



Assessing early childhood social and emotional development: Key conceptual and measurement issues



Stephanie M. Jones ^a, Martha Zaslow ^{b,c}, Kristen E. Darling-Churchill ^{c,*}, Tamara G. Halle ^c

^a Harvard University, USA

^b Society for Research in Child Development, USA

^c Child Trends, USA

ARTICLE INFO

Article history:

Received 3 July 2015

Received in revised form 24 February 2016

Accepted 24 February 2016

Available online 11 March 2016

Keywords:

Early childhood

Social and emotional development

Measurement

Constructs and subdomains

Conceptual clarity

ABSTRACT

This concluding paper takes stock of key issues that have emerged across the papers of this special issue on measuring early childhood social and emotional development, as well as in the broader literature anchoring these papers. While we include separate sections focusing on conceptual and measurement issues, it quickly becomes clear that the two are intertwined. The field lacks conceptual and definitional clarity, thus hindering our understanding of the focus and utility of specific measures. Indeed, our overarching conclusion is that the greatest progress will be made when measures of young children's social and emotional development are clearly mapped onto an agreed upon conceptual framework that both distinguishes social and emotional development from other broad domains (such as cognitive development), and includes carefully delineated and defined subdomains (broad areas within social and emotional development, such as emotional competence), constructs (specific aspects within subdomains, such as emotion knowledge), and corresponding behaviors.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The papers in this special issue focus both on conceptual and measurement issues in seeking to describe young children's social and emotional development. Key conceptual issues include the need to more clearly identify the borders of social and emotional development, those that separate it from other key domains (such as cognitive development), and to more clearly delineate and distinguish the subdomains of social and emotional development (such as emotional competence) and the specific constructs within these. Conceptual issues also include the need to more fully acknowledge that measures of social and emotional development reflect not only children's behaviors, skills and knowledge, but also features of the contexts in which children grow, learn, and play. Key methodological issues include the large and widely varying number of measures of social and emotional development identified by careful review, and the relatively small number with appropriate psychometric properties needed for administration in different types of studies. A dearth of measures that have not simply been translated but that have been standardized so as to be appropriate for important demographic subgroups has also been underscored as a central issue by the special issue papers.

While identifying serious conceptual as well as methodological issues, the papers in this special issue also point to promising new developments in the measurement of young children's social and emotional development. For example, the special issue papers point to the emergence of direct assessments in this domain, helping to address longstanding issues related to heavy reliance on parent and teacher report measures. They also note critical work towards developing measures appropriate for infancy and toddlerhood, a developmental period that until recently has had an important but limited set of measures.

The purpose of this concluding paper is to highlight and provide further discussion of what we see as some of the most important issues that have emerged across the papers of the special issue, and to identify the next steps that appear most important and potentially fruitful. While we have separate sections in this concluding paper focusing on conceptual and measurement issues, it quickly becomes clear that the two are intertwined, with the lack of conceptual and definitional clarity hindering our understanding of the target and utility of specific measures. Although we begin by focusing separately on the key conceptual and measurement issues that the special issue papers highlight, it is virtually inevitable that we must conclude with a discussion of how these intersect. Indeed, our overarching conclusion is that the greatest progress will be made when measures of young children's social and emotional development are clearly mapped onto an agreed upon conceptual framework that both distinguishes social and emotional development from other broad domains, and that includes carefully

* Corresponding author at: Child Trends, 7315 Wisconsin Avenue, Suite 1200W, Bethesda, MD 20814, USA.

E-mail address: kchurchill@childtrends.org (K.E. Darling-Churchill).

delineated and defined subdomains, constructs, and corresponding behaviors.

2. Conceptual issues

We begin our discussion of conceptual issues important to understanding young children's social and emotional development by returning to a statement from the Introduction to this special issue: "Despite this preponderance of evidence supporting the need to foster young children's positive social and emotional development, the development of psychometrically valid measures that are aligned for use **within assessment and accountability systems has lagged**" (Darling-Churchill & Lippman, this issue, p. XX). How is it that the last two decades have generated so much useful information about the importance of early social and emotional competencies for success across developmental periods, and yet we still have relatively few easy-to-use, psychometrically strong tools to measure children's discrete competencies, track them over time, and to act upon our knowledge with relevant strategies? In our view, one central reason, likely among many, lies squarely in how social and emotional competencies (skills, processes, outcomes, etc.) are defined, codified, and operationalized in research, and how that information is translated in the worlds of practice (e.g., as standards) and policy (e.g., as accountability systems).

