Measuring and Reporting Social-Emotional Development in Early Childhood

Wednesday, May 8, 2013

Natcher Conference Center, NIH Main Campus, Bethesda MD

Key Points from Culminating Activity by Session

Session	Speaker	Remarks

Welcome & Introductions
 	
Moderated by Regina Bures (Forum project lead; NIH/NICHD) and Traci Cook (Forum)
	Jen Park (Forum; OMB representative)	• While we are a "society hungry for data," the questions we have to answer have not yet been fully conceptualized and measures have not been fully developed. This session will address ways we can meet information needs for public policy and personal decision-making.
		• Agencies are increasingly suggesting data collection for program monitoring or performance evaluations of their programs over time in addition to research.
		• Important to consider the following in a Federal statistical system:
		  o Budget constraints
		  o Human capital
		  o Respondent cooperation
		  o Confidentiality and privacy
		  o Causal use of federal statistics by entrepreneurs
		  o Policy relevance
		• Think of federal data as a product from and for the public to use to make personal, business, and policy decisions *rather than* a product only to be accessed and used by trained professionals who offer their interpretations through peer review.
		• Suggested innovations:
		  o Encourage greater collaboration in planning/use of datasets and sampling frames.
		  o Explore "fitness for use" [in study designs] (indirect assessments, telephone, web, and other methods) to allocate resources to areas for precision.
		  o Explore use of administrative records (e.g., program data) to improve timeliness of data being reported.
		  o Facilitate transparency with metadata (end user will not always be the savvy user).
		  o Harness innovate ways to collect data and make data available to users. "Try new things."

Session	Speaker	Remarks
Policy Needs ***Moderated by Laura Lippman (Child Trends)***	Lauren Supplee (ACF; Head Start CARES)	- Social and emotional development matters. - What we measure says something about what matters. - What we choose to measure can dictate what programs need to focus on. Poor measurement may lead to poor conclusions. - Challenges: - Assuring the comprehensiveness of the identified domains of social and emotional development since it is also a cultural construct. - Clearly defining the purpose of the data collection. - Connecting the purpose of a measure to theory (what a program may affect vs. indicator of overall development). - Achieving excellent psychometric properties (e.g., sensitive to change, appropriate norms, predictive validity [*see Duncan paper cited at end of summary*]). - Administration (language, ease, time).
	Catherine Scott-Little (UNC)	- States look at data to monitor trends, identify gaps in services, needs for interventions, and teacher preparation programs. - States look to federal studies to see what measures were used and how to incorporate the measures in state level studies. - From a state-level policy perspective, the following considerations are important: - Predictive validity – What is measured needs to matter over time. - Face validity – Can I look at the data and understand it/how it makes sense? - Developmental significance – What makes a difference for children? How do teacher practices make a difference? - Alignment with standards – States will look to research to build their standards, as well as determine what measures can best be used to assess existing standards. - Practical, convenient, and easy measurement is important. - Expertise, financial resources, and data collection systems are challenges. - Views Kindergarten through 3rd grade levels as "uncharted territory." This age range needs standards and more measures. Kindergarten entry assessments are on state's radars right now. - The translation from research to policy/practice needs to be a part of the conversation from the beginning.

