

Memos on Measures of Social-Emotional Development in Early Childhood, by Subdomain

Prepared for

**The Federal Interagency Forum on Child and Family Statistics
Research and Innovation Committee**

Contract# 200-2012-F-51426

Presented here is a set of memos prepared by researchers with expertise in particular aspects of young children’s social and emotional development. These memos provide a broad overview of conceptual as well as measurement issues for each specific area of social and emotional development, and also discuss measures with particular potential for consideration for use in federal data collection efforts.

Table of Contents

Measuring Social-Emotional Development in Early Childhood: <i>Social Competence</i> , by Stephanie Jones & Monica Yudron, Harvard University	1
Promising Measures of Social and Emotional Development in Early Childhood (<i>sub-focus: Emotional competence</i>), by Susanne A. Denham & Grace Z. Howarth, George Mason University	6
Comments on the Measurement of Social and Emotional Development in Early Childhood (<i>sub-focus: Behavior problems</i>), by Susan B. Campbell, University of Pittsburgh	26
Recommendation for Measurement of <i>Self-Regulation</i> in Early Childhood, by C. Cybele Raver, New York University	36
Measurement of <i>Executive Function</i> in Early Childhood, by Michael Willoughby, Frank Porter Graham Child Development Institute	43

Measuring Social-Emotional Development in Early Childhood: *Social Competence*

Stephanie Jones & Monica Yudron

Harvard University

1. Why is the measurement of social competence among children 0-5 important, from a policy perspective? and
2. What is the relationship of social competence to subsequent developmental outcomes?

For young children, social-emotional competence provides a crucial foundation for the mastery of a range of skills important to academic achievement and attainment (Denham, 2002; Jones & Bouffard, 2012). As a result, social-emotional skills are included in a set of school readiness indicators commonly used to drive improvement in young children's ability to succeed in kindergarten and early elementary school (National School Readiness Indicators Initiative, 2005). Social-emotional competence is often thought of as a broad category that comprises a set of more specifically delineated skills. One of these skills is social-competence which is generally characterized as the effectiveness of a child in social interactions with peers and adults (Fabes, Gaertner & Popp, 2006). In order to be socially competent, a child must (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. Early life antecedents of social competence include temperament, self-regulatory skills, emotional understanding and communication, social information processing and communication skills (Fabes, Gaertner & Popp, 2006). It is important to understand social competence as building over time from a constellation of more basic social, emotional, and cognitive regulation skills such as basic attention, working memory, emotion understanding and regulation, and social cue detection (e.g., Jones & Bailey, 2013; Jones & Bouffard, 2012). As such, when considering approaches to supporting social competence in children, it is important to take a multiple-skill, or whole child, approach and to consider the relevant antecedents from a developmental, stage-salience perspective (e.g., Aber & Jones, 1997; Jones, 2006).

Measuring social competence should interest policymakers for two reasons. First, a large body of evidence links social competence in early childhood to a range of outcomes of interest to 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). Other research indicates that interactive play, social competence, and prosocial behavior during early childhood are associated with

motivation, liking school, attention and self-control, and literacy skills both cross-sectionally and longitudinally (e.g., Caprara, Barbaranelli, Pastorelli, Bandura, & Zimbardo, 2000; Coolahan, Fantuzzo, Miles & Stipek, 2006). Therefore, social competence may be an important lever in changing child performance on a range of behavioral and achievement tasks. Second, social competence is malleable, particularly in early childhood. In 2011, the *Journal of Applied Developmental Psychology* dedicated an entire issue to studies investigating the relationship between classroom climate features and the development of social competence. Together, these articles illustrate the powerful influence that teachers have in shaping social contexts and experiences which, in turn, can enhance children's social skills and shape emerging interactional skills (Bierman, 2011). Indeed, a number of recent experimental evaluations of classroom-based interventions in social-emotional learning, all of which target the promotion of high-quality social interactions among children and between adults and children, indicate that social competence can be built and supported with high quality intervention in preschool and elementary school (e.g., Jones Brown & Aber, 2011, Raver et al., 2009, Bierman et al., 2008), and importantly, may serve as an important link to supporting children's focus in the classroom and their academic skills (e.g., Jones Brown & Aber, 2011). In addition, to classroom environments, prior research also highlights the salience of home contexts for the optimal development of social competence (Brion-Meisels & Jones, 2011).

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, including social competence language in early learning standards and requiring relevant training (e.g., in basic child development, classroom management, and social-emotional learning) for early childhood educator certifications. Social policies, of course, also hold great potential for impacting parents' opportunities to invest in and care for children in the first five years of life. These policies range from those that provide paid maternal 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.

3. In regard to measuring social competence across ages 0-5 and among diverse populations, what issues need to be taken into account?
4. What criteria are most important to consider in selecting or developing measures of social competence?

Several issues must be taken into account when measuring social competence in young children. For example, any instruments selected should be appropriate to the children's age and development stage both in the content assessed as well as in the method used to collect information. Some early life antecedents to social competence are not fully developed in very young children. For example, communication skills are thought to be important to developing social competence. While the majority of children begin to gesture and point by fifteen months of age, few are speaking with detail and fluidity before age 3. In order to reflect upon 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. For these reasons, social competence in children younger than 3 cannot be measured in the same way as it is for older children. Two important precursor skills that can be measured in this 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 intentionality (Fabes, Gaertner & Popp, 2006).

As children age, a different set of measurement issues arise. First, social competence is a skill best observed in dyadic or group settings. Because of this a child's representations and perceptions of her own social competency are useful for understanding teacher, peer, and parent reports of social competency. Reports that do not align across reporters can be triangulated by contextualized measures such as with the use of observation protocols like 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 with each child's current level of social competence and the potential of the setting to enhance that social competence. As recent evidence supports, context characteristics are important for shaping social interactions among children in the setting. These characteristics include but are not limited to: (1) the emotional climate of the classroom, (2) relational style of the teacher, (3) nature of the pedagogic approach underlying classroom strategies, and (4) behavior management norms in school and classroom (Bierman, 2011). Several setting-level observational measures attend to these features including the Classroom Assessment Scoring System (CLASS; La Paro, Pianta & Stuhlman, 2004) and the Teaching Strategies Rating Scale (Raver & Morris, 2012).

Since the measurement of social competence hinges on the *effectiveness* or *appropriateness* of interactions a child has with others (Fabes, Gaertner & Popp, 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 in relating to 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.

Relatedly, 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. Specifically, peer

interactions in aggregate are thought to generate class norms for behavior and achievement and establish patterns through which children's experiences and competencies influence their peers, both concurrently and in future classrooms or groupings. 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 & Waldfogel (2008) find that the saturation of children in Kindergarten classes with preschool experience positively influences individual children's reading and math achievement through 3rd grade—even for children who did not attend preschool. Importantly, these authors also find 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, & Jaffee, 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 competence is both a function of individual characteristics and skills, but also very much a function of the attitudes, norms, and behaviors that are characteristic of the setting in which it is observed or about which it is reported.

5. What additional guidance would you provide for the Federal Interagency Forum on Child and Family Statistics about developing an indicator from existing data sources, or in collecting new data on social competence in early childhood?

Given the material presented above, I would 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 PIPPS); and (3) carefully considering the context in which, and perspective from which, such assessments or observations are generated.

References

- 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. doi:10.1016/j.appdev.2011.04.004
- 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), 717–35. doi:10.1017/S0954579410000416
- 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.
- 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. doi: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.
- Fabes, R.A., Gaertner, B.M., and Popp, T.K. (2006). Getting along with others: Social competence in early childhood. In K. McCartney and D. Phillips (Eds.), *Handbook of early childhood development* (pp. 297-316). Malden, MA: Blackwell.
- Jones, S.M., & Bouffard, S. (2012). Social and emotional learning in schools: From programs to strategies. *Social Policy Report, 23*(4).
- Lafreniere, P., Masataka, N., Butovskaya, M., Chen, Q., Dessen, M. A., Atwanger, K., Schreiner, S., et al. (2002). Cross-Cultural Analysis of Social Competence and Behavior Problems in Preschoolers. *Early Education and Development, 13*(2), 201–220.
- 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.
- National School Readiness Indicators Initiative. (2005). Getting ready. Findings from the National School Readiness Indicators Initiative: A 17 state partnership. Retrieved from Getting Ready website: <http://www.gettingready.org/matriarch/d.asp?PageID=303&PageName2=pdfhold&p=&PageName=Getting+Ready+-+Full+Report.pdf>
- Raver, C.C. & Morris P. A. (2012). The teaching strategies rating scale. Adapted from Domitrovich, C. E., Cortes, R., & Greenberg, M. T. (2000). The teacher style rating scale technical report. Unpublished manuscript, Pennsylvania State University

Promising Measures of Social and Emotional Development in Early Childhood
(*sub-focus: Emotional competence*)

Susanne A. Denham & Grace Z. Howarth

George Mason University

Table of Contents

Promising Measures of Social and Emotional Development in Early Childhood	7
The Importance of Measures of Social and Emotional Development.....	7
Relations with Subsequent Developmental Outcomes	8
Considerations for Social and Emotional Measurement in Diverse Populations.....	9
Criteria for Selecting and Developing Measures of Emotional Competence and Social Cognition for Early Childhood	9
Model of the Role of Assessment.	10
Plan for Using Assessment.	10
Criteria for Assessment Tools.	11
Selection of Measures for Emotional Competence and Social Cognition	12
References	14
Figure 1. Interrelated system of emotional competence and social cognitive development in educational settings.....	17
Table 1. Plan for Use of Assessment within a Larger Focus on Emotional Competence and Social Cognition.....	18
Table 2. <i>Screening Measures</i>	19
Table 3. <i>Skills Assessment (Generally Formative)</i>	20
Table 4. Examples of MPAC-R/S and ERQ Items.....	25

Promising Measures of Social and Emotional Development in Early Childhood

Although early childhood educators and parents have long recognized the importance of social cognition and emotional competence in supporting school readiness, positive relationships, and overall wellbeing and adjustment, current federal and state education initiatives focus almost exclusively on academic skills achievement (Cohen, 2006; 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 that balances positive social and emotional outcomes with more traditional academic outcomes. Specific suggestions include funding high quality early childhood education programs where teachers and administrators are trained and supported in social and emotional learning (SEL), improving access to early childhood education, early identification and intervention for socially at-risk children, and teaching social skills to children and families (Jones & Bouffard, 2012; Raver & Knitzer, 2002). For policymakers considering the proposed social and emotional education initiatives (and thus the allocation of considerable resources), a strong empirical base is important. In this report, we focus on one aspect of this base: the measurement of emotional competence and social cognition in children zero to five years of age and the relation of emotional competence and social cognition to subsequent developmental outcomes. Although we recognize that the measurement of social and emotional skills is valuable outside the context of early childhood education, we primarily focus on the early childhood education context due to the power of policy in shaping its framework.

