Roth, R. M., & Gioia, G. (2013). Contribution of rating scales to the assessment of executive functions. *Applied Neuropsychology: Child, 2*, 125–132.

Isquith, P. K., Roth, R. M., Kenworthy, L., & Gioia, G. (2014). Contribution of rating scales to intervention for executive dysfunction. *Applied Neuropsychology. Child, 3*, 197–204. http://dx.doi.org/10.1080/21622965.2013.870014 (PMID 24559500).

Izard, C. E., Fine, S., Schultz, D., Mostow, A., Ackerman, B., & Youngstrom, E. (2001). Emotions knowledge as a predictor of social behavior and academic competence in children at risk. *Psychological Science, 12*, 18–23.

Jones, S. M., & Bouffard, S. (2012). Social and emotional learning in schools: From programs to strategies. *Social Policy Report, 26*(4) (Society for Research in Child Development).

Kellam, S. G., Ling, X., Merisca, R., Brown, C. H., & Ialongo, N. (1998). The effect of the level of aggression in the first grade classroom on the course and malleability of aggressive behavior into middle school. *Development and Psychopathology, 10*, 165–185.

Kendziora, K., Weissberg, R. P., & Dusenbury, L. (2011). *Strategies for social and emotional learning: Preschool and elementary grade student learning standards and assessment.* Newtown, MA: National Center for Mental Health Promotion and Youth Violence Prevention, Education Development Center, Inc.

Kidd, C., Palmeri, H., & Aslin, R. N. (2013). Rational snacking: Young children's decision-making on the marshmallow task is moderated by beliefs about environmental reliability. *Cognition, 126*(1), 109–114.

Knudsen, B., & Liszkowski, U. (2012). 18-month-olds predict specific action mistakes through attribution of false belief, not ignorance, and intervene accordingly. *Infancy, 17*(6), 672–691.

Kochanska, G., Philibert, R., & Barry, R. A. (2009). Interplay of genes and early mother-child relationship in the development of self-regulation from toddler to preschool age. *Journal of Child Psychology and Psychiatry, 50*, 1331–1338.

Kopp, C. B. (1989). Regulation of distress and negative emotions: A developmental view. *Developmental Psychology, 25*, 343–354.

Kuczynski, L., & Kochanska, G. (1995). Function and content of maternal demands: Developmental significance of early demands for competent action. *Child Development, 66*(3), 616–628.

La Paro, K. M., Pianta, R. C., & Stuhlman, M. (2004). The classroom assessment scoring system: Findings from the prekindergarten year. *The Elementary School Journal, 104*(5), 409–426. http://dx.doi.org/10.1086/499760.

Ladd, G. W., Birch, S. H., & Buhs, E. S. (1999). Children's social and scholastic lives in kindergarten: Related spheres of influence? *Child Development, 70*, 1373–1400.

Ladd, G. W., Herald, S. L., & Kochel, K. P. (2006). School readiness: Are there social prerequisites? *Early Education and Development, 17*, 115–150.

Lafreniere, P., Masataka, N., Butovskaya, M., Chen, Q., Dessen, M. A., Atwanger, K., et al. (2002). Cross-cultural analysis of social competence and behavior problems in preschoolers. *Early Education and Development, 13*(2), 201–220. http://dx.doi.org/10.1207/s15566935eed1302_6.

Laible, D., & Thompson, R. A. (2007). Early socialization: A relationship perspective. In J. Grusec, & P. Hastings (Eds.), *Handbook of socialization: Theory and research* (pp. 181–207). New York: Guilford Press.

LeBuffe, P. A., & Naglieri, J. A. (1999). *Devereux Early Childhood Assessment Technical Manual.* Lewisville, NC: Kaplan Press.

LeBuffe, P. A., & Naglieri, J. A. (2013). *Devereux Early Childhood Assessment (P-2) User's Guide.* Lewisville, NC: Kaplan Press.