Session	Speaker	Remarks
Consensus and Criteria for Assessing Measures of Early Social-Emotional Development	Marty Zaslow *Presenter* (SRCD; Child Trends)	• Presented overview of Child Trends inventory of existing measures of social-emotional (SE) development in early childhood and review paper (*see list of substantive materials at the end of this summary*). • Based on a review of numerous frameworks, found broad agreement on SE subdomains in the literature, but less consensus on definitions of specific skills and attributes within each subdomain. • The inventory of measures is intended as a foundation, not an exhaustive list, with a focus on measures that could be used in, or are already being used by, federal surveys. (*see copy of summary slide at end of this document*)
	Jim Griffin *Discussant* (NICHD)	• In measuring social-emotional development, we need to: o Look at our conceptualization of social emotional development, before we begin measuring and reporting. o Consider advances in neuroscience, and think about how children grow and develop and how they process outside stimulation and their own internal representation. o Have better understanding of predictors, and how programs help develop these predictors. o Explore topics such as stress response in toxic environments, which has important implications for how children deal with the world and other people. • It's important and valuable to develop new measures. It's also important to see the findings people get with each existing measure. How did the measure add to the conceptual model?
Views from the Field *Moderated by Amy Madigan (ACF/DCFD)*	Tammy Mann (Campagna Center)	• What we should be doing at the federal-level is in fact influenced by what is happening at the local-level. • Programs would like to have measures that allow you to understand children from a formative perspective. • Concerned about lack of construct stability in the 0-2 year old age range and the challenge in defining what various behaviors represent. • Noted that because culture accounts for some variation in development, it is important that we incorporate it into our thinking around measurement; this includes dual language learning.

Session	Speaker	Remarks
	Susanne Denham (GMU)	• Observational and direct assessment tools are very important in understanding the young child. • Considerations for measurement: o Age and developmental appropriateness o Language and cognitive development o Non-verbal response options o Developmental disabilities o No ambiguity of the measures o Gender – Boys, in particular. Are we addressing what we need to know about them? o Temperament o Cultural background o Environmental factors (e.g., maternal depression, stress at home, etc.)
	Karen Bierman (Penn State)	• Acquisition of *skills/behavior problems* can be measured with just 8-10 items and is stable over the long term. • We know less about *regulatory behaviors* and how to measure them. • We need to think more about measures with strong (and long term) predictive validity and measures that interact with others to produce outcomes. • Teacher ratings are much better predictors of later outcomes (based on Fast Track study). Parent ratings contribute added predictability and stability, but only explain an additional 3% of the variance in Kindergarten and not at all in 1st grade. Data varies greatly depending on who you ask.
	Eva Marie Shivers (Indigo Cultural Center)	• Contextual measures needs to be considered (e.g., relating to school, family, community, poverty, etc.). • Direct assessments and observational measures are important although expensive and time consuming. • The field has few measures normed with diverse populations; she was glad to see that measures were weighted based on whether they were assessed in diverse samples in the Inventory and paper. • Concerned about criteria that rated measures with larger age spans as more positive/appropriate. She also thought the use of dyadic measures would be more appropriate than using a developmental health perspective because infants exist in the context of a caregiver relationship. However, she approves of using a caregiver/teacher report as child care providers tend to be more culturally matched to the children.

Session	Speaker	Remarks
Promising Measures *Moderated by Tamara Halle (Child Trends)*	Susanne Denham (GMU) *Emotional Competence*	• There are few measures of emotional competence. o Among the available measures, there is <u>not</u> a measure that captures all the facets of emotional competence. • "Screening measures" – questionnaires that can be completed easily by teachers and/or parents — are worth considering but they lack specificity. • Computerized direct assessments can be used with ease in the classroom with the advantage that no assessor training would be required. **COMMENTS ON MEASURES:** • Infant Toddler Social Emotional Assessment (ITSEA) is a great measure but did not get at emotional competence. • Ages & Stages Questionnaire (ASQ) assesses emotional competence and has valuable subscales. • Expressiveness = a sub-domain of a sub-domain → Infant Behavior Questionnaire (IBQ), Early Childhood Behavior Questionnaire (ECBQ), Children's Behavior Questionnaire (CBQ) • Emotion regulation → Ages & Stages Questionnaire: Social Emotional (ASQ:SE), Emotion Regulation Checklist • Emotion knowledge – did not include anything for infant/toddler → Affective Knowledge Test (AKT) has concurrent and predictive validity. • Social Problem Solving → Social Skills Improvement-Rating Scales (SSIS-RS); Challenging Situations Task (CST); Schultz Test of Emotion Processing
	Celene Domitrovich (CASEL) *Social competence and behavior problems*	• It is important to look at *both* emotional competence *and* problem behaviors when looking at behavior problems. • Remember that social and emotional competence involves the integration of emotion, cognition, and behavior. • More discussion is needed on the purpose of measurement. There are opportunities and a need to put direct assessment on the table. • Biggest concerns: o Goodness of fit (not all skills emerge in all contexts) o Reporter bias o Questions about cross sectional vs. longitudinal study design o Different approaches to intervention: at the child vs. teacher levels o Accurately matching the content of the measure to the identified sub-domain • Important to consider how measures cut across infant/toddler and preschool – assessing these skills at different ages with the same measure might be good or bad depending on the context. **COMMENTS ON MEASURES**: *See below for screen shot of slide distilling Inventory paper Table 7, particularly the measures that assess across age spans.*