The Importance of Measures of Social and Emotional Development

Broadly, social cognition refers to an individual's perceptions of and beliefs about his or her social world. Emotional competence is a more specific, but related, skill that we define 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 (Denham, 2006a). Social cognition and emotional competence undergo dramatic developmental changes within the first five years of life, with the expression of more sophisticated, nuanced, and regulated emotions throughout infancy, learning to look to others for social information, and beginning to understand, identify, and empathize with others' emotions (Warren, Bassett, & Denham, 2008). Measures that assess this development inform policymakers and educators about what emotions and behaviors are developmentally appropriate across early childhood and provide insight into when and even how to target interventions for children who lag behind. Additionally, measures of social cognition and emotional competence are valuable in assessing the effectiveness of SEL interventions.

At a fundamental level, social-emotional competence is a key component of school readiness and empirically supported SEL is thus important in early childhood classrooms

(Denham, 2006b). For at-risk children, SEL appears to be particularly important; for example, high levels of behavioral and emotion regulation appear to buffer family- and SES-related risk (Denham et al., 2012b; Sektnan, McClelland, Acock, & Morrison, 2010). Further, from both family and education policy perspectives, it is valuable to recognize that a primary educational aim of many parents and educators is to endow children with a sense of social responsibility (Cohen, 2006). Social cognition and emotional competence are not only foundational skills for socially responsible behaviors and worldviews, but also promote safe and supportive learning environments. From a policy perspective, an evidence base that guides curricula important to families, educators, and society as a whole is also important. Research on social and emotional interventions is optimistic, showing that positive social cognition and emotional competence are teachable skills, and that early interventions are effective long term (Boyd, Barnett, Bodrova, Leong, Gomby, 2006; Cohen, 2006).

Relations with Subsequent Developmental Outcomes

Social cognition and emotional competence are reciprocal constructs that are linked in support of various academic and adjustment outcomes. Emotion knowledge, a component of both social cognition and emotional competence, and social problem-solving (i.e., more mature social cognitive processing and more adaptive emotional and behavioral responses in response to challenging social situations), are related; furthermore, they are unique predictors of both concurrent and future adjustment and academic success in early childhood—above and beyond the influence of gender, age, and risk status (Denham et al., 2012b; Denham et al., 2011). In fact, all aspects of emotional competence described here, as well as social cognition, work together to promote school success (Denham et al., 2012a). Children who enter school with more positive profiles of emotional competence and social cognition (i.e., are emotionally positive and regulated, understand emotions, and can solve social problems) are more likely to develop positive and supportive relationships with peers and teachers, participate more in classroom activities, and achieve at higher levels throughout the early school years. Conversely, children who enter school with less positive profiles are more likely to be disliked and rejected by peers, develop less supportive relationships with teachers, participate and enjoy school less, achieve at lower levels, and ultimately may be at risk for later school failure (Ladd, Birch, & Buhs, 1999; NICHD, 2004; Raver & Knitzer, 2002).

Unfortunately, kindergarten teachers report that a significant proportion of students enter their classrooms without these important skills (Rimm-Kaufman, Pianta, & Cox, 2000). Given the crucial effects of early childhood emotional competence and social cognition on a variety of outcomes across developmental domains, their support is a promising area for research, policy, and educational action. Below we discuss innovative measures of social and emotional development that are useful in research and, in some cases, classroom settings.

Considerations for Social and Emotional Measurement in Diverse Populations

In designing and selecting measures for social cognition and 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 (Denham et al., 2011, in press). Across early childhood, direct assessments need to be unambiguous and allow children the option of responding non-verbally (Denham, Way, Kalb, Warren-Khot, & Bassett, 2012). Additionally, gender often affects children's scores on social-emotional measures with boys often behaving more aggressively during play or providing more aggressive responses (Denham et al., 2012b). 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 and social cognition. 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 social cognition.

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 emotions from parents, than children from the US (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 and social cognition during this age range (Denham et al., 2011, 2012a, in press). 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.

Criteria for Selecting and Developing Measures of Emotional Competence and Social Cognition for Early Childhood

Emotional competence and social cognitive skills form integral facets of developmental tasks during the infant/toddler/preschool period. Research reviewed here 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 and social cognitive skills of this child?* Due to this clear importance, means to measure these skills in both research and educational arenas are crucially needed. Adequate – hopefully excellent – assessment tools must be identified. In this memo, we focus most forcefully on the need to use emotional competence and social cognition assessments in

applied settings, for screening and evaluation (potentially serving functions of formative, and perhaps even summative assessment), but we acknowledge, too, that assessment tools are required for policy-related research purposes. Emotional competence and social cognition have historically been measured via a wide variety of mechanisms, including informant ratings, direct assessment, and observation. In many cases, however, the time required for direct assessments and observations, as well as resources required for training and observer/coder reliability, make these tools less practical for applied usage. Although rating systems, when carefully crafted, may be more appropriate for certain usage in applied settings (as well as useful for research purposes), our position is nonetheless that observational and direct assessment tools are very important in best understanding the young child. Our recent work focuses on shortening and ultimately computerizing such tools; thus, we will include such assessments.

To bridge the information and understanding gaps regarding assessment of early emotional competence and social cognition that exist amongst 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. It is important to note that applied assessment of any domain 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, interim, summative). It is important to see emotional competence and social cognitive assessment within this integrated framework.

Model of the Role of Assessment. Figure 1 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 memo.

Plan for Using Assessment. Table 1 one shows a potential planning sequence in which the use of assessment is embedded. We will follow this sequence in selecting assessment tools. Ideally, screening of emotional competence and social cognitive strengths and weaknesses by both parents and teachers would be first steps, 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 whenever possible psychometrically

adequate measures should be utilized, even if not yet norm-referenced; this view guides our choice of measures. Formative assessment is followed by programming for both classrooms and parenting, 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 these are typically administered statewide, and are usually used as part of an accountability program or to otherwise inform policy; alternatively, they 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.

The terrain of inserting assessment tools into the system depicted in Figure 1 and Table 1 is very new and largely uncharted, especially for competence-based measures of emotional competence and social cognition, as opposed to broader measures of social behavior and behavior *problems*. To begin moving toward such a model and assessment plan, however, criteria for “best-bet” measures are now outlined, followed by choices of assessment tools and commentary.

Criteria for Assessment Tools. We know that some states are acknowledging the importance of the skills focused upon in this memo, 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 and social cognitive functioning. Denham et al. (2009) and Kendziora, Weissberg, & 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 of programming); this area of concern is far from systematized. Specifically for the skills focused upon here, documentation should make clear that the assessment is appropriate for an age period from infancy through kindergarten.

Families of measures that cut across age periods may be desirable, but more important is age-appropriate assessment.

Second, qualities of the actual assessment tool must be considered. 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 in this memo meet psychometric requirements, but issues of fairness and generalizability require more study. For example, norms and psychometric data for measures must be obtained for diverse samples representing the demographics of US, 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 to help interpret results and interpret score change over time, to make assessment tools useful in tracking the results of instruction and programming. All such tools should be administrable within a reasonable time frame (e.g., 10–20 minutes). 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, classrooms, and school's competencies (for possible use in formative and summative assessment), but time constraints may indicate that only rating a random sample of students in each classroom is possible. Such restricted usage could still be valuable for interim and summative functions.

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 and Social Cognition

Given these criteria, it is important to note that no one brief indicator stands a chance of meeting all criteria, and especially impossible is finding indicators that are brief and comprehensive of all aspects of emotional competence and social cognition across the pertinent age range. Also, as noted elsewhere, 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. Thus, we present choices that retain as many of the above criteria as possible; we realize that it is

unlikely that all could be used in any one endeavor, but feel that these capture best the developmentally appropriate aspects of emotional competence and social cognition.

We present these measures according to the plan depicted in Table 1, and for the functions shown in Figure 1. It should be noted that perusal of the SE Measures Inventory Chart showed no new measures to choose other than those with which we were not already familiar, or in fact authors of, with the possible exception of Pettit, Dodge, and Brown's (1988) Social Problem-Solving Measure; we are not aware of other measures from nationwide surveys. Thus, in Table 2 we illustrate potential screening measures for both the 0 to 3 and 3 to 5 year age ranges, along with very brief descriptions and summaries of strengths and weaknesses. It is important that these screeners do not cover all the aspects of social cognition and emotional competence; nonetheless, we feel they would be quite useful for screening purposes, and potentially for limited use for the next aspect of our assessment plan, that of formative (perhaps summative) assessment of skills. Measures for this step of our assessment plan are shown in Table 3, again along with very brief descriptions and summaries of strengths and weaknesses. It is obvious that there are not ratings, direct assessments, and observational methods for each aspect of emotional competence and social cognition. This is, however, a sensible caveat, because raters or observers could not know details of a child's social cognition – but could observe emotional expression and emotion regulation, for example. Similarly, because of the developmental nature of these skills, we include no measures of infant/toddler emotion knowledge or social cognition, and were unable to find a useful measure (short, requiring little or no training) of infant/toddler emotion regulation (although possible the ERQ could be used with two-year-olds). To give more detail, in Table 4 we give examples of observed behaviors for the MPAC-R/S and the ERQ.

Finally, we include attachments, including two compendia of measures, an article that discusses social-emotional measures in infancy and early childhood, and the manuals for the AKT, CST, and MPAC-R/S. We hope that these efforts will be found useful.

