Compendium of Preschool Through Elementary School Social-Emotional Learning and Associated Assessment Measures

Susanne A. Denham

George Mason University

Peter Ji

University of Illinois at Chicago

Bridget Hamre

University of Virginia

October 2010

Deliverable for "Assessments for Social, Emotional, and Academic Learning with Preschool /Elementary-School Children"

Funded by the Bill & Melinda Gates Foundation, the Kirlin Charitable Foundation, the NoVo Foundation, the Rauner Family Foundation, and the University of Illinois at Chicago

Department of Psychology
UIC UNIVERSITY OF ILLINOIS AT CHICAGO
Social and Emotional Learning Research Group

CASEL

Table of Contents

Introduction

In this compendium we focus on tools to assess the social and emotional learning (SEL) of preschool and elementary school students (i.e., five- to ten-year-olds), along with aspects of the contexts in which they learn and their learning behaviors. These assessment measures are intended for researchers and those in the applied research community – for example, educators and social workers – who may find them useful in their work with groups of children.

Increasingly, SEL has been shown to be critical to children's success in school, both academically and socially (Greenberg et al., 2003; Payton et al., 2007). We have identified five core SEL competencies: self-awareness, self-management, social awareness, responsible decision-making, and relationship/social skills (Payton et al., 2000; Zins et al., 2007).

Self-awareness. Self-awareness refers to the ability to accurately assess personal feelings, interests, values, and strengths. As children move into elementary school, their feelings of self-efficacy become even more important. This aspect of SEL also includes identifying and labeling one's own feelings. Although preschoolers have a well-defined, stable sense of self, much development in this area occurs during the preschool through primary years.

Self-management. Self-management includes the ability to handle one's emotions in productive ways, being aware of feelings, monitoring them, and modifying them when necessary, so that they aid rather than impede the ways in which the child is able to cope with varying situations. This aspect of SEL also includes handling stress, persevering despite obstacles, and expressing emotions appropriately. At the same time, important non-emotional aspects of self-management are paramount to success in the preschool through elementary years. These include being able to use some non-SEL skills (e.g., working memory, attention, and inhibitory control) to regulate one's social and academic behavior.

Social Awareness. This aspect of SEL includes the ability to take others' perspectives, understand their feelings, and empathize with them, and to appreciate others' similarities and differences. Children are constantly attempting to understand their own and others' behavior, and emotions play a role in this understanding, conveying crucial interpersonal information that can guide interaction. Inability to interpret emotions can make the classroom a confusing, overwhelming place. Much development occurs in this area of SEL across this age period.

Relationship Skills. The goal in this aspect of SEL is to promote positive and effective exchanges with others and, ultimately, relationships that last over time. Numerous skills are crucial at this level, including making positive overtures to play with others, initiating and maintaining conversations, cooperating, listening, taking turns, seeking help, and developing friendship skills (e.g., joining another child or small group, expressing appreciation, negotiating, and giving feedback). In addition, asserting oneself, resolving conflict, and addressing others' needs through negotiation develop during the preschool-to-primary period.

Responsible decision-making. Responsible decision-making becomes important as the everyday social interactions of preschoolers increase in frequency and complexity. Young children must learn to solve social problems – to analyze social situations, identify problems, set prosocial goals, and determine effective ways to solve differences that arise within their peer group. Responsible decision-making increasingly includes the ability to make appropriate ethical decisions that consider and respect others and promote the well-being of the school and community. At the ages considered here, complying with classroom rules, resisting peer pressure, and controlling aggression or other disruptive behavior also reflect such responsibility.

All of these components of SEL are interrelated. To perform research on these components and their relation to academic and social success, to focus on the needs of individual children, and to show the success of SEL programming, assessment tools are required. This compendium provides a guide to such instruments.