Lee, K., Bull, R., & Ho, R. M. (2013). Developmental changes in executive functioning. *Child Development, 84*(6), 1933–1953. http://dx.doi.org/10.1111/cdev.12096.

Leerkes, E. M., Paradise, M., O'Brien, M., Calkins, S. D., & Lange, G. (2008). Emotion and cognition processes in preschool children. *Merrill-Palmer Quarterly, 54*, 102–124.

Lerner, M. D., & Lonigan, C. J. (2014). Executive function among preschool children: Unitary versus distinct abilities. *Journal of Psychopathology and Behavioral Assessment, 36*(4), 626–639. http://dx.doi.org/10.1007/s10862-014-9424-3.

Love, J. M., Chazen-Cohen, R., Raikes, H., & Brooks-Gunn, J. (2013). What makes a difference: Early Head Start evaluation findings in developmental context. *Monographs of the Society for Research in Child Development, 78* (whole no. 1).

Mahone, E. M., & Hoffman, J. (2007). Behavior ratings of executive function among preschoolers with ADHD. *The Clinical Neuropsychologist, 21*(4), 569–586. http://dx.doi.org/10.1080/13854040600762724.

McAuley, T., Chen, S., Goos, L., Schachar, R., & Crosbie, J. (2010). Is the behavior rating inventory of executive function more strongly associated with measures of impairment or executive function? *Journal of the International Neuropsychological Society, 16*(3), 495–505. http://dx.doi.org/10.1017/S1355617710000093.

McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acock, A. C., Miao, A., et al. (2014). Predictors of early growth in academic achievement: The Head-Toes-Knees-Shoulders task. *Frontiers in Psychology, 5*, 599.

McClelland, M. M., Morrison, F. J., & Holmes, D. H. (2000). Children at-risk for early academic problems: The role of learning-related social skills. *Early Childhood Research Quarterly, 15*, 307–329.

McWayne, C. M., & Cheung, K. (2009). A picture of strength: Preschool competencies mediate the effects of early behavior problems on later academic and social

adjustment for Head Start children. *Journal of Applied Developmental Psychology, 30*, 273–285.

McWayne, C. M., Fantuzzo, J. W., & McDermott, P. A. (2004). Preschool competency in context: An investigation of the unique contribution of child competencies to early academic success. *Developmental Psychology, 40*, 633–645.

Meisels, S. J. (1999). Assessing readiness. In R. C. Pianta, & M. Cox (Eds.), *The transition to kindergarten* (pp. 39–66). Baltimore: Brookes.

Miller, M. R., Muller, U., Giesbrecht, G. F., Carpendale, J. I., & Kerns, K. A. (2013). The contribution of executive function and social understanding to preschoolers' letter and math skills. *Cognitive Development, 28*(4), 331–349. http://dx.doi.org/10.1016/j.cogdev.2012.10.005.

Mische, W., et al. (2011). 'Willpower' over the life span: Decomposing self-regulation. *Social Cognitive and Affective Neuroscience, 6*(2), 252–256.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21*(1), 8–14. http://dx.doi.org/10.1177/0963721411429458.

Moffitt, E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences, 108*(7), 2693–2698.

Moffitt, T. E., Caspi, A., Dickson, N., Silva, P., & Stanton, W. (1996). Childhood onset versus adolescent onset antisocial behavior: Natural history from ages 3 to 18. *Development and Psychopathology, 8*, 399–424.

Mooney, B., Walmsley, C., & McFarland, K. (2006). Factor analysis of the self-report Dysexecutive (DEX-S) Questionnaire. *Applied Neuropsychology: Adult, 13*(1), 12–18. http://dx.doi.org/10.1207/s15324826an1301_2.

Morrison, F. J., & Grammer, J. K. (2016). Conceptual clutter and measurement mayhem: Proposals for cross disciplinary integration in conceptualizing and measuring executive function. In J. Griffin, L. Freund, & P. McCardle (Eds.), *Executive function in preschool age children: Integrating measurement, neurodevelopment and translational research.* Washington DC: American Psychological Association Press (in press).