References

- Boyd, J., Barnett, S., Bodrova, E., Leong, D. J., & Gomby, D. (2005, March). *Promoting children's social and emotional development through preschool education* (Preschool Policy Brief). New Burnswick, NJ: National Institute for Early Education Research.
- Brassard, M. R., & Boehm, A. E. (2007). *Preschool assessment: Principles and practices*. New York: Guilford Press.
- Cohen, J. (2006). Social, emotional, ethnical, and academic education: Creating a climate for learning, participating in democracy, and well-being. *Harvard Educational Review, 76*, 201-237.
- Denham, S. A. (2006a). Emotional competence in preschoolers: Implications for social functioning. In J. Luby (Ed.), *Handbook of Preschool Mental Health: Development, Disorders and Treatment* (pp. 23-44). New York: Guilford.
- Denham, S. A. (2006b). Social-emotional competence as support for school readiness: What is it and how do we assess it? *Early Education and Development, 17*, 57-89. doi: 10.1207/s15566935eed1701_4
- Denham, S. A. Bassett, H. H., Mincic, M.M., Kalb, S. C., Way, E., Wyatt, T., & Segal, Y. (2012a). Social-emotional learning profiles of preschoolers' early school success: A person-centered approach. *Learning and Individual Differences, 22*, 178-189. doi: 10.1016/j.lindif.2011.05.001
- Denham, S. A., Bassett, H. H., Thayer, S. K., Mincic, M., Sirotkin, Y.S, & Zinsser, K. (2012b). Observing preschoolers' social-emotional behavior: Structure, foundations, and prediction of early school success. *Journal of Genetic Psychology, 173*, 246-278. doi: 10.1080/00221325.2011.597457
- Denham, S. A., Bassett, H. H., Way, E., Kalb, S. C., Warren-Khot, H. K., & Zinsser, K. (in press). "How would you feel? What would you do?" Development and underpinnings of social information processing. *Journal of Research in Childhood Education*.
- 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. doi: 10.1080/02699931.2011.602049
- Denham, S. A., Way, E., Kalb, S. C., Warren-Khot, H. K. and Bassett, H. H. (2012), Preschoolers' social information processing and early school success: The challenging situations task. *British Journal of Developmental Psychology, 17*, 57-89. doi: 10.1111/j.2044-835X.2012.02085.x
- 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*, 137-152. doi:10.1136/jech.2007.070797
- 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. doi:10.1007/s10935-007-0081-0
- 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. doi:10.1111/j.1467-8624.2010.01564.x

- Gresham, F. M., & Elliott, S. N. (2008). *Social Skills Improvement System Rating Scales manual*. Minneapolis, MN: NCS Pearson.
- 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. doi:10.1002/jcop.20345
- 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.
- Jones, S. M. & Bouffard, S. M. (2012). *Social and emotional learning in schools: From programs to strategies* (Social Policy Report 26, 4). Society for Research in Child Development.
- 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.
- 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. doi: 10.1111/1467-8624.00101
- LaFreniere, P. J., & Dumas, J. E. (1996). Social competence and behavior evaluation in children ages 3 to 6 years: The short form (SCBE-30). *Psychological Assessment, 8*, 369–377. doi:10.1037/1040-3590.8.4.369
- Morgan, J. K., Izard, C. E., & King, K. A. (2010). Construct validity of the Emotion Matching Task: Preliminary evidence for convergent and criterion validity of a new emotion knowledge measure for young children. *Social Development, 19*, 52–70. doi:10.1111/j.1467-9507.2008.00529.x
- NICHD Early Care Research Network. (2004). Multiple pathways to early academic achievement. *Harvard Educational Review, 74*, 1-29.
- Pettit, G. S., Dodge, K. A., & Brown, M. M. (1988). Early family experiences, social problem solving patterns, and children's social competence. *Child Development, 59*, 107-120.
- 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* (Promoting the Emotional Well-being of Children and Families Policy Paper No.3). New York, NY: National Center for Children in Poverty.
- 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. doi: [http://dx.doi.org/10.1016/S0885-2006\(00\)00049-1](http://dx.doi.org/10.1016/S0885-2006(00)00049-1)
- Rothbart, M. K., Ahadi, S. A., Hershey, K., & Fisher, P. (2001). Investigations of temperament at three to seven years: The Children's Behavior Questionnaire. *Child Development, 72*, 1394-1408. doi: 10.1111/1467-8624.00355
- Rothbart, M. K., & Hwang, J. (2002). Measuring infant temperament. *Infant Behavior and Development, 25*, 113–116. doi:10.1016/S0163-6383(02)00109-1
- Schultz, D., Ambike, A., Logie, S. K., Bohner, K. E., Stapleton, L. M., VanderWalde, H., Betkowski, J. A. (2010). Assessment of social information processing in early childhood: Development and initial

- validation of the Schultz Test of Emotion Processing—Preliminary Version. *Journal of Abnormal Child Psychology*, *38*, 601–613. doi:10.1007/s10802-010-9390-5
- Sektnan, M., McClelland, M. M., Acock, A., & Morrison, F. J. (2010). Relations between early family risk, children's behavioral regulation, and academic achievement. *Early Childhood Research Quarterly*, *25*, 464-479. doi: 10.1016/j.ecresq.2010.02.005
- 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. doi: 10.1037//0012-1649.33.6.906
- Squires, J., Bricker, D., Heo, K., & Twombly, E. (2001). Identification of social-emotional problems in young children using a parent-completed screening measure. *Early Childhood Research Quarterly*, *16*, 405–419. doi:10.1016/S0885-2006(01)00115-6
- Wang, Q. (2003). Emotion situation knowledge in American and Chinese preschool children and adults. *Cognition and Emotion*, *17*, 725-746. doi: 10.1080/02699930244000156
- Warren, H. K., Bassett, H. H., & Denham, S. A. (2008). Early emotional development, social cognition and relationships with others. *Zero to Three*, May, 32-39.
- Watanabe, N., Kobayashi, T., Bassett, H. H., & Denham, S. A. (2012, March). *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.
- 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. doi: 10.1037//0022-3514.54.6.1063

Figure 1. Interrelated system of emotional competence and social cognitive development in educational settings

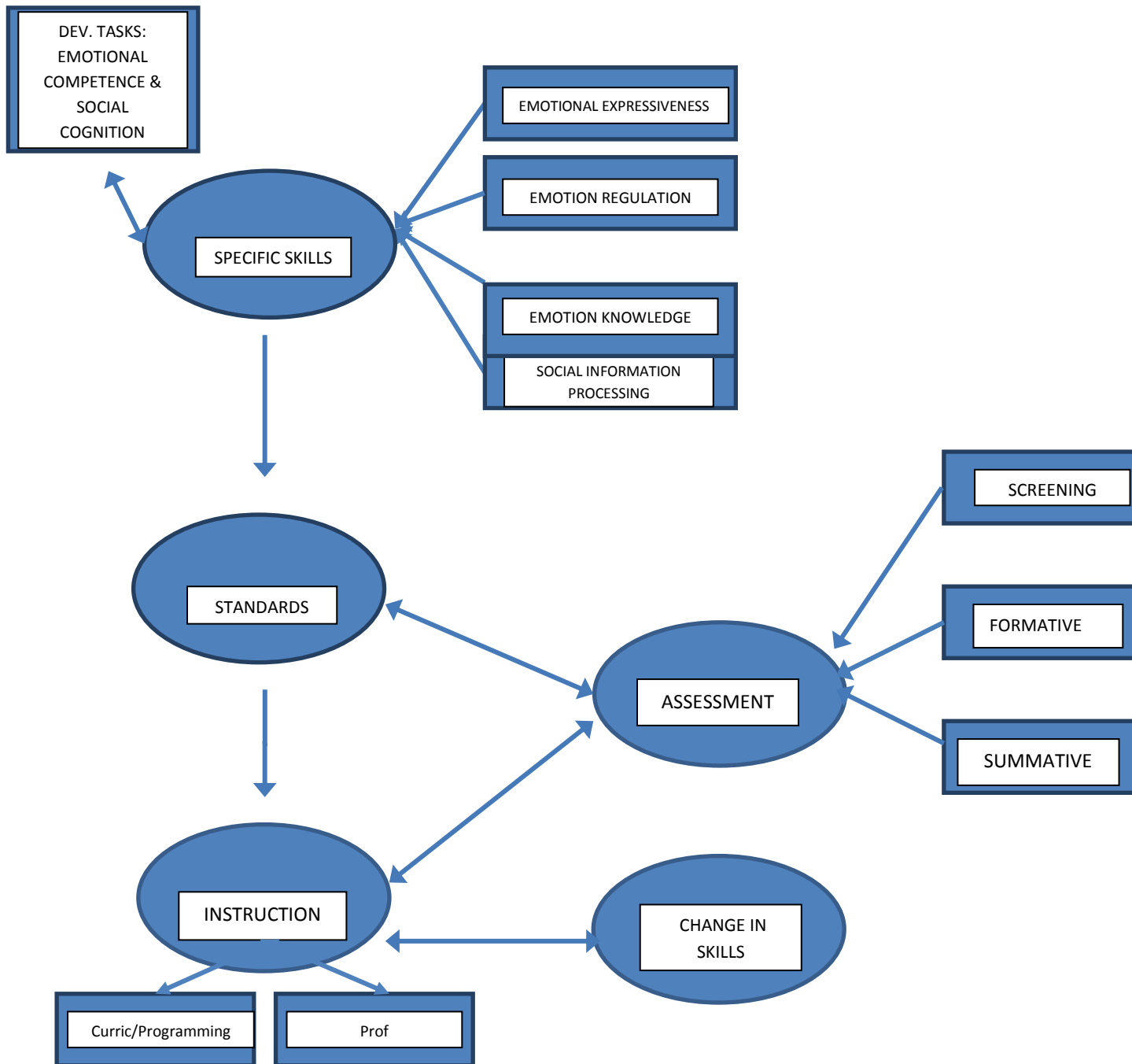


Table 1. Plan for Use of Assessment within a Larger Focus on Emotional Competence and Social Cognition

1. Annual completion of screening instruments by parents
2. Annual/semi-annual screening by teachers
3. Assessment of milestones and skills by teachers
4. Utilization of social-emotional curriculum that focuses upon elements of emotional competence and social cognition (e.g., Preschool PATHS; Domitrovich, Cortes, & Greenberg, 2007)
5. Implementation of parent education in this area (e.g., Havighurst, Wilson, Harley, & Prior, 2009)
6. Diagnostic workup for children identified as having difficulties

-Adapted from Brassard and Boehm (2007).