In addition, knowledge has increased about the characteristics of school contexts and how they relate to both academic and social success. For example, healthy relationships in the classroom both predict SEL skills and are predicted by them over time. Aspects of the classroom that are important in this regard include**,** along with supportive relationships among teachers and children, effective classroom management and good instruction (including instruction in SEL skills and opportunities for children to apply those skills)**.**

The assessment tools in this compendium have been gathered to provide researchers, especially educational and policy practitioners, a means to gauge the SEL skills of the preschool/ elementary students in their care, as well as to evaluate contexts that promote SEL and its long-term outcomes. Many compendia of measurement tools are currently available, but we believe this compilation has special value because (1) it follows our framework of SEL-related inputs (see Table 1, next page); and (2) it is more comprehensive than most.

Assessment tools are included for the leftmost two columns of Table 1 only (we do not consider the rightmost column because our focus here is on more proximal aspects of SEL and schooling in the preschool and elementary years). Rather than attempting to be completely exhaustive, we have included measures that match the core SEL constructs for each age range and for multiple reporters and methods, where possible.

The tools in this compendium are organized into three sections, corresponding with Context (Section 1); the Core SEL Competencies of Self-Awareness, Self-Management, Social Awareness, Relationship Skills, and Responsible Decision-Making (Section 2), and Academic-related SEL Competencies, i.e., Feelings about School/School Climate, and Academic Competencies (Section 3). We have created a table for each section so users can have an overview of what SEL assessments are available, for which student grades, and for which SEL core competency.

Within each table, the assessments were first sorted by grade level (preschool, preschool/elementary, elementary school). Specifics regarding age and grade range for each measure are further detailed in the measure description. Within each grade level assessment tools were then sorted alphabetically by grade level group and basic rating type – including teacher rating, parent rating, student report, performance-based, observational, and "other." For each measure, we provide information about the measure itself, including administration (rater, format, and length), scoring, psychometrics, overall strengths and weaknesses, pricing, source, and references.

We acknowledge there are gaps in what we offer in each section. For some constructs and some age ranges, few assessment tools are available. These gaps are noted in the introductions to each section. The SEL assessment field is evolving, and although we applaud what has been done so far, we recognize the need for more work to create new and better assessment tools. Finally, we chose, for reasons of space and focus, not to include in this compendium curriculum-based, diagnostic, or work sampling assessment tools.

For other information on SEL assessment, please see the following:

Denham, S. A. (2006). Social-emotional competence as support for school readiness: What is it and how do we assess it? *Early Education and Development, Special Issue: Measurement of School Readiness, 17*, 57-89.

Denham, S. A., & Weissberg, R. P. (2004). Social-emotional learning in early childhood: What we know and & where to go from here? In E. Chesebrough, P. King, T. P. Gullotta, & M. Bloom (Eds.), A blueprint for the promotion of prosocial behavior in early childhood (pp. 13-50). New York: Kluwer/Academic Publishers.

Denham, S. A., Wyatt, T., Bassett, H. H., Echeverria, D., & Knox, S. (2009). Assessing social-emotional development in children from a longitudinal perspective. *Journal of Epidemiology and Community Health, 63,* 37-52.

Greenberg, M. T., Weissberg, R.P., O'Brien, M.U., Zins, J. E., Fredericks, L., Resnik, H., & Elias, M. J. (2003). School-based prevention: Promoting positive social development through social and emotional learning. *American Psychologist***,** *58(6/7)*, 466-474.

Payton, J. W., Wardlaw, D. M., Graczyk, P. A., Bloodworth, M. R., Tompsett, C. J., & Weissberg, R. P. (2000). Social and emotional learning: A framework for promoting mental health and reducing risk behaviors in children and youth. *Journal of School Health, 70(5)*, 179-185.

Payton, J., Weissberg, R.P., Durlak, J.A., Dymnicki, A.B., Taylor, R.D., Schellinger, K.B., & Pachan, M. (2008). *The positive impact of social and emotional learning for kindergarten to eighth-grade students: Findings from three scientifi c reviews.* Chicago, IL: Collaborative for Academic, Social, and Emotional Learning.

Zins, J. E., Bloodworth, M. R., Weissberg, R. P., & Walberg, H. J. (2007). The scientific base linking social and emotional learning to school success. *Journal of Educational and Psychological Consultation, 17,* 191-210.