Session	Speaker	Remarks
	Sally Atkins-Burnett (Mathematica) *Self-Regulation*	- Social and emotional development is important for adult outcomes but not always for academic achievement. - A problematic issue is that most self-regulation measures look at dysregulation instead of regulation. - It is important to note that children can be high on both problem behaviors and social competence, i.e., a high score on a measure of problem behavior does not necessarily mean that the child has poor social competence. - Important to take into account cultural variance. - Problems with direct assessment: - Measures using tablet technology may tap into different constructs because tablets keep children's attention and lack normal distraction. - Direct assessments give you a good idea of what a child can do in a short amount of time, but might not capture children's capabilities at their best. - Advantages to assessor reports (over direct assessment): - Easy to collect - Low cost - Able to capture behavior over a long period of time - Assessors tend to have a broader idea of what is normative than teachers - Helpful to have parent reports if you are thinking longitudinally (to eliminate variance from different teachers reporting on the same child). **COMMENTS ON MEASURES**: - Arousal regulation is important at the infant-toddler stage, so measures of sleep should be considered. - Social Skills Rating System (SSRS) for kindergarten and up is a solid measure.

Session	Speaker	Remarks
	Karen Bierman (Penn State) ***Executive Function***	• There is a lot of excitement about Executive Function (EF), but there is still a lack of consensus and lack of clarity about how it should be defined and which measures are best. Specifically, there is debate in the field of developmental psychology about whether we're talking about effortful control (a bottom up phenomena) or EF (development of pre frontal cortex). • Many rating scales measure overlapping but different constructs. • Issues in direct assessment of EF: o Direct assessment picks up on something not obtained from behavior questionnaires. o With EF, direct assessments are designed to have children do something they have not done before. If they are given the same assessment multiple times, it's not tapping EF anymore – few EF measures are good over 4 years of development. • EF measures are better than impulsivity measures at predicting academic achievement. **COMMENTS ON MEASURES**: • SSRS is overused because it has great labels. It seems to measure compliance and obedience (not 'control'). • Dimensional Change Card Sort (DCCS) and Head Toes Knees Shoulders (HTKS) add new rules as children get older and can be used over multiple years.
Options for Data Reporting and Collection *Moderated by Bob Kominski (Census)*	Gail Mulligan (NCES)	• The biggest issue in selecting and administering measures is the time and burden on participants. • National Center for Education Statistics (NCES) looks for assessments that are short by design, but typically adapts existing measures. • There are drawbacks in shortening existing measures: o May become less reliable. o May be less related to the outcome measure than the original scale. o Conversely, participants may become annoyed at having to answer too many questions that seem to assess the same thing. • Staff training is intensive and not always successful. • NCES has a demonstrated record of collaboration with ACF and others. **COMMENTS ON MEASURES**: • SSRS became SRS – Social Rating Scale, a hybrid of items. • For EF, NCES is currently fielding DCCS and Backward digit Span (appropriate measures for 5 years and older).