Table 2. Screening Measures

Screener	Strengths and Weaknesses
Reporter: Parent Ages and Stages Questionnaire: Social-Emotional (ASQ-SE) ^a	<ul style="list-style-type: none"> • Eight questionnaires for ages 6 to 60 months (19 to 33 items) • Addresses Regulation (often emotional), Compliance, Communication (about internal states), Adaptive Functioning, Autonomy, Affect, Interactions with People • Thus, does not address emotion knowledge except for internal state language or social problem-solving – does assess emotional expressiveness and regulation • Readable, age-appropriate • Easy to score • Good-to-excellent psychometrics • Spanish version
Reporter: Teacher Social Competence and Behavior Evaluation (SCBE-30) ^b	<ul style="list-style-type: none"> • 30-item short form • Describes behavior in context • Assesses Anger/Aggression, Anxiety/Withdrawal, Cooperation/Sensitivity • Does not address emotion knowledge or social problem-solving • Excellent norming and psychometrics • Spanish version
Reporter: Teacher or Parent Social Skills Improvement System Rating Scales (SSIS-RS) ^c	<ul style="list-style-type: none"> • Cooperation, Assertion, and Self-control scales; parent version also includes Responsibility and Empathy scales • Includes norming information • 15 – 25 minutes to complete • Does not directly address emotional expressiveness or emotion knowledge • Good psychometric properties • Direct link to intervention

^a Squires, Bricker, Heo, & Twombly (2001)^b LaFreniere & Dumas (1996)^c Gresham & Elliot (2008)

Table 3. Skills Assessment (Generally Formative)

	Ratings – Screening or Skill Assessment		Direct Assessment		Observation	
	Measure	Strengths/Weaknesses	Measure	Strengths/Weaknesses	Measure	Strengths/Weaknesses
Expressiveness	0 – 3 Infant/Early Childhood Behavior Questionnaires ^a (IBQ, ECBQ)	<ul style="list-style-type: none"> • Very short versions available • Psychometric properties good-to-excellent • Anger/Frustration, Sadness, Smiling/Laughter, 	None	---	None chosen	---
	3 – 5 Child Behavior Questionnaire (CBQ)	<ul style="list-style-type: none"> • Fear • Parent-report; also teacher-report for CBQ 	None	---	MPAC-R/S ^c	<ul style="list-style-type: none"> • 18 – 21 items • Positive Affect • Negative Affect • Trained observers watch children’s behaviors for four 5-minute intervals, noting the presence of items. • Reliability and validity adequate • Standardized training materials • Now computerized • Weakness is time needed for training; this is being streamlined and systematized for teacher use
	3 - 5 Positive and Negative Affect Scale ^b	<ul style="list-style-type: none"> • Parent Report • Excellent psychometrics • assesses how the child/adult “feels on average,” for 12 negative emotions (e.g., sad, angry) and 3 positive emotions (e.g., excited, enthusiastic). 				

Table 3 (continued)

	Ratings – Screening or Skill Assessment		Direct Assessment		Observation	
	Measure	Strengths/Weaknesses	Measure	Strengths/Weaknesses	Measure	Strengths/Weaknesses
Emotion Regulation	See Table 2; ASQ:SE	---	None	---	None	---
0 – 3						
3 - 5	Emotion Regulation Checklist ^d	<ul style="list-style-type: none"> • 24-items • Taps both prevalent emotional expressiveness and regulation, including affect lability, intensity, valence, flexibility, and contextual appropriateness of expressiveness • Good psychometrics but conflates expressiveness and regulation 	None selected	<ul style="list-style-type: none"> • Some direct assessments purport to evaluate emotion regulation, but these either are insufficient, still under development, or require too much training to be useful 	MPAC-R/S	See above

Table 3 (continued)

	Ratings – Screening or Skill Assessment		Direct Assessment		Observation	
	Measure	Strengths/Weaknesses	Measure	Strengths/Weaknesses	Measure	Strengths/Weaknesses
Emotion Knowledge 0-3	None	--	None	---	None	---
3 - 5	None	---	Affect Knowledge Test (AKT) ^e	<ul style="list-style-type: none"> • Widely used, large research base • Good reliability and validity • Minimizes language requirement • Ecologically valid and fun via play • Two parallel shortened version available; 10 minutes and 15 items • Computerized version under development; needs no manual scoring or coding 	None	---
			Emotion Matching Test ^f	<ul style="list-style-type: none"> • Photograph alternative to AKT • Children match emotion expressions with expressions of the same category, with situations or causes, and with spoken emotion labels. • Children also produce emotion Labels • Good psychometric properties • 15 minutes • Not much history but appears promising 		

Table 3 (continued)

	Ratings – Screening or Skill Assessment		Direct Assessment		Observation	
	Measure	Strengths/Weaknesses	Measure	Strengths/Weaknesses	Measure	Strengths/Weaknesses
Social Problem-Solving 0 – 3	None	---	None	---	None	---
3 - 5	SSIS-RS	See Table 2	Challenging Situations Task (CST) ^g	<ul style="list-style-type: none"> • Takes 10 min; two parallel 6-story versions • Pictorial vignettes on peer provocation, both overt and relational; children indicate how they would feel (happy, sad, angry, or just ok) & what they would do (socially competent, aggressive, dysregulated, and passive) • Portable, dev. appropriate • Coding not needed; • Earlier age than most other social cognitive instruments • Includes emotional aspect of social information processes • Adequate reliability, good validity • Computerized version under development; needs no manual scoring or coding 	None	---
			Schultz Test of Emotion Processing-Preliminary	<ul style="list-style-type: none"> • Video vignettes • Assesses Cue Interpretation, Response Access, and Response Decision • Psychometric properties good • Needs to be streamlined for use as an indicator 		

			Version ^h			
--	--	--	----------------------	--	--	--

^a e.g., Rothbart & Hwang (2002), Rothbart, Ahadi, Hershey, & Fisher (2001)

^b Watson, Clark, & Tellegen (1988)

^c Minnesota Preschool Affect Checklist-Revised/Shortened (MPAC-R/S; Denham et al., 2012b)

^d Shields and Cicchetti (1997)

^e Denham et al. (2011)

^f Morgan, Izard, and King (2010)

^g Denham et al. (in press); Denham, Way, et al. (2012)

^h Schultz et al. (2010)

Table 4. Examples of MPAC-R/S and ERQ Items

MPAC-R/S “Mega”-Scales	Exemplars of behaviors observed
Expression and regulation of positive affect	<ul style="list-style-type: none"> • Displays positive affect in any manner—facial, vocal, bodily
Expression and regulation of negative affect	<ul style="list-style-type: none"> • Uses negative affect which in social interaction
Productive involvement in purposeful activity	<ul style="list-style-type: none"> • Engrossed, absorbed, intensely involved in activity • Independent—involved in an activity that the child organizes for himself
Unproductive, unfocused use of personal energy	<ul style="list-style-type: none"> • Wandering • Listless
Lapses in impulse control (negative reactions to frustration)	<ul style="list-style-type: none"> • Context-related, physical, interpersonal aggression
Positive reactions to frustration	<ul style="list-style-type: none"> • Promptly expresses, in words, feelings arising from problem situation, then moves on
Skills in peer leading and joining	<ul style="list-style-type: none"> • Successful leadership • Smoothly approaches an already ongoing activity
Isolation	<ul style="list-style-type: none"> • No social interaction continuously for 3 minutes or more
Hostility/Aggression	<ul style="list-style-type: none"> • Unprovoked, physical, interpersonal aggression
Prosocial response to needs of others	<ul style="list-style-type: none"> • Taking turns • Cooperating • Sharing

Emotion Regulation Checklist Scale	Example Items
Lability/Negativity	Exhibits wide mood swings; is easily frustrated; is prone to angry outbursts
Emotion Regulation	Is a cheerful child; responds positively to neutral or friendly overtures by adults; can say when s/he is feeling sad, angry or mad, fearful or afraid

Comments on the Measurement of Social and Emotional Development in Early Childhood (sub-focus: Behavior problems)

Susan B. Campbell

Department of Psychology, University of Pittsburgh

Introductory Comments and Conceptual Framework

Under the broad rubric of social-emotional development in early childhood a number of constructs are important to consider. These include both positive aspects of development including emotion regulation and social competence, and less adaptive aspects reflected in behavior problems, which are partly defined by *poor* emotion regulation and *lower* levels of social competence. Although my charge is to focus on behavior problems in children 0-5, these can only be meaningfully assessed within the broader context of other child characteristics (e.g., language and cognitive development), 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 (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 must also change and be changed by the developing child. The ecological model places these dynamic developmental and family processes in the broader context of the neighborhood, community, and culture in which the child and family reside.

In addition, a developmental perspective on emerging social competence and behavior problems must also take into account the profound changes in cognitive, social, 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 stage salient tasks and transitions that characterize typical development during infancy and early childhood (Cicchetti, 1990; Sroufe, 1990). During infancy these include regulating impulses, communicating wants and needs, and establishing attachment relationships with primary 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). Thus, behavior problems and social competence must both 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 to meet their rapidly changing developmental needs, and when signs of problems are often unstable.

A policy perspective on measuring behavior problems in young children

The measurement of behavior problems in infants and preschoolers is important 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 and later outcomes have been substantiated in numerous studies over the past several decades. Thus, early identification and early intervention are important goals to pursue.

At the same time, some problems that are 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. 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 will provide only incomplete and potentially inaccurate information on the prevalence of serious behavior problems, as they will not be able to differentiate between transient adjustment difficulties and more persistent and severe problems.

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. 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.

Furthermore, we know unequivocally that poor prenatal care and pediatric care, chronic poverty, poor nutrition, high levels of family stress, 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 et al., 1998; NICHD Early Child Care Research Network [ECCRN], 2004, 2005; Shaw, Hyde, & Brennan, 2012; Shonkoff & Phillips, 2000). Thus, research needs to focus on how we help families at risk to support young children's development by building on child and parenting competence (e.g., Love, Chazen-Cohen, Raikes, & Brooks-Gunn, 2013; Shaw et al., 2006).

What do we mean by behavior problems in children 0-5?

I would argue that the concept of behavior problems in children prior to the second birthday, except in very extreme circumstances, is not meaningful. Behavior problems in children from ages 2 to 5 raise different issues, although the caveats discussed above apply; thus, both the defining constructs and measurement need to be different.

Adjustment Indicators in Infancy (0-24 months). 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, 1987; Sroufe, 1997). 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 *infant temperament*, assessing individual differences in reactivity and regulation that are the building blocks of personality (Rothbart, 2007). Infant temperament, measured during the second half of the first year, predicts later behavior problems in some children, but research also indicates that high levels of negative affect and poor self-regulation are more likely to predict later difficulties in the context of insensitive or overly harsh parenting and that temperamental difficultness can be moderated by parenting that is a good fit with children's needs (e.g., Belsky et al., 1998; Kochanska et al., 2009).

Measurement. Measures of infant temperament are readily available, albeit time-consuming to complete. The most widely used is the Infant Behavior Questionnaire (IBQ, Gartstein & Rothbart, 2003) which is listed on the Childtrends spreadsheet. A short form of the IBQ (37 items) is now available from the authors (Gartstein, personal communication, March

16, 2013) and seems appropriate for a wide survey study. 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, 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.

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, children are also going through a major developmental transition marked by struggles over autonomy versus dependence that may be reflected in transient aggression toward peers, difficulty sharing 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 (Campbell et al., 2000). Moreover, problems are more likely to become chronic in the context of less positive, engaged and proactive parenting and other indicators of family risk (e.g., 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.