To begin, as suggested by several of the commentary writers for this special issue, the area of social and emotional development suffers from what have been called the "jingle and jangle fallacies" (e.g., Borghans, Duckworth, Heckman, & ter Weel, 2008; Garcia, 2014). In short, the "jingle fallacy" refers to the use of a single term or names for constructs to represent a wide variety of skills, and the "jangle fallacy" refers to the use of different terms or construct names to refer to the same skill (Duckworth, Quinn, & Tsukayama, 2012; Jones & Bailey, *unpublished manuscript*; Reeves & Venator, 2014). The social and emotional domain is rife with these two challenges.

The use of the term "self-control" in the literature on social and emotional development provides a good example of this conceptual clutter. In an important and much cited paper, Moffitt and colleagues report **self-control in childhood to be a strong predictor of a variety of life outcomes** (Moffitt et al., 2011). In their paper, they describe self-control as "an umbrella construct that bridges concepts and measurements from different disciplines (e.g., impulsivity, conscientiousness, self-regulation, delay of gratification, inattention, hyperactivity, executive function, willpower, intertemporal choice)" (Moffitt et al., 2011, p. 2693). Each of the concepts and constructs under this umbrella holds its own distinct set of definitions, operationalizations, and measurement tools.

In another recent paper, in contrast to Moffitt et al., Diamond (2013) presents self-control as a component of self-regulation, and as a feature of the inhibitory control dimension of executive function. Indeed, in the broader literature, self-regulation itself is characterized by similar levels of complexity (e.g., Burman, Green & Shanker, 2015). As noted by Williford and Vick Whittaker (article 3 of this issue), the distinctions among emotion regulation and emotion expression, for example, and between behavioral regulation and aggression, as another example, are not always clear.

A further key complication is that the borders between social and emotional development and other important domains of young children's development are not always clear or agreed upon. The area of research on executive function (EF), as discussed in the paper by Willoughby (article 3 of this issue), suffers both from a lack of clarity on where EF fits in terms of domains of development, and also in terms of what specific skills comprise EF. Executive function has become prominent in the literature on young children's development over the last decade, and with good reason, as EF skills have been linked to a variety of positive outcomes for children (e.g., Blair & Razza, 2007; Carlson & Wang, 2007). The challenge is that, in both research and popular writing, the term EF has been used to refer exclusively to mental tasks

performed in a lab-like environment (e.g., computerized memory and attention tasks) as well as to a broader set of skills including how children manage their feelings, exercise self-control, and interact with peers or adults (e.g., delay gratification, listen and follow rules, cope with frustration).

In lumping all of these skills together and using the term EF to refer to any of them, the meaning of each and the specific research findings tied to it, are obscured (Jones, Bailey, & Partee, 2015). In fact, in a comprehensive review of the EF and self-regulation literatures, over 50 distinct construct terms were employed to represent and operationalize EF. In some cases different terms were used to refer to the same underlying phenomena (and deploying the same measurement tools to represent them) and in other cases the same terms represented different phenomena (and deployed different tools to represent them; Jones et al., 2015). As such, one could say that both the jingle and the jangle fallacies are represented in the EF literature.

It may be that some of this conceptual clutter makes sense given the rapid expansion of research in this area over the last two decades, and the intensive interest in the field from practice and policy circles. It may also be that a variety of terms are legitimately deployed across development to represent a core underlying process (e.g., self-control in children may be willpower in adults), and that developmental growth and change is truly a process of hierarchical reorganization and differentiation over time (Cicchetti & Rogosch, 2002). If this is the case, the result is a real overlap of skills, especially early in development, as also suggested by Willoughby (article 3 of this issue).

What are the implications of such definitional messiness for measurement? One direct consequence is that we may lose sight of the actual behavior or skill we think is important. For example, we might be seeking to develop a set of early childhood standards focused on essential regulatory skills necessary to success in the classroom (e.g., managing self in face of frustration, collaboration, dealing with distractions, etc.). If we were to *define and measure* these regulatory skills using EF terminology and referencing the EF literature (e.g., inhibitory control, working memory), we wouldn't necessarily be capturing the **basic skills and behaviors we think are actually central to success in the classroom environment** (Jones et al., 2015).

Two of the guiding principles (outlined more fully in a subsequent section of this paper) that emerged from the review of the state of measurement of young children's social and emotional development are that "what gets measured matters," and "know what you want to know." These are very important principles for making progress in the domain of social and emotional development, but we must also proceed with care and caution while keeping these in mind. If measurement drives what matters, we may miss the mark because of the definitional clutter and misalignment that currently characterizes the field. If, however, we begin with "know what you want to know" and then build measurement around it, we are certain to be more successful in providing a measure with sufficient clarity for the field (both for research and the worlds of practice and policy) (National Research Council, 2008).