National Research Council (2008). *Early childhood assessment: Why, what, and how.* Washington, DC: National Academies Press.

National School Readiness Indicators Initiative (2005). Findings from the National School Readiness Indicators Initiative: A 17 state partnership. Retrieved from http://www.gettingready.org/

Neidell, M., & Waldfogel, J. (2008). Cognitive and Non-cognitive peer effects in early education. *National Bureau of Economic Research (NBER) Working Paper Series #14277* (Available at http://www.nber.org/papers/w14277.pdf).

Nelson, J. A., Leerkes, E. M., Perry, N. B., O'Brien, M., Calkins, S. D., & Marcovitch, S. (2013). European-American and African-American mothers' emotion socialization practice relate differently to their childrens' academic and social-emotional competence. *Social Development, 22*(3), 485–498.

NICHD Early Child Care Research Network (2003). Social functioning in first grade: Associations with earlier home and child care predictors and with current classroom experiences. *Child Development, 74*, 1639–1662.

NICHD Early Child Care Research Network (2004). Trajectories of aggression from toddlerhood to middle childhood: Predictors, correlates, and outcomes. *Monographs of the Society for Research in Child Development, 69* (whole no. 4).

NICHD Early Child Care Research Network (2005). Duration and developmental timing of poverty and children's cognitive and social development from birth through third grade. *Child Development, 76*, 795–810.

Nix, R. L., Bierman, K. L., Domitrovich, C. E., & Gill, S. (2013). Promoting children's social-emotional skills in preschool can enhance academic and behavioral functioning in kindergarten: Findings from Head Start REDI. *Early Education and Development, 24*, 1000–1019.

Passler, M. A., Isaac, W., & Hynd, G. W. (1985). Neuropsychological development of behavior attributed to frontal lobe functioning in children. *Developmental Neuropsychology, 1*(4), 349–370.

Ponitz, C. E. C., McClelland, M. M., Jewkes, A. M., Connor, C. M., Farris, C. L., & Morrison, F. J. (2008). Touch your toes! Developing a direct measure of behavioral regulation in early childhood. *Early Childhood Research Quarterly, 23*(2), 141–158.

Putnam, S. P., Gartstein, M. A., & Rothbart, M. K. (2006). Measurement of fine-grained aspects of toddler temperament: The Early Childhood Behavior Questionnaire. *Infant Behavior & Development, 29*, 386–401.

Putnam, S. P., & Rothbart, M. K. (2006). Development of short and very short forms of the children's behavior questionnaire. *Journal of Personality Assessment, 87*, 103–113.

Raver, C. C. (2012). Low-income children's self-regulation in the classroom: Scientific inquiry for social change. *American Psychologist, 67*(8), 681–689. http://dx.doi.org/10.1037/a0030085.

Raver, C. C., Jones, S. M., Li-Grining, C. P., Zhai, F., Bub, K., & Pressler, E. (2011). CSRP's impact on low-income preschoolers' pre-academic skills: Self-regulation and teacher-student relationships as two mediating mechanisms. *Child Development, 82*(1), 362–378.

Raver, C. C., & Knitzer, J. (2002). *Ready to enter: What research tells policymakers about strategies to promote social and emotional school readiness among three- and four-year-old children.* New York, NY: National Center for Children in Poverty.

Raver, C. C., & Zigler, E. F. (1997). Social competence: An untapped dimension evaluating Head Start's success. *Early Childhood Research Quarterly, 12*, 363–385.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: two approaches for exploring measurement invariance. *Psychological Bulletin, 114*(3), 552–566.

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior assessment system for children* (2nd ed.). Circle Pines, MN: American Guidance Service.

Rimm-Kaufman, S. E., Curby, T. W., Grimm, K. J., Nathanson, L., & Brock, L. L. (2009). The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. *Developmental Psychology, 45*, 958–972.