The developmental course of physical aggression clearly illustrates the importance of using longitudinal data and considering problem behaviors in developmental context. Tremblay (2000) has demonstrated that physical aggression peaks at age two and declines rapidly as children use words instead of physical means to make their wants and needs understood (Dionne et al., 2003). In a similar vein, in the NICHD Study of Early Child Care, we examined trajectories of physical aggression from ages 2 to 9 (NICHD ECCRN, 2004). Maternal ratings of aggression toward others and destruction of property were highest at 24 months and showed a linear decline with age in the sample as a whole ($n = 1195$). However, 3% of the sample, mostly boys, exhibited high and stable aggression that was first evident at age 2 and persisted through age 9; high aggression over time also predicted adjustment problems through age 12 (Campbell et al., 2006). Moreover, membership in this high trajectory group was predicted by family and parenting risk including low income, low maternal education, maternal depression, and less engaged and stimulating parenting from infancy through 24 months.

Thus, behavior problems can be identified in young children, but 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 and even beyond. Behavior problems that are identified early and persist across the preschool period are especially likely to signal risk for ongoing difficulties into middle childhood.

Based on these considerations, measurement of social and emotional development in toddlers (2-3) and preschoolers (3-5) should include dimensions of social competence as well as problem behavior. Standardized measures are available to assess these constructs.

Measurement

Social Competence: The Social Skills Rating System (SSRS, Gresham & Elliott, 1991) is the most comprehensive and best standardized measure of social competence available and it is listed on the resource guide from Childtrends, so I will not reiterate information about this measure. In the NICHD Study of Early Child Care we used the SSRS starting at age 5, and we used only the social competence items, not the behavior problem items which are redundant with the Child Behavior Checklist (CBCL, see below). The SSRS has now been standardized on children, ages 3 and up, so it is a useful measure across the 3-5 year age range.

In the NICHD Study of Early Child Care, we used the Adaptive Social Behavior Inventory (ASBI, Hogan, Scott, & Bauer, 1992) at 24 and 36 months to assess social competence (self-expression and compliance) according to both mother and teacher reports. The Express Scale (13 items, e.g., understands others' feelings) had good alpha levels (.77-.84) based on maternal and teacher ratings at both ages; similarly, the Comply Scale (10 items, e.g., is helpful to other children; shares) showed good internal consistency (alphas from .76-.87). The other items (7) that assess disruptive behavior showed lower alphas and overlap with the CBCL to some degree (see below), so they are not likely to be useful. This measure was standardized in the 1980's, so the better standardization sample and the broader age coverage make the SSRS a better choice.

Behavior Problems: The CBCL (Achenbach & Rescorla, 2000) is the most widely used and best standardized measure of behavior problems available; different versions are appropriate for ages 2-5 and for older children. It is detailed on the Childtrends resource guide. Although it is long (about 100 items), the CBCL has been translated into multiple languages and it has extensive normative data by age and gender. It was standardized on a relatively diverse sample. It also has a good deal of data attesting to the predictive validity of elevated aggression and externalizing scales. Finally, the data from any survey using the CBCL can be compared with many other data sets, given its wide use in studies of children's problems in the US and internationally. Other measures are available, but they cannot begin to compete with the CBCL

in terms of comparative data across cultures, ages, and a variety of risk factors. In addition, the CBCL has been used extensively in longitudinal studies.

Other measures are shorter, such as the Strengths and Difficulties Questionnaire (SDQ, Goodman, 1999; www.sdqinfo.org/). The SDQ is a useful screening measure, but it is only normed on children ages 5 and older and the standardization was conducted in the United Kingdom. It consists of five 5-item scales assessing emotional problems, conduct problems, hyperactivity-inattention, peer relationship problems, and prosocial behavior. It has been translated in many languages and is available at no charge from Dr. Robert Goodman. However, it lacks broader age coverage and complete standardization data.

The Devereux Early Childhood Assessment Form looks promising partly because it assesses positive adjustment (initiative, self-control, attachment) as well as problems and it taps appropriate dimensions of behavior problems for this age range (attention problems, aggression, withdrawal-depression, and emotional control problems). It is also shorter than the CBCL. However, it has not been used in large-scale survey studies and it may not be available in languages other than English. It also does not have the wealth of construct and predictive validity data that the CBCL provides.

Overall, to measure social competence and problem behavior in toddlers and preschoolers, the SSRS and CBCL provide standardized and well-validated measures. Their length is a major drawback. Possibly, items from specific scales of interest, such as aggression, attention problems, and rule-breaking behavior on the CBCL and cooperation, empathy, and self-control from the SSRS can be extracted in order to assess key constructs, using shorter measures. Presumably this would require approval from the developers of these measures.

Other considerations

Among the criteria to consider in selecting or adapting existing measures are their age-appropriateness, cultural sensitivity, availability in languages other than English, short-term stability and predictive validity.

Summary of main points:

- To understand the emergence of social competence and behavior problems, behaviors need to be considered from a developmental perspective based on the changing developmental tasks and accomplishments of infancy, toddlerhood and the preschool period.
- The importance of the child's family and social environment cannot be under-estimated.
- Distinct, albeit overlapping, constructs best capture individual differences in adjustment and adaptation in infancy and the preschool period: temperamental variation is best

assessed in infancy and social competence and problem behavior are best assessed at preschool age.

- When considering the assessment of problem behavior in young children, it would be ideal to obtain longitudinal data so that transient and age-related problems can be differentiated from more chronic problems that merit intervention.

References

- Achenbach, T. & Rescorla, L. (2000). *Manual of the ASEBA Preschool Forms and Profiles*. Burlington, VT.: University of Vermont, Research Center for Children, Youth, and Families.
- 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.
- Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist, 32*, 513-531.
- Campbell, S.B. (2002). *Behavior problems in preschool children: Clinical and developmental issues, 2nd edition*. New York: Guilford Press.
- 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.
- Campbell, S. B., Spieker, S., Burchinal, M., Poe, M., and the NICHD Early Child Care Research Network. (2006). Trajectories of aggression from toddlerhood to age 9 predict academic and social functioning through age 12. *Journal of Child Psychology and Psychiatry, 47*, 791-800.
- Cicchetti, D. (1990). Perspectives on the interface between normal and atypical development. *Development and Psychopathology, 2*, 329-334.
- Cummings, E. M., Davies, P., & Campbell, S. B. (2000). *Developmental Psychopathology and Family Process: Research, Theory, and Clinical Implications*. New York: Guilford Press.
- 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.
- Dionne, G., Tremblay, R., Boivin, M., Laplante, D., & Perusse, D. (2003). Physical aggression and expressive vocabulary in 19-month-old twins. *Developmental Psychology, 39*, 261-273.
- Gartstein, M. A. (2013). Personal Communication. Department of Psychology, Washington State University, Pullman, WA.
- Gartstein, M. A. & Rothbart, M. K. (2002). Studying infant temperament via the Revised Infant Behavior Questionnaire. *Infant Behavior and Development, 26*, 64-86.
- Gresham, F.M. & Elliott, S.N. (1991). *Social Skills Rating System*. Circle Pines, MN: American Guidance Service.

- Goodman, R. (1999). The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. *Journal of Child Psychology and Psychiatry, 40*, 791-799.
- 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.
- 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.
- 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.
- 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.
- Rothbart, M.K. (2007). Temperament, development, and personality. *Current Directions in Psychological Science, 16*, 207-212.
- Sameroff, A.J (2009). *The transactional model of development: How children and contexts shape each other*. Washington, DC: American Psychological Association.
- 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., Hyde, L. W., & Brennan, L.M. (2012). Early predictors of boys' antisocial trajectories. *Development and Psychopathology, 24*, 871- 888.
- Shonkoff, J. & Phillips, D. (2000). *From neurons to neighborhoods*. Washington, DC: National Academy Press.
- 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.

Tremblay, R. E. (2000). The development of aggressive behavior during childhood: What have we learned in the past century? *International Journal of Behavioral Development*, 24, 129-141.

Recommendation for Measurement of *Self-Regulation* in Early Childhood

C. Cybele Raver

New York University



NEW YORK UNIVERSITY

A private university in the public service

Office of the Provost

Elmer Holmes Bobst Library
70 Washington Square South, Room 1230
New York, New York 10012-1091
Telephone: 212 998 2274
Facsimile: 212 995 4300
Cybele.Raver@nyu.edu

C. Cybele Raver

Vice Provost for Academic, Faculty and Research Affairs

To: Federal Interagency Forum on Child and Family Statistics and Child Trends

From: C. Cybele Raver

Subject: Recommendation for measurement of self-regulation in early childhood.

Date: April 4th, 2013

Rationale for inclusion of self-regulation at the construct level:

It is clear from the last decade of neuroscience and developmental science research that self-regulation has major implications for children's opportunities to learn. Cutting-edge neuroscience research highlights self-regulation *at the individual level*, where key areas of the brain (including the amygdala, the anterior cingulate, and the prefrontal cortex) are implicated in "top down" processes that support children's attention, working memory, and cognitive control; these "top down" processes work conjointly with "bottom up" processes involved managing emotion and behavioral self-control (see Blair & Raver, 2012; Ursache, Blair, & Raver, 2012). Two decades of research in developmental science highlight ways that children's competent self-regulation is prospectively associated with a large number of later positive outcomes in school settings, including higher social competence, higher academic competence, and lower levels of proneness to externalizing and internalizing problems. Conversely, children's cognitive and emotional dysregulation represent early indicators of emerging behavioral problems for some children, with deleterious consequences in both the shorter and longer term (Greenberg, Kusche, & Speltz, 1991; Webster-Stratton & Taylor, 2001).

In addition, self-regulation functions to support or constrain children's opportunities for learning *at the classroom level*. For example, teachers may simply be able to spend less time on

instruction when they experience trouble maintaining order and focused attention in the classroom. Children may also have more trouble attending to cognitively challenging tasks and teachers' instruction when their attention is pulled to peers' misbehavior and emotional distress (see Raver, Blair, & Li-Grining, 2012, for review). Chaotic classroom contexts may not only result from children's self-regulatory difficulty, but may exacerbate it: Recent research from child care studies suggests that the stressful nature of low-quality care may get "under the skin" for some children, altering children's regulatory set points (Watanura, Phillips, Morrissey, McCartney, & Bub, 2011; Ursache et al., 2012). On the other hand, positive classrooms that are well-staffed, well-managed, and cognitively engaging have been found (in both experimental and correlational studies) to support higher self-regulation and greater academic gain among low-income children (Bierman, 2011; Morris et al., 2013; Raver et al., 2011).