A further and no less important conceptual challenge is related to context. As several authors of the commentaries (article 3 of this issue) note, skills and competencies in the social-emotional domain are highly attuned to, or susceptible to, characteristics of the immediate environment, and may actually be more so than academic or achievement-related skills. Indeed, as indicated by Jones and Yudron (article 3 of this issue), social competence, for example, is defined in part by the nature of the social dynamics and opportunities embedded in the relevant context. Said another way, children's social and emotional competencies are likely to vary in meaningful ways depending on where and when they are measured. That is not simply a problem of measurement, but a genuine feature of this domain that must be adequately captured by measurement and carried into the design of and use of assessment and accountability systems.

With sufficient understanding of contexts, in conjunction with individual children's skills, we will be better able to understand and make

judgments about what may be a young child's legitimate responses to harsh and challenging environments, and perhaps identify a need for a setting-directed intervention. Alternatively, we also need to better understand any enduring challenges a child might face across contexts that would suggest an individually-oriented intervention. In the absence of this more nuanced understanding, some children may be mischaracterized and interventions mis-targeted. Indeed some have taken this notion even further, proposing "integrated assessments" in which child functioning is understood in the context of a setting viewed as a dynamic system comprising social relations and processes, compositional features, and structural characteristics (Bouffard & Jones, 2011). Defining and then capturing, with reliable and valid measurement tools, characteristics of context holds its own set of challenges and opportunities. The last decade has seen expanding interest in defining features of settings important to child and youth development (e.g., Jones, Brown, & Aber, 2008), and there are an expanding number of research-based, practice-oriented tools (e.g., the very well-known Classroom Assessment Scoring System [CLASS] system; La Paro, Pianta & Stuhlman, 2004), but there is still a great deal of work to be done to capture both the enduring, static features of context as well as their dynamic nature.

3. Measurement issues

Turning now to measurement issues, we note that this special issue highlights topics that are specific to large-scale national surveys, as well as measurement issues that need focus when conducting in-depth studies, whether small or large. There are important connections between measures developed for in-depth studies and those used in national surveys. For example, we have sometimes seen a progression in which a measure that is first developed and its psychometric properties examined in more complex or focused studies is then brought to scale, sometimes in abbreviated form, in national surveys. We also see an important potential for national survey work to provide critical and nationally representative information on how a measure is functioning in particular subgroups and contexts, which has been a challenge for smaller in-depth studies. Thus, our summary below highlights issues specific to national surveys and also in-depth studies, and discusses potential bridges across these.

3.1. Measurement issues in national surveys

Multiple authors in this special issue, and especially the paper by Halle and Darling-Churchill, help to focus attention on specific measures of young children's social and emotional development that are promising for use in national surveys or other large-scale studies. The review of candidate measures considered both psychometric criteria such as validity and reliability, and practical considerations specific to national surveys (e.g. brevity, appropriateness for as much of the age range from birth to kindergarten entry as possible to facilitate comparison across age groups, and ability to administer the measure without needing special background or training). This review of 75 measures found only six with strengths in both psychometric and other important characteristics to make them sound candidates for use in large-scale, nationally representative surveys.

Halle and Darling-Churchill (this issue) also highlight additional challenges for measurement in national surveys. They point out the scarcity of measures of social and emotional development for use with infants and toddlers, although as noted, there are some exciting recent developments in measurement for this age range. They also address the lack of comprehensive coverage of the subdomains of social and emotional development and differences in the quality and reliability of data collected from parents vs. teachers vs. direct child observation (see also Darling-Churchill & Lippman, this issue). Furthermore, they point out the questionable appropriateness of measures for use with diverse populations, the lack of developmental continuity between assessments of abilities at earlier and later ages, and (related to the

discussion above of the jingle and jangle fallacies) the seeming overlap in operationalization of distinct competencies. Clearly, there is more conceptual as well as methodological work that needs to be done to enhance the usefulness and effectiveness of social and emotional measures within national surveys.

3.2. The potential of national surveys to inform measurement within in-depth studies

Interestingly, while the review of measures in the paper by Halle and Darling-Churchill (this issue) identifies candidate measures for use in national surveys, the commentary writers also reveal ways in which collecting data with these measures could be extremely informative for measurement challenges that go beyond the purview of federal data collection. As one example, Jones and Yudron, in their discussion of social competence (article 3 of this issue), note that behaviors that reflect social competence may differ by cultural group and context. Regarding culture, they indicate that assertiveness with adults may be seen as appropriate in one cultural group but not another; regarding context, they observe that assertiveness may be appropriate in interactions with peers but not adults in a school setting. Similarly, Denham and Howarth (article 3 of this issue) note differences in norms for emotion expression by different cultural groups. For example, approaches to discussing emotions may vary across cultural groups, and in some cultural groups the expression of emotion may be frowned upon. Differences in social competence, emotion recognition, and emotion expression, as well as the ways in which these skills predict outcomes such as academic achievement, could be examined directly across cultures in national surveys that have sufficient samples of key demographic subgroups and that include appropriately normed measures of social and emotional development.