Rimm-Kaufman, S. E., Pianta, R. C., & Cox, M. J. (2000). Teachers' judgments of problems in the transition to kindergarten. *Early Childhood Research Quarterly, 15*, 147–166.

Rispoli, K. M., McGoey, K. E., Koziol, N. A., & Schreiber, J. B. (2013). The relation of parenting, child temperament, and attachment security in early childhood to social competence at school entry. *Journal of School Psychology, 51*(5), 643–658.

Rose-Krasnor, L. (1997). The nature of social competence: a theoretical review. *Social Development, 6*, 111–135.

Rothbart, M. K. (2007). Temperament, development, and personality. *Current Directions in Psychological Science, 16*, 207–212.

Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at 3 to 7 years: The Child Behavior Questionnaire. *Child Development, 72*, 1394–1408.

Rudolph, K. D., Hammen, C., & Burge, D. (1995). Cognitive representations of self, family, and peers in school-age children: Links with social competence and sociometric status. *Child Development, 66*, 1385–1402. http://dx.doi.org/10.2307/1131653.

Saarni, C. (1999). *Children's emotional competence.* New York: Guilford Press.

Sameroff, A. J. (2009). *The transactional model of development: How children and contexts shape each other.* Washington, DC: American Psychological Association.

Santos, A. J., Vaughn, B. E., Peceguina, I., Danial, J. R., & Shin, N. (2014). Growth of social competence during the preschool years: A 3-year longitudinal study. *Child Development, 85*(5), 2062–2073.

Sbordone, R. J. (2001). Limitations of neuropsychological testing to predict the cognitive and behavioral functioning of persons with brain injury in real-world settings. *NeuroRehabilitation, 16*(4), 199–201.

Sbordone, R. J. (2014). The hazards of strict reliance on neuropsychological tests. *Applied Neuropsychology: Adult, 21*(2), 98–107. http://dx.doi.org/10.1080/09084282.2012.762630.

Sbordone, R. J., & Purisch, A. D. (1996). Hazards of blind analysis of neuropsychological test data in assessing cognitive disability: the role of confounding factors. *NeuroRehabilitation, 7*(1), 15–26. http://dx.doi.org/10.3233/NRE-1996-7103.

Schoemaker, K., Bunte, T., Wiebe, S. A., Espy, K. A., Dekovic, M., & Matthys, W. (2012). Executive function deficits in preschool children with ADHD and DBD. *Journal of Child Psychology and Psychiatry, 53*(2), 111–119. http://dx.doi.org/10.1111/j.1469-7610.2011.02468.x.

Schussler, L. E. (2012). *The development of the Children's Behavior Questionnaire, Teacher Version (CBQ-T).* (Doctoral dissertation) University of Maryland College Park.

Shaw, D. S., Dishion, T. J., Supplee, L., Gardner, F., & Arnds, K. (2006). Randomized trial of a family-centered approach to the prevention of early conduct problems: 2-year effects of the family check-up in early childhood. *Journal of Consulting and Clinical Psychology, 74*, 1–9.

Shaw, D. S., Gilliam, M., Ingoldsby, E. M., & Nagin, D. S. (2003). Trajectories leading to school-age conduct problems. *Developmental Psychology, 39*, 189–200.

Shaw, D. S., Hyde, L. W., & Brennan, L. M. (2012). Early predictors of boys' antisocial trajectories. *Development and Psychopathology, 24*, 871–888.

Shields, A., & Cicchetti, D. (1997). Emotion regulation in school-age children: The development of a new criterion Q-sort scale. *Developmental Psychology, 33*, 906–916.

Shields, A., Dickstein, S., Seifer, R., Giusti, L., Magee, K. D., & Spritz, B. (2001). Emotional competence and early school adjustment: A study of preschoolers at risk. *Early Education and Development, 12*, 73–96.