Rationale for inclusion of precise measurement of children's self-regulation:

Given the importance of self-regulation for early learning and health, how do we measure it? In my view, the fields of applied developmental, prevention, and education sciences have made major methodological strides in the last several years by developing new and highly innovative research designs (such as the use of cluster-randomized control trials, or RCT) to strengthen our capacity to make causal inferences about what works in improving children's educational and health outcomes. However, we have focused less on increasing the precision of measurement of key mediators and child outcomes such as self-regulation (see Raver et al., 2012 for discussion, as well as Shadish, Cook, & Campbell, 2002). This may be because social scientists have not, until recently, had ready access to standardized and "portable" measures of children's self-regulatory skills that have been extensively piloted or validated outside of university research laboratories. In a recent chapter on this topic, my colleagues and I have argued that, although young children's self-regulation has increasingly been identified as an important predictor of their ability to navigate the social and academic worlds of early schooling, there have been few tools available to use to directly assess these key skills (see Raver et al., 2012).

As just one example, results from recent meta-analyses of Head Start impacts clearly indicate that the "yardstick" used for measurement of the benefits of the program matters (Shager et al., 2013). That is, higher quality direct assessments with greater precision for capturing children's self-regulation are likely to maximize the chance of empirically detecting whether educational contexts have a significant impact on both cognitive and emotional domains of children's self-regulation. However, it has not been until recently that measures of children's self-regulation have been included in most educational evaluations (see Bierman, Nix, Greenberg, Blair, & Domitrovich, 2008; Diamond, Barnett, Thomas, & Munro, 2007; and Morris, Raver, Millensky, Jones, & Lloyd, 2010 as exceptions).

Recommendations for measurement of self-regulation:

The matrix that was submitted for our review represents an excellent starting point for considering multiple methods of measurement of self-regulation. These different measures can be categorized or "binned" into categories of *I. Adult Report* and *II. Direct Assessment*, and in domains that include A) social skills measures, B) measures of externalizing and internalizing

behavior problems, C) emotional regulatory skills, and D) executive function skills.

To address this need for more precise, direct assessment of children's self-regulation, I recommend that surveys include at a minimum, 1 adult report of each of the 4 types of socioemotional skills (A through D). You have listed a wide array of different types of measures that cover this empirical "waterfront," and future teams will have the expertise and past experiences of both successes and challenges faced by recent research teams to draw from, as they make selections from among those measures. My research team and I have found that reports from the assessor or research team member (in our case, on a measure called the Preschool Self-Regulation Assessment- Assessor Report (or PSRA-AR) filled out directly after cognitive and behavioral assessments such as the PPVT have been completed) have been invaluable. Measures such as the PSRA-AR (or other measures filled out by assessors, home visitors, or research staff) are relatively inexpensive to collect, reasonably to highly reliable, offer substantial predictive and criterion validity, and most critically, represent an independent empirical "window" into children's behavioral profiles that complements the perspectives of parents and teachers (Smith-Donald, Raver, Hayes, & Richardson, 2007).

Inclusion of Direct Assessment of Executive Function and Emotion Regulation:

If possible, I strongly support the inclusion of direct assessment of domains C (emotion regulation, or ER) and D (executive function, EF). In our research team's experience, we have had success in adapting direct assessments for field-based survey use, drawing from lab-based measures that have a strong "track record" in producing valid and reliable, finely-grained behavioral data. Our work suggests that direct assessment of ER and EF is feasible with large samples of low-income children in classroom-based settings. Moreover, inclusion of those measures has arguably yielded "value added" to both prevention science and basic developmental science: Inclusion of those measures has allowed our team to test for a) impact of intervention on underlying developmental processes that are central for later adjustment and academic outcomes, while also supporting b) theoretically-driven tests of developmental processes and mechanisms among at-risk samples of low-income children. As mentioned earlier, our inclusion of EF measures allowed us to demonstrate that early intervention at the classroom-based level significantly supported low-income children's school readiness in Head Start centers in urban neighborhoods of concentrated disadvantage (Raver et al., 2011). *In addition*, our recent re-analysis of CSRP data allowed us to demonstrate the deleterious impact of local homicides on children's attention and impulsivity, because we were able to capitalize on inclusion of those EF-related outcomes in pre- and post-test CSRP data collection (see Sharkey, Tirado-Strayer, Papachristos, & Raver., 2012 for example). Recent analyses of those same data have yielded convincing evidence of the deleterious consequences of household instability (including high numbers of moves) for young CSRP-enrolled children's effortful control, a skill related to EF (McCoy & Raver, resubmitted).

One question is whether measures are available for field-based use. The answer is yes: For example, you have listed several excellent direct assessments of children's executive function in the matrix of measures that was provided (measures 25 through 30), citing Phil Zelazo's, Mike Willoughby's and Clancy Blair's recent validation work on those measures. These

measures have excellent potential for yielding high empirical “payoff” in demonstrating possible benefits of preschool interventions for low-income children in early educational settings, as has been demonstrated by my own and others’ impact evaluations of classroom-based interventions (see Raver, 2012 for review). Children’s executive function has consistently been found in both correlational and experimental research to be malleable (i.e. to show change as a function of exposure to supportive vs. stressful environments), with long-term implications for children’s opportunities to learn in early educational settings.

While measures of executive function (domain D) have received the bulk of recent attention in recent reports of the benefits of early intervention, it’s important to note that emotion regulation may be equally important as a target of prevention science (and therefore equally important to measure), particularly for vulnerable, at-risk populations. You have listed, for example, a direct assessment of children’s knowledge of different emotional states developed by Denham and colleagues (measure #20, in your large list). My work as a team leader on measurement selection for the large, multi-site randomized trial of preschool interventions (Project CARES) led by Pamela Morris and MDRC suggests that those emotional regulatory skills are often the primary target of many “SEL” preschool curricula, and are likely to serve as the most proximal, directly targeted outcomes. (We adapted additional measures of children’s emotion knowledge, based on work by Celene Domitrovich and Karen Bierman, that might be of use in future measurement effort). Our team members (as well as other teams) are currently adapting measures that are more commonly part of affective neuroscience methods used in research labs, for use with school-aged children in field-based settings. We anticipate that these measures will be of high utility in testing hypotheses regarding children’s exposure to “toxic stress” in both home and school contexts and their ability to attend to, interpret, and respond to emotionally salient social cues in peer and teacher-student interactions (see Raver, 2012 and Blair & Raver, 2012).

While measures of young children’s emotion regulation may seem less immediately relevant to educational outcomes such as academic achievement, they are very salient for long-term measures of health and psychological well-being. Some of the most compelling new work in my field is on the etiology and sequelae of depression and anxiety among at-risk children. Children’s interpretation of emotional information and their ability to use higher-order cognitive processing when coping with difficulties in modulating or regulating negative emotion represents an innovative, and cutting-edge area of research that will benefit from better measurement, in larger field-based studies.

Finally, it is important to note the importance of measurement equivalence across all measures to be selected for assessment of children’s self-regulation. I strongly endorse development, piloting, and use of measures that have demonstrated measurement equivalence across groups of girls and boys, across younger and older children, and across groups differing by race/ethnicity. This has now become more standard practice in program evaluation as well as in basic developmental research. That said, it is worth noting that the groups across which measures must demonstrate psychometric robustness have changed as the U.S. population changes. For example, language minority status and recency of immigration have become

increasingly important as factors to consider as teams select measures for inclusion in large survey studies and program evaluations.

Conclusion:

Twenty years ago, our field was arguably less clear on what exactly we meant when we discussed the importance of children's self-regulation. In the last two decades, remarkable strides have been made across diverse areas of developmental neuroscience, developmental psychology, clinical psychology, and prevention science on the measurement of discrete domains of children's social skills, internalizing and externalizing problems, emotion regulation, and executive function. We are now well-positioned to capitalize on a range of precise, reliable, and valid measures of those domains for future research in education, health, and child welfare. In so doing, we will strengthen the science that supports evidence-based decision-making, as well as our understanding of basic cognitive and emotional processes in human development,

References

- 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. doi: [10.1016/j.appdev.2011.04.004](https://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. doi:10.1017/S0954579408000394
- Blair, C. & Raver, C. C. (2012). Child development in the context of adversity: Experiential canalization of brain and behavior. *American Psychologist, 67*(4), 309-318. doi: [10.1037/a0027493](https://doi.org/10.1037/a0027493)
- Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318*(5855), 1387-1388. doi: [10.1126/science.1151148](https://doi.org/10.1126/science.1151148)
- Greenberg, M. T., Kusche, C. A., & Speltz, M. (1991). Emotional regulation, self-control and psychopathology: The role of relationships in early childhood. In D. Cicchetti & S. Toth (Eds.), *Rochester Symposium on Developmental Psychopathology* (Vol. 2), (pp. 21-56). New York: Cambridge University Press.
- McCoy, D. C., & Raver, C. C. (resubmitted). Household instability and child self-regulation: Quasi-experimental links for low-income children. *Developmental Psychology*.
- Morris, P., Lloyd, C. M., Millenky, M., Leacock, N., Raver, C. C., & Bangser, M. (2013). *Using classroom management to improve preschoolers' social and emotional skills: Final impact and implementation findings from the Foundations of Learning Demonstration in Newark and Chicago*. New York, NY: MDRC.
- Morris, P., Raver, C. C., Millenky, M., Jones, S. M., & Lloyd, C. (2010). *Making preschool more productive: How classroom management training can help teachers*. New York, NY: MDRC.
- Raver, C. C. (2012). Low-income children's self-regulation in the classroom: Scientific inquiry for social change. *American Psychologist, 67*(8), 681-689. doi: [10.1037/a0030085](https://doi.org/10.1037/a0030085)
- Raver, C. C., Blair, C., & Li-Grining, C. P. (2012). Extending models of emotional self-regulation to classroom settings: Implications for professional development. In C. Howes, B. Hamre, & R. Pianta (Eds.), *Effective early childhood professional development: Improving teacher practice and child outcomes*. Baltimore: Brookes.
- 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 as a mediating mechanism. *Child Development, 82*(1), 362-378. doi: [10.1111/j.1467-8624.2010.01561.x](https://doi.org/10.1111/j.1467-8624.2010.01561.x)
- Raver, C. C., Smith-Carter, J., McCoy, D. C., Roy, A., Ursache, A., & Friedman, A. H. (2012). Testing models of children's self-regulation within educational contexts: Implications for measurement. *Advances in Child Development and Behavior, 42*, 245-270.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton-Mifflin.