Table 1. Core Constructs for Student Social, Emotional, and Academic Competencies; Classroom & School Context; and Long-term Student Outcomes

Context	Student's SEL Competencies	Short-term Student Outcomes
• **Effective Classroom Management** (e.g., materials clearly organized, student matters dealt with efficiently and appropriately, proactively) • **Instructional Support** (e.g., teacher has students prepared) • **Healthy Relationships** (e.g., teachers emotionally support students, respond to student concerns) • **Social and Emotional Skills Instruction** (e.g., class time devoted to sequenced, active, focused, and explicit social and emotional skills development) • **Opportunities for Social and Emotional Skill Application** (e.g., extra-curricular programming, student government, service-learning) • **Prosocial Norms and Behaviors** (e.g., clear standards and expectations, prosocial and healthy behaviors are established and rewarded)	• **Self-Awareness** (e.g., identifying feelings, beliefs, and values, self-confidence, self-esteem, appropriate emotional expression, curiosity, optimism, perceived competence, self-efficacy) • **Self-Management** (e.g., attention, emotion regulation, self-control, anger management, stress management, setting and achieving goals, flexibility and adaptability, initiative and persistence) • **Social Awareness** (e.g., emotion knowledge, empathy, social awareness, perspective taking, respect for differences, belongingness to family, community culture) • **Relationship skills** (e.g., peer competence, effective communication and listening skills, ability to work well with others, cooperation with others, conflict-resolution skills, negotiation skills) • **Responsible Decision-Making** (e.g., follows rules, takes responsibility for mistakes, resists negative peer pressure, responsible decision-making and social problem-solving skills, leadership skills)	• **Lack of internalized problems** (e.g., depression, anxiety) • **Lack of external behavior problems** (e.g., acting out, discipline problems) • **School Engagement** (e.g., relationships with teachers, social connections in the classroom, classroom participation, academic motivation, attendance) • **Academic Competence** (e.g., GPA, achievement test scores) **Long-term (Secondary School and Beyond) Outcomes** • **Academic Success** (e.g., high school graduation, college graduation) • **Mental Health** (e.g., reduced behavioral and emotional disorders, positive life outlook) • **Physical Health** (e.g., engage in behaviors that promote healthy lifestyles, does not engage in risky behaviors such as substance abuse or unprotected sex) • **Positive Social Relationships** (e.g., having network of friends, positive family relationships, and parenting skills) • **Prepared for workforce** (e.g., skills and dispositions to gain, maintain, and advance in employment) • **Civic and Community Engagement** (e.g., community service, civic participation, or contribution)

Section 1: Aspect of School Context

This includes measures in Table 2 on:

- **Effective Classroom Management** (e.g., materials clearly organized, student matters dealt with efficiently, appropriately, proactively)
- **Instructional Support** (e.g., teacher has students prepared)
- **Healthy Relationships** (e.g., teachers emotionally support students, respond to student concerns)
- **SEL-Supportive Environment**
    - **Social and Emotional Skills Instruction** (e.g., class time devoted sequenced, active, focused, explicit social and emotional skills development)
    - **Opportunities for Social and Emotional Skill Application** (e.g., extra-curricular programming, student government, service-learning)

Table 2 summarizes the following measures. Although it may appear that there is a dearth of means to assess healthy relationships in the classroom, those that exist (the CLASS and the STRS) are gold standards in the field.

Table 2. Context Measures, Sub-Construct Assessed, Age Range, Pre-school or Elementary School, and Rating Type*

Scale Name	Environmental Skills and Instruction				School/Grade Level		Rating Type						
	Effective Classroom Management	Instructional Support	Healthy Relationships	SEL-supportive Environment	Preschool	Elementary School	Teacher	Parent	Peer	Self	Observational	Performance-based	Other (e.g., interview)
Arnette Caregiver Interaction Scale (CIS)				■	■						■		
Assessment Profile For Early Childhood Programs (APECP)	■	■		■	■		■						■
Early Childhood Environment Rating Scale –Revised Edition (ECERS-R)	■	■		■	■						■		■
Preschool program quality assessment, 2nd edition (PQA)		■		■	■						■		
Classroom Assessment Scoring System (CLASS)	■	■	■		■	■					■		
Early Childhood Classroom Observation Measure (ECCOM)	■	■		■	■	■					■		
School-age care environment rating scale (SACERS)				■	■	■					■		
Student-Teacher Relationship Scale			■		■	■	■						
Assessment Of Practices In Early Elementary Classrooms (APEEC)		■				■					■		■

* Shaded cells indicate area of context assessed, age level, or rating type.