Shing, Y. L., Lindenberger, U., Diamond, A., Li, S. C., & Davidson, M. C. (2010). Memory maintenance and inhibitory control differentiate from early childhood to adolescence. *Developmental Neuropsychology, 35*(6), 679–697.

Shonkoff, J., & Phillips, D. (2000). *From neurons to neighborhoods.* Washington, DC: National Academy Press.

Silver, C. H. (2014). Sources of data about children's executive functioning: review and commentary. *Child Neuropsychology, 20*(1), 1–13. http://dx.doi.org/10.1080/09297049.2012.727793.

Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of the Preschool Self-Regulation Assessment (PSRA) for field-based research. *Early Childhood Research Quarterly, 22*, 173–187.

Song, Q., & Wang, Q. (2013). Mother-child reminiscing about peer experiences and children's peer-related self-views and social competence. *Social Development, 22*(2), 280–299.

Sroufe, L. A. (1990). Considering normal and abnormal together: The essence of developmental psychopathology. *Development and Psychopathology, 2*, 335–347.

Sroufe, L. A. (1997). Psychopathology as an outcome of development. *Development and Psychopathology, 9*, 251–268.

Sroufe, L. A., & Rutter, M. (1984). The domain of developmental psychopathology. *Child Development, 55*, 15–24.

Sullivan, J. R., & Riccio, C. A. (2006). An empirical analysis of the BASC Frontal Lobe/Executive Control scale with a clinical sample. *Archives of Clinical Neuropsychology, 21*(5), 495–501. http://dx.doi.org/10.1016/j.acn.2006.05.008.

Teglasi, H., Schussler, L., Gifford, K., Annotti, L. A., Sanders, C., & Liu, H. (2015). *Child Behavior Questionnaire–Short Form for Teachers: informant correspondences and divergences.* (Assessment, pii: 1073191114562828. [Epub ahead of print]. PMID:25573857).

Thompson, R., & Lagattuta, K. H. (2006). Feeling and understanding: Early emotional development. In K. McCartney, & D. Phillips (Eds.), *Blackwell handbook of early childhood development* (pp. 317–338). Malden, MA: Blackwell Publishing.

Thorell, L. B., & Nyberg, L. (2008). The Childhood Executive Functioning Inventory (CHEXI): A new rating instrument for parents and teachers. *Developmental Neuropsychology, 33*(4), 536–552.

Toplak, M. E., Bucciarelli, S. M., Jain, U., & Tannock, R. (2009). Executive functions: performance-based measures and the behavior rating inventory of executive function (BRIEF) in adolescents with attention deficit/hyperactivity disorder (ADHD). *Child Neuropsychology, 15*(1), 53–72. http://dx.doi.org/10.1080/09297040802070929.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner Review: Do performance-based measures and ratings of executive function assess the same construct? *Journal of Child Psychology and Psychiatry*, 54(2), 131–143. http://dx.doi.org/10.1111/Jcpp.12001.

Trentacosta, C. J., & Fine, S. E. (2009). Emotional knowledge, social competence, and behavioral problems in childhood and adolescence: A meta-analytic review. *Social Development*, 19(1), 1–29. http://dx.doi.org/10.1111/j.1467-9507.2009.00543.x.

U.S. Department of Health and Human Services [USDHHS], Administration for Children and Families, Office of Head Start (2011). *The Head Start child development and early learning framework: Promoting positive outcomes in early childhood programs serving children 3–5 years old.* Washington, DC: Author.

Ursache, A., Blair, C., & Raver, C. C. (2012). The promotion of self-regulation as a means of enhancing school readiness and early achievement in children at risk for school failure. *Child Development Perspectives*, 6(2), 122–128. http://dx.doi.org/10.1111/j.1750-8606.2011.00209.x.