Session	Speaker	Remarks

	Stephen Blumberg (NCHS)	• Issues in the current questions included in National Center for Health Statistics (NCHS) surveys:
		   o Discriminating between at-risk, flourishing, and other
		   o Length of the scales
		   o Ease of use for analysts
		   o Validation of scoring algorithms
		   o Each year, there is a discussion of what to cut, and long lists of questions typically get cut
		• Opportunities for social and emotional questions in *Division of Health Interview Statistics* at NCHS:
		   o National Survey of Children's Health (NSCH) redesign of sample and questionnaire
		   o National Health Interview Survey (NHIS) annual supplements
		   o NHIS follow-back surveys
		   o NHIS questionnaire design (redesign to be completed in 2017)
		• An important consideration is whether to release data from a questionnaire as agencies feel they must also be able to release the question. If a copyright holder refuses, agencies won't use it in any surveys.
		• In many cases, NCHS validates a full scale then comes up with a shorter version that isn't subject to as rigorous of an evaluation.

	Yanique Edmond (SAMHSA)	• Project Launch funds states that work on improving child service systems and communities that look at evidence-based practices.
		   o Not a national survey, but it builds on work refining and defining social and emotional competence; seeks to translate a public health approach by providers to what is happening in communities to enhance SE development for children.
		   o Their dilemma is that they don't have robust measures and community levels for social and emotional development.
		• The goal is to understand how different providers understand social and emotional development.
		• Project focuses on different ages (0-1, 2-3 and school setting).
		• The program is built on academic readiness but the overarching goal is for children to be successful over their lifespan.
		• Realizing that they could not measure all of this across different/diverse communities, the Project conducted a special study (grantees chose particular outcomes and measures to assess those outcomes). The result was a decision to focus on the age range without good measures of SE development (infants).
		• Currently looking to research for recommendations on good measures of social and emotional development for age groups without good measures (infants, in particular).
		• Looking for measures that can be modified for tribal and urban communities.

Session	Speaker	Remarks
	Maria Woolverton (FACES)	• Family and Child Experiences Survey (FACES) is one of the only sources of child level outcomes data for Head Start (and demonstrates increases in skills and decreases in problem behaviors). • FACES includes a parent interview (social skills and problem behaviors), teacher-report (cooperative classroom behavior, problem behaviors, and approaches to learning), and a direct child assessment (assessor ratings and executive functioning as of 2009). • Measurement considerations and challenges: o Differences in age, gender, risk factors. o Likelihood of relationship between teacher-reported social skills and the classroom organization based on classroom observations. o Interpreting differences across reporters on social and emotional measures (culture, context, expectations). o Availability of appropriate norms because they use part of full scales. o Time limits for direct child assessments. o Tension between maintaining consistency of measurement over time and updating battery with state of the art measures. • FACES redesign was launched in 2011to determine if the information being collected is still relevant in the field, as well as to increase the efficiency of data collection.
	Lynda Laughlin (SIPP)	• Survey of Income and Program Participation (SIPP): Household survey on household economic well-being (topics include child care, child well-being, fertility, assets, etc.) that follows respondents for a minimum of three years. The next collection begins in 2014. • The child well-being topical module includes daily parent interactions, performance in school, and participation in activities, academic performance, neighborhood characteristics, etc. for children 0-17 years old. They also collect data on family living arrangements and household income dynamics. • Laughlin advocated for supplements in the survey from other federal agencies.

Session	Speaker	Remarks
	Bob Kominski & Panel	Question posed to expert panel: Due to space and time limitations, would you 1. Pick one narrow domain within the larger set of domains? 2. Identify a series of SE domains and touch on everything with a small set of questions (possibly just 1 per domain)? OR 3. Another option? • **Karen Bierman:** o There is no point in doing a 1 item measure. No one is going to be able to use it. o On the other hand, some measures in the Inventory were designed for other purposes. o Due to time constraints, most researchers use 5 to 6 selected items from longer scales. These modified scales aren't being published as assessment measures but there could be a process to see how many people are using shortened versions. • **Gail Mulligan:** Federal data collectors have content review panels where experts provide input. • **Stephen Blumberg:** Peer-reviewed journals tend to publish complex scales and may be skeptical about taking a 30 item scale and reducing it.

Reference: Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., et al. (2007). *School readiness and later achievement*. Developmental Psychology, 43, 1428-1446.