National surveys could also provide a context for making measures of young children's social and emotional development available in multiple languages and for studying how the measures function in different subgroups. Willoughby (article 3 of this issue) notes, for example, the importance of examining whether versions of a measure in different languages, as they are used in different racial/ethnic or socioeconomic groups, show equivalent psychometric properties across groups at any one point in time, as well as over time. Understanding whether measures show equivalent psychometric properties at different points in development, in turn, is critical to understanding when and how specific aspects of young children's social and emotional development change with age.

An important distinction to be made here is in the nature of repeated, cross-sectional national surveys in contrast to large-scale longitudinal studies, such as the Early Childhood Longitudinal Study (ECLS; National Center for Education Statistics, n.d.-a) or the Head Start Family and Child Experiences Survey (Child Care and Early Education Research Connections, 2010; Office of Planning Research, n.d.). One current, federal, cross-sectional household survey, the National Survey of Children's Health (NSCH), focuses on parent report of behavioral and emotional difficulties (Data Resource Center for Child & Adolescent Health, n.d.). Others, such as the National Household Education Survey Early Childhood Program Participation and School Readiness modules (NHES ECPP; NHES SR) focus on early childhood program participation and school readiness (National Center for Education Statistics, n.d.-b). As discussed by Halle and Darling-Churchill (this issue), these cross-sectional surveys currently aim to reduce respondent burden by fielding only short lists of questions related to social and emotional development.

In contrast, the ECLS and FACES study teams have collected more in-depth, longitudinal information using direct assessment measures of the social and emotional domain. These studies have required specially trained observers to administer the measures, including for example the Leiter International Performance Scale-Revised (LEITER-R) Examiner Rating Scale (Roid, Miller, Pomplun, & Koch, 1997), the Peer Play

Observation Scale (Howes & Stewart, 1987), the Toddler Attachment Sort (TAS-45; National Center for Education Statistics, 2007), and the Two-Bags Task (National Center for Education Statistics, 2010). These more time-consuming measures of young children's development complement self-report of parents' and teachers' perceptions of and experiences with the children in the study.

Because these studies combine more in-depth measurement with specific sampling approaches, they make possible the consideration of another key measurement issue discussed by Jones and Yudron, the need to measure saturation when considering young children's social and emotional development. Saturation refers to the number and proportion of children in a classroom, home-based early care and education setting, or even in a region or district who show specific social and emotional strengths or challenges. The availability of nested data for children within classrooms or other early care and education settings in such national studies provides an important opportunity to consider such issues as how saturation relates to teacher effectiveness as well as to children's development over time. While this obviously excludes children not enrolled in out-of-home education and care arrangements or school settings, these data would prove tremendously valuable for policy and for setting program priorities.

3.3. Joint consideration of the goals of national surveys and in-depth studies

It is useful to recall that the initial measures review (Federal Interagency Forum on Child & Family Statistics, 2015-a,b) which inspired this separate collection of papers served one purpose: to provide information and options for federal data collection and reporting on the social and emotional development of young children. The commentaries in the third article of this special issue serve a different purpose: to reflect upon the methods and findings presented by Halle and Darling-Churchill (this issue), and to highlight other important factors related to measure conceptualization and quality and to identify gaps in the inventory of measures and the research more broadly.

What emerges from the joint consideration of priorities for national surveys and the broader consideration of measure conceptualization and quality? Here we present four guiding principles (two of which have already been mentioned above). First, *what gets measured matters* in that what we choose to measure can dictate what programs and policymakers will target. Poor measurement decisions may lead to poor conclusions; thus it is important to have precise, rigorous measures that lead to accurate conclusions. Second, it is important to "know what you want to know" when selecting a measure. In other words, measurement must be associated with a defined purpose consistent with a child- or program-level objective. Another way to think of this is to use the popular phrase "begin with the end in mind." Third, *there is a need for national measures that provide high quality data*. Brief, valid, and reliable scales need to be developed for use in federal surveys to provide data on social and emotional development for both local and nationwide estimates.

Fourth and finally, *collaboration and coordination among data-collecting agencies around dataset and sampling frame planning can ease tensions tied to limited resources*. The same could be said for collaboration between federal and non-federal researchers; further dialog and collaboration across the spectrum of stakeholders may serve as a win-win for advancing the field of early childhood measurement.