Usai, M. C., Viterbori, P., Traverso, L., & De Franchis, V. (2014). Latent structure of executive function in five- and six-year-old children: A longitudinal study. *The European Journal of Developmental Psychology*, 11(4), 447–462. http://dx.doi.org/10.1080/17405629.2013.840578.

van der Ven, S. H., Kroesbergen, E. H., Boom, J., & Leseman, P. P. (2013). The structure of executive functions in children: a closer examination of inhibition, shifting, and updating. *British Journal of Developmental Psychology*, 31(Pt 1), 70–87. http://dx.doi.org/10.1111/j.2044-835X.2012.02079.x.

Wang, Q. (2003). Emotion situation knowledge in American and Chinese preschool children and adults. *Cognition and Emotion*, 17, 725–746.

Watanabe, N., Kobayashi, T., Bassett, H. H., & Denham, S. A. (2012). Measuring emotion knowledge in Japanese children: Cross-cultural comparison on the reliability of the Affect Knowledge Test. *Poster presented at the 2012 Society for Research in Human Development Biennial Meeting, New Orleans, LA.*

Waterman, C., McDermott, P. A., Fantuzzo, J. W., & Gadsden, V. L. (2012). The matter of assessor variance in early childhood education – Or whose score is it anyway? *Early Childhood Research Quarterly*, 27, 46–54.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality and Social Psychology*, 54, 1063–1070.

Weintraub, S., Bauer, P. J., Zelazo, P. D., Wallner-Allen, K., Dikmen, S. S., Heaton, R. K., et al. (2013). I. Nih Toolbox Cognition Battery (Cb): Introduction and Pediatric Data. *Monographs of the Society for Research in Child Development*, 78(4), 1–15. http://dx.doi.org/10.1111/Mono.12031.

Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology*, 7(2), 131–149.

Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, 4(1), 10–18.

Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology*, 44(2), 575–587.

Wiebe, S. A., Sheffield, T., Nelson, J. M., Clark, C. A. C., Chevalier, N., & Espy, K. A. (2011). The structure of executive function in 3-year-olds. *Journal of Experimental Child Psychology*, 108(3), 436–452. http://dx.doi.org/10.1016/j.jecp.2010.08.008.

Williford, A. P., Maier, M., Downer, J. T., Pianta, R. C., & Howes, C. (2013a). Understanding how children's engagement and teachers' interactions combine to predict school readiness. *Journal of Applied Developmental Psychology*, 34, 299–309.

Williford, A. P., Whittaker, J. E., Vitiello, V. E., & Downer, J. T. (2013b). Children's engagement within the preschool classroom and their development of self-regulation. *Early Education and Development*, 24, 162–187.

Willoughby, M. T., Blair, C. B., Wirth, R. J., Greenberg, M., & Investigators, F. L. P. (2010). The measurement of executive function at age 3 years: Psychometric properties and criterion validity of a new battery of tasks. *Psychological Assessment*, 22(2), 306–317.

Willoughby, M. T., Blair, C. B., Wirth, R. J., Greenberg, M., & Investigators, F. L. P. (2012). The measurement of executive function at age 5: Psychometric properties and relationship to academic achievement. *Psychological Assessment*, 24(1), 226–239.

Xu, F., Han, Y., Sabbagh, M. A., Wang, T., Ren, X., & Li, C. (2013). Developmental differences in the structure of executive function in middle childhood and adolescence. *PLoS ONE*, 8(10), e77770. http://dx.doi.org/10.1371/journal.pone.0077770.

Yudron, M., Jones, S. M., & Raver, C. C. (2014). Implications of different methods for specifying classroom composition of externalizing behavior and its relationship to social–emotional outcomes. *Early Childhood Research Quarterly*, 29(4), 682–691. http://dx.doi.org/10.1016/j.ecresq.2014.07.007.

Zhou, Q., Chen, S. H., & Main, A. (2012). Commonalities and differences in the research on children's effortful control and executive function: A call for an integrated model of self-regulation. *Child Development Perspectives*, 6(2), 112–121. http://dx.doi.org/10.1111/j.1750-8606.2011.00176.x.