- Shager, H. M., Schindler, H. S., Magnuson, K. A., Duncan, G. J., Yoshikawa, H., & Hart, C. M. D. (2013). Can research design explain variation in Head Start research results? A meta-analysis of cognitive and achievement outcomes. *Education Evaluation and Policy Analysis, 35*(1), 76-95. doi:10.3102/0162373712462453
- Sharkey, P., Tirado-Strayer, N., Papachristos, A. & Raver, C. C. (2012). The effect of local violence on children's attention and impulse control. *American Journal of Public Health, 102*(12), 2287-2293. doi: 10.2105/AJPH.2012.300789
- 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*(2), 173-187. doi: [10.1016/j.ecresq.2007.01.002](https://doi.org/10.1016/j.ecresq.2007.01.002)
- 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. doi: 10.1111/j.1750-8606.2011.00209.x
- Watamura, S. E., Phillips, D. A., Morrissey, T. W., McCartney, K., & Bub, K. (2011). Double jeopardy: Poorer social-emotional outcomes for children in the NICHD SECCYD experiencing home and child-care environments that confer risk. *Child Development, 82*(1), 48-65. doi: 10.1111/j.1467-8624.2010.01540.x
- Webster-Stratton, C., & Taylor, T. (2001). Nipping early risk factors in the bud: Preventing substance abuse, delinquency, and violence in adolescence through interventions targeted at young children (0-8 years). *Prevention Science, 2*(3), 165-192. doi: 10.1023/A:1011510923900

Measurement of *Executive Function* in Early Childhood

Michael Willoughby

Frank Porter Graham Child Development Institute

MEMORANDUM

To: Child Trends, Federal Interagency Forum on Child & Family Statistics

From: Michael Willoughby, Ph.D.

Re: Measurement of Executive Function in Early Childhood

Date: March 18, 2013

Updated: November 25, 2013

Abstract

This memo provided a justification and rationale for the measurement of executive functions (EF) in future federal surveys that involve children age 0-5 years. It also summarized issues related to the selection of measurements including (1) challenges in interpretation that result from the low correlations between caregiver ratings versus child direct performance on EF tasks, (2) the lack of standardized norm-referenced tests that provide full coverage of EF abilities, and (3) limited psychometric information regarding measurement equivalence of EF tasks across time and subpopulations (e.g., gender, poverty, race/ethnicity). Despite an explosion of interest in EF across the life course, fundamental questions remain regarding the optimal way to conceptualize this construct. Future federal studies should adopt multi-informant and multi-method measurement of EF and would likely benefit from the selection of specific tasks and ratings that have demonstrated reliability and validity at the specific ages of interest.

Q1: Why is the measurement of executive function (EF) among children 0-5 important from a policy perspective?

EF refers to a set of domain general skills that facilitate novel problem solving and goal directed pursuits. EF undergoes substantial developmental change during the 0-5 age period. EF may represent one final common pathway through which a variety of risk factors—including poverty-related risks, traumatic brain injury, prematurity, and toxicant, alcohol or drug exposure—negatively impinge on child developmental outcomes, including school readiness. To the extent that this is true, the careful measurement of individual differences in EF from 0-5

years may inform policy-relevant interventions (e.g., the provision of Universal Pre-Kindergarten) that are directed towards children at risk.

Q2: What is the relationship of EF to subsequent developmental outcomes?

EF has been implicated as an important predictor of school readiness (Blair, 2002). Children who perform better on EF tasks also perform better on tests of academic achievement, and these associations hold even when pretest levels of academic achievement are included as a covariate in models relating early EF to later academic achievement (Bierman, Torres, Domitrovich, Welsh, & Gest, 2009; Brock, Rimm-Kaufman, Nathanson, & Grimm, 2009; Smith-Donald, Raver, Hayes, & Richardson, 2007; Thorell & Wahlstedt, 2006; Welsh, Nix, Blair, Bierman, & Nelson, 2010). Individual differences in EF appear to be more strongly related to math than reading achievement, which is theoretically interesting given the presumed involvement of the prefrontal cortex in both solving math problems and completing inhibitory control tasks (Blair & Razza, 2007; Bull, Espy, & Wiebe, 2008; Bull & Scerif, 2001; Espy et al., 2004).

In addition to academic outcomes, EF has long been implicated in developmental models of psychopathology, especially attention deficit/hyperactivity disorder (ADHD) and Autism (Kenworthy, Yerys, Anthony, & Wallace, 2008; Pennington & Ozonoff, 1996; Sonuga-Barke, Dalen, & Remington, 2003; Willcutt, Doyle, Nigg, Faraone, & Pennington, 2005). In both disorders, EF deficits appear to emerge early in life. Indeed, at least for ADHD, there have been suggestions that EF may serve as a possible endophenotype of the disorder, with the expectation that measuring “core” processes of the disorder may facilitate etiologic research (Doyle et al., 2005; Kebir, Tabbane, Sengupta, & Joober, 2009; Nigg, Blaskey, Stawicki, & Sachek, 2004).

Nonetheless, it is important to note that most of what is known about the contributions of EF to academic outcomes has been derived from studies that employed passive longitudinal designs and that did not consider a wide range of potential confounder variables. It is not clear whether and to what extent these associations might change if (un)measured confounders were considered (Willoughby, Kupersmidt, & Voegler-Lee, 2012). Moreover, although EF may be related to ADHD, it is clear that it is neither necessary nor sufficient for the disorder (Willcutt et al., 2005). The important point is that although EF has been implicated for multiple developmental outcomes, the current evidence base does not permit causal inferences.

Q3: What issues should be taken into account with regard to the measurement of EF from 0-5 years in diverse populations?

EF is a nebulous construct that lacks a definitive consensus definition. EF is often defined in broad terms to encompass all cognitive abilities that facilitate goal directed problem solving pursuits. For example, Anderson (2002) proposed an Executive Control System that consisted of four inter-related domains of functioning including cognitive flexibility, attentional control, goal setting, and information processing. This model is broader in scope than the typical measurement of EF in early childhood (e.g., goal setting is too advanced of a skill to assess among preschoolers; speed of processing is distinct from but correlated with typical measures

of EF). In early childhood, the assessment of EF has more commonly been understood to refer to individual differences in inhibitory control, working memory, and cognitive flexibility/attention shifting. Whereas these sub-dimensions are understood to reflect distinct but correlated constructs in middle childhood through adult aged samples (Lehto, Juujarvi, Kooistra, & Pulkkinen, 2003; Miyake et al., 2000), the available factor analytic evidence indicates that they are undifferentiated constructs in early childhood (Hughes, Ensor, Wilson, & Graham, 2010; Wiebe, Espy, & Charak, 2008; Willoughby, Blair, Wirth, Greenberg, & Investigators, 2010), though perhaps only in normative samples (Schoemaker et al., 2012). The important point is that the assessment of EF from age 0-5 years requires precise definitions that should constrain measurement decisions.

It is well established that caregiver (parent or teacher) ratings and direct assessments of EF correlate poorly, with $r_s \approx .20$ (Toplak, West, & Stanovich, 2013). In contrast to conventional measurement wisdom which posits that multiple methods/informants should help to “triangulate” on true EF ability (Campbell & Fiske, 1959), ratings and direct assessments of EF appear to measure different things. Barkley and colleagues have provocatively suggested that researchers and clinicians should abandon direct assessments of EF and rely exclusively on ratings of EF (Barkley, 2012; Barkley & Fischer, 2011; Barkley & Murphy, 2011). Although this represents an extreme and minority position, it underscores the importance of clarifying the method of EF assessment. The inclusion of both assessment approaches in future federal surveys would provide opportunities to clarify the relative merits of each assessment approach, as well as provide opportunities to test how these approaches may complement each other (e.g., by testing interactions between rated and directly assessed EF in the prediction of developmental outcomes).

In her synopsis and empirical comparison of 24 tasks that have been commonly used to measure EF in early childhood, Carlson (2005) observed that many tasks yield scores that exhibit “binary distributions” which correspond to pass/fail distinctions in ability but that mask individual differences in ability. Binary or bimodal score distributions likely result from a task having too few items and/or items that do not sufficiently vary in difficulty level for a given age (ability level). Future federal surveys that measure EF should ensure that any selected EF tasks exhibit good variability at the age(s) at which they will be used. Many tasks that may “work” well at one age will do poorly at other ages (e.g., due to floor or ceiling effects). To the extent that this is true, such tasks are not well suited for making inferences about individual differences in intra-individual change in EF during the early childhood period.

The most widely used EF tasks in the early childhood literature were developed by individual researchers. Although some norm-referenced standardized assessments of general cognitive ability that are appropriate for children 0-5 years of age include subtests that measure specific aspects of EF (e.g., NEPSY-II; WPPSI-IV), as Anderson and Reidy (2012) recently noted “There is currently no commercially available, norm-referenced test battery that enables a comprehensive assessment of executive function in preschoolers (p. 350)”. To be clear, this refers specifically to direct assessments of EF. There are norm-referenced measures of rated EF available for use with children 0-5 years old (Gioia, Espy, & Isquith, 2003). The lack of norm-

referenced scores is a limitation because the resulting scores can only be interpreted relative to the sample in which they are used.

A final issue to consider is whether the selected measures have been demonstrated to work with children from diverse (e.g., poverty, language, race/ethnicity) backgrounds. Two issues are relevant. First, the selected measures of EF should have undergone formal tests of measurement invariance. Measurement invariance refers to formal statistical tests of whether the psychometric properties of a given assessment are comparable across distinct groups. Mungas and colleagues provided an exemplar approach to testing measurement invariances of EF in a diverse sample of older adults (Mungas, Widaman, Reed, & Farias, 2011). Second, to the extent that large federal surveys would likely include children for who English is a second language, it would be important to determine whether the measures of interest have been translated into other languages.

Q4: What criteria are most important to consider in selecting/developing measures of EF?

The first criterion that should inform measure selection is whether direct assessments of EF are feasible or whether resource (including child assessment time) constraints limit the assessment of EF to caregiver or teacher ratings. Given the inherent limitations of assessing EF by a single approach, multi-method assessments will provide the most flexibility in terms of evaluating the utility of EF as an early developmental outcome or predictor of later outcomes.

The second criterion to inform measurement selection is whether the goal is to measure individual differences in EF ability at a target age versus characterizing individual differences in the trajectory of EF acquisition across time. It has only been recently that researchers have utilized prospective designs to characterize developmental trajectories of EF acquisition from 3-5 years (Wiebe, Sheffield, & Espy, 2012; Willoughby, Wirth, Blair, & Investigators, 2012). Due to the rapid changes in EF capacity/ability, it is not feasible to make inferences regarding the trajectories of ability across a broader age period because the tasks that work with very young children (i.e., 1-2 years) are typically not appropriate (too easy) for older children (i.e., 4-5 years). Given the current state of the literature, it is likely more appropriate/feasible to identify discrete ages at which EF will be measured and to select measurement tools with demonstrated reliability and validity at those particular ages.