To begin to operationalize these principles, developing short- and long-term priorities to help fulfill the goal of measuring early childhood social and emotional development in federal data collections may be useful (e.g. Federal Interagency Forum on Child & Family Statistics, 2013). Short-term priorities that have already been noted include providing greater clarity on the subdomains to be assessed, and continuing the work to develop short but also valid and reliable measures of young children's development for use in federal surveys. In the short-term, there is also an opportunity for the federal government to set the bar with respect to benchmarks, or thresholds, describing what a young

child should know and be able to do for measures of early childhood social and emotional development.

In the longer-term, we hope this work will contribute to the development of strategies that align multiple interests. Specifically, it is desirable to have both smaller scale, in-depth studies that provide very rich data (though in need of replication in multiple samples to be generalizable) as well as larger population-based measures that provide a broader picture of trends in social and emotional development that can inform policy and practice. To move the field closer to achieving this, we elaborate now on some of the themes that have been developed throughout this special issue.

3.4. The central role of purpose of measurement

We have made a primary distinction between measures collected for the purpose of providing a nationally representative descriptive picture via national surveys, and smaller studies focusing in-depth on a particular issue. However, the papers of this special issue make clear that purpose of measurement drives measures selection beyond this key distinction. Specifically, other features of the goals for data collection also drive the psychometric requirements of a measure. As Denham and Howarth note (article 3 of this issue), when a measure is selected to guide ongoing instruction in a classroom, the focus is on whether the child is attaining knowledge and skills that align with the teacher's instructional goals. Assuming an assessment is based on ongoing observation completed for the purpose of making decisions about how best to approach instruction with a given child, a teacher who is uncertain of her rating of a child's skills and behavior is not violating the guidelines of assessment for this purpose if she asks a follow-up question or observes in other situations to be certain of her rating. However, such probing would not be at all appropriate if a measure of a child's social and emotional development is administered for the purposes of research or providing reliable indicators of children's progress throughout a geographical area. Different forms of reliability, involving consistency over time or across raters, must be much stronger when measures are used for purposes of research or accountability in a policy context.

The need for more complete documentation when presenting technical information about measures goes beyond the need to identify the purposes for which a measure is appropriate. Willoughby (article 3 of this issue) notes a lack of uniform information available on specific measures. Williford and Vick Whittaker (article 3 of this issue) observe that if we are to take seriously that some measures have been developed with the purpose of guiding instruction, documentation is needed that goes beyond the assessment itself to the specific courses of action a teacher could take to help a child progress on the specific aspect of social and emotional development considered.

The papers collected in this special issue point to the need for agreement on the documentation that should be included in technical manuals or when a measure is made available through journal publication. In addition to the purpose for which a measure has been developed and is appropriate, perhaps the criteria summarized in the paper by Halle and Darling-Churchill (this issue), including psychometric information, information about the age range and demographic subgroups for which a measure is appropriate, and information about administration requirements, could provide a basis for standardizing the information made available about measures.

3.5. Measures development

The papers of this special issue identify some exciting new steps being taken in the measurement of young children's social and emotional development. As Jones and Yudron note (article 3 of this issue), measures are being developed that involve direct observation rather than informant report. For example, measures such as the Minnesota Preschool Affect Checklist (MPAC-R; Bassetta, Thayera, Mincic, Sirotkina, & Zinsser, 2012), are being used to record observed emotion

expression, eliminating the need to rely on ratings that might be more subject to respondent bias. As another example of the development of direct child assessments and child report measures, Denham and Howarth (article 3 of this issue) describes a direct assessment of young children's understanding of emotion using the Affect Knowledge Test (Denham, Bassett, Brown, Way, & Steed, 2015), in which children respond to a puppet's behavior to assess their understanding of emotions. Additionally, a measure by Santos (Santos, Vaughn, Peceguina, Danial, & Shin, 2014) described by Denham and Howarth (article 3 of this issue) uses child report of peer social networks. This measure takes into account variation in child verbal ability at young ages by using photographs of the children in the classroom rather than requiring verbal identification.

In his discussion of executive function, an area related to but not necessarily falling within the domain of social and emotional development, Willoughby (article 3 of this issue) notes the importance of examining directly the correlations among performance-based measures (gathered through direct assessment or observation) and questionnaire measures. He further calls for an integrated approach, utilizing both performance-based measures and questionnaires, rather than attempting to replace informant report measures entirely with performance-based measures. Some national data collections, particularly those that involve longitudinal designs such as the ECLS-B and FACES, currently do involve direct assessments. It will be important to determine the potential for inclusion of data from multiple informants and perspectives in the full range of federal survey data collection.