<u>**Substantive Materials Prepared for the Event**</u>

Characteristics of Existing Measures of Social and Emotional Development in Early Childhood: Applications for Federal Reporting and Data Collection (a.k.a. the "Inventory Paper"), by Child Trends

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

Also available electronically: *Inventory of Measures of Social-Emotional Development in Early Childhood*, compiled by Child Trends. This document presents descriptive information, in tabular format, about prominent measures of social-emotional development.

Celene Domitrovich's distillation of Inventory Paper Table 7

This excerpt from Celene Domitrovich's presentation summarizes Table 7 of the Inventory Paper authored by Child Trends. This shows how the measures included in that paper cut across infant/toddler and preschool, which may be good or bad depending on what the measure is used for*.*

Table 7: Measures Meeting 5 or 6 Criteria

Infant/Toddler Period (0-2)
- ITSEA   (SC, BP, SR)
- IBQ-R   (SC, BP, SR)
- NCATS   (SC, EC)
- TBAQ   (BP, SR)

Preschool Period (= or > 3)
- SSIS-RS   (SC, BP)
- PIPPS   (SC, BP)
- EDI   (SC, EC)
- BST   (EC, SR)
- AKT   (EC)
- PLBS   (EF)

Crosses Both Periods
- CCTI   (SC, BP, SR, EF)
- MESSY   (SC, BP, SR)
- Two Bags   (SC, BP, SR)
- DECA-C   (EC, BP, SR)
- BASC-2   (BP, SR, EF)
- CBCL   (BP, SR, EF)
- CDI   (SC)
- CBQ   (SR)

SC: Social Competence
EC: Emotional Competence
BP: Behavior Problems
SR: Self Regulation
EF: Executive Function

KEY TO ACRONYMS:		EDI	Early Development Instrument
AKT	Denham's Affect Knowledge Test	IBQ-R	Infant Behavior Questionnaire Revised
BASC-2	Behavior Assessment System for Children, Second Edition	ITSEA	Infant Toddler Social Emotional Assessment
BST	Brief Scale of Temperament	MESSY	Matson Evaluation of Social Skills with Youngsters
CBCL	Child Behavior Checklist	NCATS	Nursing Child Assessment Teaching Scale
CCTI	Colorado Childhood Temperament Inventory	PIPPS	Penn Interactive Preschool Play Scales
CDI	Child Development Inventory	PLBS	Preschool Learning Behaviors Scale
CBQ	Children's Behavior Questionnaire	SSIS-RS	Social Skills Improvement System-Rating Scales
DECA-C	Devereux Early Childhood Assessment Clinical Form	TBAQ	Toddler Behavior Assessment Questionnaire

Marty Zaslow's Summary of the Inventory Paper (excerpted from slides)

This chart summarizes the number of measures reviewed for each subdomain, and how many of the reviewed measures met half or more of the ten criteria used to determine recommendations. See **Characteristics of Existing Measures of Social and Emotional Development in Early Childhood: Applications for Federal Reporting and Data Collection** (a.k.a. the "Inventory Paper") for details regarding these criteria.

Subdomain	Total # reviewed	# for consideration as a regular indicator that met > ½ of criteria	# for consideration as a one time indicator that met > ½ of criteria	# for consideration for first use in federal surveys that met > ½ of criteria
Social Competence	27	none	1	4
Emotional Competence	18	none	none	2
Behavior Problems	18	1	2	5
Self Regulation	26	1	3	6
Executive Function	16	1	2	1

Summary:

- There are many measures of **social competence**, but there are no measures appropriate for use on a recurrent basis.
- **Self-regulation** also has many measures, but only one is deemed appropriate for use as a regular indicator.
- **Behavior Problems** and **Executive Function** also have only one measure each that could be used for regular data collections.
- We are lacking **emotional competence** measures in the federal statistical system.
- Across subdomains, there is only a **small set** of measures that could be used as a **one-time indicator**. There is, however, a **large set** of measures that could be **considered for first use** in federal data collections.