A third criterion concerns the pragmatics of data collection and task scoring. Many widely used tasks involve idiosyncratic test materials or involve staff training requirements that may not be amenable to federal surveys. Computerized and norm-referenced assessments provide important standardization to direct assessments of EF, provide scores that have clear meaning, and in many cases benefit from software solutions for task/questionnaire scoring.

Q5: What additional guidance would you provide for the Federal Interagency Forum on Child and Family Statistics about developing an indicator from existing data sources, or in collecting new data on EF in early childhood?

I would discourage conflating caregiver ratings of attention, persistence, self-control, effortful control, etc. as indicators of EF. Although these constructs are important in their own right,

they will be highly correlated with other aspects of social behavior but not performance on EF tasks.

A distinction has been made between “hot” and “cool” EF tasks (Zelazo & Müller, 2002). Hot EF tasks require the resolution of novel problems that are emotionally arousing, including tasks with appetitive demands, and are understood to engage the orbitofrontal cortex (OFC), an area with strong connections to the limbic system (Happaney, Zelazo, & Stuss, 2004). In contrast, cool EF tasks require the resolution of novel problems that are emotionally neutral and are understood to engage the dorsolateral prefrontal (DL-PFC) cortex (Happaney et al., 2004). The distinction between hot and cool EF is consistent with similar distinctions made in reviews of the self-control (Metcalf & Mischel, 1999) and inhibition (Nigg, 2000) literatures. Studies of preschool-aged children indicate that whereas cool (but not hot) EF is uniquely associated with academic functioning, hot (but not cool) EF may be uniquely related to behavioral functioning (Brock et al., 2009; Kim, Nordling, Yoon, Boldt, & Kochanska, 2013; Willoughby, Kupersmidt, Voegler-Lee, & Bryant, 2011). The distinction between hot and cool EF tasks may be germane to future federal surveys that have interest in prediction of behavioral and academic aspects of school readiness. Nonetheless, the assessment protocols used to measure “hot” EF are even less well developed than those used to measure “cool” EF.

Supplementary Q: Are there any measures in current federal surveys that you feel are important to consider for use as an indicator in the America’s Children report? What are their strengths/weaknesses?

Assessor ratings of children’s behavior during direct assessments of EF (or more generally problem solving) tasks may serve as a “middle ground” between caregiver ratings of behavior and children’s performance on EF tasks. The LEITER items that were normed in FACES (measure 34 in the SE Measures Inventory that you provided) appear promising (Smith-Donald et al., 2007). I am not aware of EF measures for children 0-5 years old that have been used in current federal surveys.

Supplementary Q: Among measures that have been successfully used in non-federal national surveys or small scale studies, which do you recommend for inclusion in a federal survey? What are their strengths/weaknesses?

The NIH Toolbox represents a standard set of tasks that purportedly measure, among other things, EF from age 3-85 years (Weintraub et al., 2013). As noted by the authors,

The NIH-TB Cognition Battery is intended to serve as a brief, convenient set of measures to supplement other outcome measures in epidemiologic and longitudinal research and clinical trials. With a computerized format and national standardization, this battery will provide a “common currency” among researchers for comparisons across a wide range of studies and populations.

It will be important for future research to ensure that the EF tasks work sufficiently well with children 3-5 years.

I would advocate for supplemental ratings by parents and/or teachers of EF ability at the same ages. Although the BRIEF-P is among the mostly widely used measure (and benefits from the availability of norms), non-commercial instruments (that lack norms) also exist (e.g., CHEXI; Thorell & Nyberg, 2008).

Supplementary Q: Are there gaps in the measurement of EF that should be addressed through federal data collection efforts?

Confusion abounds regarding the appropriate interpretation of the poor agreement between ratings and direct assessments of EF. Moreover, children's performance on multiple direct assessments of EF is often poorly correlated within a single testing session. Future federal studies that included ratings and multiple, direct assessments of EF would provide an invaluable resource for resolving these discrepancies.

Supplementary Q: Are there publications or other resources that you recommend we refer to for this effort?

Two reviews provide a broad coverage of EF in early childhood (Anderson & Reidy, 2012; Garon, Bryson, & Smith, 2008).

References

- Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology, 8*(2), 71-82.
- Anderson, P. J., & Reidy, N. (2012). Assessing Executive Function in Preschoolers. *Neuropsychology Review, 22*(4), 345-360. doi: DOI 10.1007/s11065-012-9220-3
- Barkley, R. A. (2012). *Executive functions: What they are, how they work, and why they evolved*. New York: Guilford Press.
- Barkley, R. A., & Fischer, M. (2011). Predicting Impairment in Major Life Activities and Occupational Functioning in Hyperactive Children as Adults: Self-Reported Executive Function (EF) Deficits Versus EF Tests. *Developmental Neuropsychology, 36*(2), 137-161. doi: Pii 933894190
- Doi 10.1080/87565641.2010.549877
- 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. doi: DOI 10.1007/s10862-011-9217-x
- Bierman, K. L., Torres, M. M., Domitrovich, C. E., Welsh, J. A., & Gest, S. D. (2009). Behavioral and Cognitive Readiness for School: Cross-domain Associations for Children Attending Head Start. *Social Development, 18*(2), 305-323.
- Blair, C. (2002). School readiness - Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American Psychologist, 57*(2), 111-127.
- 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*(2), 647-663.
- Brock, L. L., Rimm-Kaufman, S. E., Nathanson, L., & Grimm, K. J. (2009). The contributions of 'hot' and 'cool' executive function to children's academic achievement, learning-related behaviors, and engagement in kindergarten. *Early Childhood Research Quarterly, 24*(3), 337-349.
- Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology, 33*(3), 205-228.
- Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology, 19*(3), 273-293.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 5*, 81-105.

- Carlson, S. A. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*(2), 595-616.
- Doyle, A. E., Willcutt, E. G., Seidman, L. J., Biederman, J., Chouinard, V. A., Silva, J., & Faraone, S. V. (2005). Attention-deficit/hyperactivity disorder endophenotypes. *Biological Psychiatry, 57*(11), 1324-1335.
- Espy, K. A., McDiarmid, M. M., Cwik, M. F., Stalets, M. M., Hamby, A., & Senn, T. E. (2004). The Contribution of Executive Functions to Emergent Mathematic Skills in Preschool Children. *Developmental Neuropsychology, 26*(1), 465-486.
- 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.
- 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.
- Happaney, K., Zelazo, P. D., & Stuss, D. T. (2004). Development of orbitofrontal function: Current themes and future directions. *Brain and Cognition, 55*(1), 1-10.
- 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.
- Kebir, O., Tabbane, K., Sengupta, S., & Joobor, R. (2009). Candidate genes and neuropsychological phenotypes in children with ADHD: review of association studies. *Journal of Psychiatry & Neuroscience, 34*(2), 88-101.
- Kenworthy, L., Yerys, B. E., Anthony, L. G., & Wallace, G. L. (2008). Understanding Executive Control in Autism Spectrum Disorders in the Lab and in the Real World. *Neuropsychology Review, 18*(4), 320-338.
- Kim, S., Nordling, J. K., Yoon, J. E., Boldt, L. J., & Kochanska, G. (2013). Effortful Control in "Hot" and "Cool" Tasks Differentially Predicts Children's Behavior Problems and Academic Performance. *Journal of Abnormal Child Psychology, 41*(1), 43-56. doi: DOI 10.1007/s10802-012-9661-4
- Lehto, J. E., Juujarvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology, 21*, 59-80.
- Metcalfe, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: Dynamics of willpower. *Psychological Review, 106*(1), 3-19.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49-100.
- Mungas, D., Widaman, K. F., Reed, B. R., & Farias, S. T. (2011). Measurement Invariance of Neuropsychological Tests in Diverse Older Persons. *Neuropsychology, 25*(2), 260-269. doi: Doi 10.1037/A0021090

- Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: Views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin*, *126*(2), 220-246.
- Nigg, J. T., Blaskey, L. G., Stawicki, J. A., & Sachek, J. (2004). Evaluating the endophenotype model of ADHD neuropsychological deficit: Results for parents and siblings of children with ADHD combined and inattentive subtypes. *Journal of Abnormal Psychology*, *113*(4), 614-625.
- Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry*, *37*(1), 51-87.
- 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. doi: DOI 10.1111/j.1469-7610.2011.02468.x
- 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*(2), 173-187.
- Sonuga-Barke, E. J. S., Dalen, L., & Remington, B. (2003). Do executive deficits and delay aversion make independent contributions to preschool attention-deficit/hyperactivity disorder symptoms? *Journal of the American Academy of Child and Adolescent Psychiatry*, *42*(11), 1335-1342.
- 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.
- Thorell, L. B., & Wahlstedt, C. (2006). Executive functioning deficits in relation to symptoms of ADHD and/or ODD in preschool children. *Infant and Child Development*, *15*(5), 503-518.
- 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. doi: Doi 10.1111/Jcpp.12001
- Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., . . . Gershon, R. C. (2013). Cognition assessment using the NIH Toolbox. *Neurology*, *80*(11 (Supplement 3)), S54-S64.
- Welsh, J. A., Nix, R. L., Blair, C., Bierman, K. L., & Nelson, K. E. (2010). The Development of Cognitive Skills and Gains in Academic School Readiness for Children From Low-Income Families. *Journal of Educational Psychology*, *102*(1), 43-53.
- 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. D., & Espy, K. A. (2012). Separating the Fish From the Sharks: A Longitudinal Study of Preschool Response Inhibition. *Child Development*, *83*(4), 1245-1261. doi: DOI 10.1111/j.1467-8624.2012.01765.x

- Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., & Pennington, B. F. (2005). Validity of the executive function theory of attention-deficit/hyperactivity disorder: A meta-analytic review. *Biological Psychiatry, 57*(11), 1336-1346.
- Willoughby, M., Kupersmidt, J., Voegler-Lee, M., & Bryant, D. (2011). Contributions of hot and cool self-regulation to preschool disruptive behavior and academic achievement. *Developmental Neuropsychology, 36*(2), 162-180.
- 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., Kupersmidt, J. B., & Voegler-Lee, M. E. (2012). Is Preshcool Executive Function Causally Related to Academic Achievement? *Child Neuropsychology, 18*(1), 79-91.
- Willoughby, M. T., Wirth, R. J., Blair, C. B., & Investigators, F. L. P. (2012). Executive function in early childhood: Longitudinal measurement invariance and developmental change. *Psychological Assessment, 24*(2), 418-431.
- Zelazo, P. D., & Müller, U. (2002). Executive function in typical and atypical development. In U. Goswami (Ed.), *Handbook of childhood cognitive development* (pp. 445-469). Oxford: Blackwell