While significant new steps are being taken in measures development, the papers of this special issue also point to gaps in the measurement of young children's social and emotional development, especially in specific subdomains. For example, Williford and Vick Whittaker (article 3 of this issue) note a lack of depth and precision in measures addressing children's development in the subdomain of self-regulation. They also note that existing measures generally have not yet been normed on large samples and thus are not ready for use at scale. This is precisely where piloting some of the more promising measures of self-regulation within federal data collection efforts might help to advance the field. Other serious gaps are noted in the area of emotional competence. Denham and Howarth (article 3 of this issue) call attention to an overall lack of strong measures in this subdomain. They identify, in particular, a lack of brief emotional expressiveness scales appropriate for teacher report. They also identify as a significant challenge that emotion understanding may be difficult for a teacher or parent to describe or rate, as it does not involve observation of behavior.

As these gaps are addressed through measures development, it will be important to keep in mind the different purposes of assessment, and to identify the specific purposes for which a measure is appropriate. This consideration should encompass the need for measures appropriate for large-scale survey administration as well as for those appropriate for guiding instruction, conducting research, and for establishing accountability in light of policy initiatives.

4. At the intersection of conceptualization and measurement

Some of the most important progress in understanding young children's social and emotional development is occurring at the intersection of conceptualization and measurement. In concluding, we note several examples of exciting work occurring at this intersection (some of which have been briefly anticipated above), and then identify a new step at this intersection that could build on and substantially extend the work of this special issue.

First, we are beginning to see the systematic examination of children's social and emotional development in light of social context, with measures that intentionally encompass both. To illustrate, Campbell (article 3 of this issue) notes that a behavior problem is a child's way of adapting to an environment that is challenging. To understand the child's behavior problem, and to help address it, requires knowledge

of the child's key social contexts as well as of his or her behaviors. In this regard, we are beginning to see measures that seek to capture and describe both the behavior of the child and of the social context, such as the Individualized Classroom Assessment Scoring System (inCLASS; Downer, Booren, Lima, Luckner, & Pianta, 2010).

A combined measurement focus on individual differences and context can, in turn, support consideration of such key issues as whether a child's challenging social behaviors are elicited by specific contexts or are consistent across context, with challenging behavior that persists across context being of much greater concern (Campbell, article 3 of this issue). An important extension of this joint consideration of individual differences and social context is the emerging understanding that some children are more vulnerable to harsh or unresponsive social contexts. We are just beginning to see measurement of child social and emotional behavior and understanding in social contexts that are systematically varied. Williford and Vick Whittaker (article 3 of this issue) describe systematic variation of the behavior of an adult (varying the degree to which the adult is dependable) with whom a child interacts prior to being asked to participate in the marshmallow self-control task. They note large differences in latency to eating the forbidden marshmallow according to the behavior of the adult. Consideration of individual differences in response to differing social contexts will be a key next step.

Second, we are seeing more systematic consideration of whether patterns of social and emotional development persist or are transient. For example, Campbell (article 3 of this issue) calls attention to the danger, in describing young children's problem behaviors, of failing to distinguish transient and more enduring problems. She notes that challenges may emerge during periods of transition (such as change in early care and education setting) or during specific periods of development (such as noncompliance and limit testing in the second year of life). Appropriate measurement may need to include thresholds to demarcate higher levels of severity, as well as repeated measurement to assess the persistence of behavior problems.

As the field moves forward, we also see a need for further intentional and systematic attention to the intersection between conceptual and measurement issues. In particular, we see a need for work that directly addresses the problem of conceptual clutter, noted by many of the authors in this special issue. We suggest tackling this issue through a mapping project that starts with a clear articulation of subdomains of social and emotional development and the constructs and relevant behaviors within them, and then aligns existing measures (and in the future, planned measures) with this clearly stated framework.

As illustrated in Table 1, the papers in this special issue provide an important starting point through the authors' clear articulation of the constructs that fall in each subdomain. Their work helps to pinpoint problematic areas of overlap and where definitional clarifications are needed. For example, what is the distinction between "impulse control" noted by Campbell (article 3, this issue) and "inhibitory control" noted by Willoughby (article 3, this issue)? Similarly, what are the distinctions between "attentional control" as an aspect of problem behaviors and "attention shifting" as an aspect of executive function? The jingle and jangle fallacies are also evident in "emotion regulation" being identified by both Denham and Howarth (article 3, this issue) and Williford and Vick Whittaker (article 3, this issue) for two distinct subdomains of social and emotional development: emotional competence and self-regulation. Furthermore, "cognitive regulation," a construct identified as part of self-regulation by Williford and Vick Whittaker, is understood by these authors to include skills also noted as part of executive function.

Further work would be needed to address the problems of definition and overlap surfaced in this special issue and illustrated in Table 1. If consensus can be achieved on a refined framework, clearly defining subdomains and constructs and minimizing or eliminating overlap, a critical next step would be the re-cataloging of existing measurement tools against this new map. Such an exercise would make clear where

Table 1

Mapping of domains, subdomains, and constructs as defined by commentators.

Domain	Subdomain	Constructs	Author
Cognitive	Executive function	Inhibitory control Working memory Cognitive flexibility/attention shifting	Willoughby (this issue)
Social and emotional	Social competence	Understanding social cues Resolving conflict Cooperating Positive relationships with others Communicating actions and feelings with social partners Recognizing and regulating emotions and actions	Jones & Yudron (this issue)
Social and emotional	Emotional competence	Emotion expression Emotion regulation Emotion knowledge	Denham & Howarth (this issue)
Social and emotional	Behavior problems	Impulse control ^a Attentional control ^b Aggressive behavior Compliance Internalizing problems Externalizing problems	Campbell (this issue)
Cognitive, social and emotional	Self-regulation	Emotion regulation Behavior regulation Cognitive regulation ^c	Williford & Vick Whittaker (this issue)

^a Potentially related to the executive function aspect of inhibitory control.^b Potentially related to the executive function aspects of working memory and/or attentional shifting.^c Often includes skills that fall within the domain of executive functioning.

the major gaps and redundancies are set the stage for a measurement agenda that is strategic and efficient. A recent example of such a mapping process is the EF Mapping Project conducted by Jones and colleagues in collaboration with the Office of Planning, Research and Evaluation within the Administration for Children and Families, U.S. Department of Health and Human Services (OPRE/ACF/HHS) (Jones et al., 2015).

5. Conclusion

This effort has moved the measurement of social and emotional development in young children forward by:

- Creating an inventory of measures, outlining common subdomain classifications to group extant measures, identifying key criteria for reviewing the quality of the measures, and making recommendations based upon the review.
- Clearly outlining key conceptual and measurement issues confounding the advancement of the study of young children's social and emotional development.
- Providing preliminary thoughts on a mapping project for the field to consider.

Going forward, this work has the potential to bear more fruit. We hope that the critical thinking about conceptual and methodological issues as well as their intersection, will lay the foundation for future research. Specifically, this special issue makes clear that work on measures of social and emotional development should more sharply delineate and define constructs within subdomains, strive to provide documentation of key aspects of measures in thorough and consistent detail, make clear how existing (and forthcoming) measures align with a map of subdomains and constructs, and prioritize the development of psychometrically strong measures as a non-negotiable feature of the larger research agenda.

Acknowledgments

Research reported in this special issue was partially supported by the Federal Interagency Forum on Child and Family Statistics [the Forum] under CDC contract# 200-2012-F-51426 to Child Trends. The content

of the special issue is solely the responsibility of the authors and does not represent the official views of the Forum.

Laura Lippman, Project Director, conceptualized this volume of papers and provided substantive contributions throughout the planning and review process. Kristen Darling-Churchill managed the project and all of the editing and submissions for this special issue.

References

- Bassett, H. H., Thayera, S. K., Mincic, M. S., Sirotnik, Y. S., & Zinsser, K. (2012). Observing preschoolers' social-emotional behavior: Structure, foundations, and prediction of early school success. *The Journal of Genetic Psychology: Research and Theory on Human Development*, 173(3), 246–278.
- Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, 78, 647–663. <http://dx.doi.org/10.1111/j.1467-8624.2007.01019.x>.
- Borghans, L., Duckworth, A. L., Heckman, J. J., & ter Weel, B. (2008). The economics and psychology of personality traits. *The Journal of Human Resources*, XLIII(4) (972–1059).
- Bouffard, S., & Jones, S. M. (2011). The whole child, the whole program: Toward integrated measures of quality. In M. Zaslow, I. Martinez-Beck, K. Tout, & T. Halle (Eds.), *Quality measurement in early childhood settings*. Baltimore: Paul H. Brookes Publishing (281–296).
- Burman, J. T., Green, C. D., & Shanker, S. (2015). On the meanings of self-regulation: Digital humanities in service of conceptual clarity. *Child Development*. <http://dx.doi.org/10.1111/cdev.12395> (Published online: 2015, July 31).
- Carlson, S. M., & Wang, T. W. (2007). Inhibitory control and emotion regulation in preschool children. *Cognitive Development*, 22(4), 489–510. <http://dx.doi.org/10.1016/j.cogdev.2007.08.002>.
- Child Care and Early Education Research Connections. (2010). Head Start Family and Child Experiences Survey (FACES) instrument matrix. Retrieved May 31, 2015, from <http://www.researchconnections.org/childcare/resources/18961/pdf>
- Cicchetti, D., & Rogosch, F. A. (2002). A developmental psychopathology perspective on adolescence. *Journal of Consulting and Clinical Psychology*, 70(1), 6–20.
- Data Resource Center for Child & Adolescent Health. (n.d.). National Survey of Children's Health. Retrieved May 31, 2015, from <http://childhealthdata.org/learn/NSCH>
- Denham, S. A., Bassett, H. H., Brown, C. A., Way, E., & Steed, J. (2015). "I know how you feel": Preschoolers' emotion knowledge contributes to early school success. *Journal of Early Childhood Research*, 13(3), 252–262.
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135–168.
- 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.
- Duckworth, A. L., Quinn, P. D., & Tsukayama, E. (2012). What no child left behind leaves behind: The roles of IQ and self-control in predicting standardized achievement test scores and report card grades. *Journal of Educational Psychology*, 104(2), 439–451. <http://dx.doi.org/10.1037/a0026280>.
- Federal Interagency Forum on Child and Family Statistics. (2013). Key points from measuring and reporting social-emotional development culminating activity,

- by session. Natcher Conference Center, NIH Main Campus, Bethesda MD. Available at <http://www.childstats.gov/forum/deliverables.asp>
- Federal Interagency Forum on Child and Family Statistics. (2015-a). Characteristics of existing measures of social and emotional development in early childhood: Applications for federal reporting and data collection. Retrieved from <http://www.childstats.gov/forum/deliverables.asp>
- Federal Interagency Forum on Child and Family Statistics. (2015-b). Inventory of existing measures of social and emotional development in early childhood. Washington, DC: Author. Retrieved from <http://www.childstats.gov/forum/deliverables.asp>
- Garcia, E. (2014). The need to address non-cognitive skills in the education policy agenda: Economic policy institute. Briefing paper #386. Available at <http://www.epi.org/publication/the-need-to-address-noncognitive-skills-in-the-education-policy-agenda/>.
- Howes, C., & Stewart, P. (1987). Child's play with adults, toys, and peers: An examination of family and child care influences. *Developmental Psychology, 23*(3), 423–430.
- Jones, S. M., & Bailey, R. (unpublished manuscript). The taxonomy project: Defining and codifying "non-cognitive" skills for learning and life success.
- Jones, S. M., Bailey, R., & Partee, A. (2015). *Mapping executive function: From constructs to measures to behaviors: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services*.
- Jones, S. M., Brown, J. L., & Aber, J. L. (2008). Classroom settings as targets of intervention and research. In M. Shinn, & H. Yoshikawa (Eds.), *Toward positive youth development: Transforming schools and community programs* (pp. 58–77). New York: Oxford University Press.
- La Paro, K. M., Pianta, R. C., & Stuhlman, M. (2004). The classroom assessment scoring system: Findings from the prekindergarten year. *The Elementary School Journal, 104*(5), 409–426. <http://dx.doi.org/10.1086/499760>.
- Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences of the United States of America, 108*(7), 2693–2698. <http://dx.doi.org/10.1073/pnas.1010076108>.
- National Center for Education Statistics. (2007). Early childhood Longitudinal Study, Birth cohort (ECLS-B) Psychometric report for the 2-year data collection. Washington, DC. Information available from: http://nces.ed.gov/pubs2007/2007084_c8.pdf.
- National Center for Education Statistics. (2010). Early childhood longitudinal study, birth Cohort (ECLS-B) preschool-kindergarten 2007 psychometric report. Washington, DC. Information available from: <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2010009>.
- National Center for Education Statistics. (n.d.-a). Early Childhood Longitudinal Study (ECLS) program. Retrieved May 31, 2015, from <https://nces.ed.gov/ecls/>.
- National Center for Education Statistics. (n.d.-b). National Household Education Surveys (NHES) program, early childhood surveys. Retrieved May 31, 2015, from https://nces.ed.gov/nhes/surveyttopics_early.asp
- National Research Council. (2008). In C. E. Snow, & S. B. Van Hemel (Eds.), *Early childhood assessment: Why, what, and how*. Washington, DC: Board on Children, Youth, and Families, Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education: The National Academies Press.
- Office of Planning Research and Evaluation. (n.d.). Head Start Family and Child Experiences Survey (FACES), 1997–2018. Retrieved May 31, 2015, from <http://www.acf.hhs.gov/programs/opre/research/project/head-start-family-and-child-experiences-survey-faces>
- Reeves, R. R., & Venator, J. (2014). *Jingle-jangle fallacies for non-cognitive factors*. Washington, DC: Brooking Institution.
- Roid, G. H., Miller, L. J., Pomplun, M., & Koch, C. (1997). Leiter International Performance Scale. Available from: <http://www4.parinc.com/Products/Product.aspx?ProductID=LEITER-3>.
- Santos, A. J., Vaughn, B. E., Peceguina, I., Danial, J. R., & Shin, N. (2014). Growth of social competence during the preschool years: A 3-year longitudinal study. *Child Development, 85*(5), 2062–2073.