Measure	Arnett Caregiver Interaction Scale (CIS)
Constructs	Classroom Emotional Environment, SEL-Supportive Environments
Age range	Originally validated for preschool
Rating type	Observational
Description of measure as related to construct of interest	"To rate the emotional tone, discipline style, and responsiveness of teachers and caregivers in a classroom. The items focus on the emotional tone and responsiveness of the caregiver's interactions with children. The scale does not address issues of curriculum or other classroom management issues (such as grouping or flow of activities)" (U.S. Department of Education, 1997, p. 78).
Administration	Training Required: You must achieve a .70 inter-rater reliability for two consecutive visits to be a certified Arnett Caregiver Interaction Scale observer (Jaeger & Funk, 2001).
Scoring	The Caregiver Interaction Scale (CIS) consists of 26 items usually divided into 4 subscales. Some researchers have conducted factor analyses on the 26 items and have found different subscales (e.g., Whitebook, Howes, & Phillips, 1989). Observers are asked to rate the extent to which 26 items are characteristic of the child care provider whom they are observing. Items are scored on a 4-point scale from (1) Not at all characteristic to (4) Very much characteristic of the child care provider. The measure usually contains the following subscales: Sensitivity (10 items) Harshness (8 items) Detachment (4 items) Permissiveness (4 items)
Reliability	Jaeger and Funk (2001) reported inter-rater reliability coefficients ranging from .75 to .97 between a certified observer and trainees. For internal consistency, Cronbach's alphas from the Observational Study of Early Childhood Programs (Layzer, 1993): Warmth/responsiveness rating (10) = .91, and Harshness rating (7) = .90 Jaeger and Funk (2001) reported coefficients of .81 and higher for the sensitivity (positive interaction), punitiveness, and detachment subscales.
Validity	Concurrent Validity: Layzer (1993) found correlation coefficients of .43 to .67 between the CIS and several other measures of child care quality (i.e., Early Childhood Environment Rating Scale (ECERS), Assessment Profile for Early Childhood Programs, Description of Preschool Practices. However, the authors did not expect large coefficients because the CIS focuses more narrowly on an aspect of teacher behavior than the other observation measures.

Strengths	The training is short and there is no cost for the measure.
Weaknesses	---
Publication/Pricing	Source: Arnett, J. (1989). Caregivers in day-care centers: Does training matter? Journal of Applied Developmental Psychology, 10, 541- 522. (Note that this article does not contain a list of the items on the scale. However, this is the article that is typically cited when the CIS is used.) Publisher: A copy of the scale can be found in Jaeger and Funk (2001). Cost: None

Jaeger, E. & Funk, S. (2001). *The Philadelphia Child Care Quality Study: An examination of quality in selected early education and care settings*. Philadelphia. Philadelphia, PA: Saint Joseph's University.

Layzer, J. I. (1993). *Observational study of early childhood programs. Final report. Volume I: Life in preschool*. (ERIC # ED366468). Washington, DC: US Department of Education.

Love, J. M., Meckstroth, A., & Sprachman, S. (1997). *Measuring the quality of program environments in Head Start and other early childhood programs: A review and recommendations for future research (Working Paper No. 97-36).* Washington, DC: U.S. Department of Education National Center for Education Statistics.

US Department of Education. (1997) *The pocket condition of education.* National Center for Education Statistics; Washington, DC: National Center for Education Statistics, Aug 1997

Whitebook, M., Howes, C., & Phillips, D. (1989). *Who cares? Child care teachers and the quality of care in America. Executive summary of the National Child Care Staffing Study.* Oakland, CA: Child Care Employee Project.

Measure	Assessment Profile for Early Childhood Programs (APECP)
Age range	Preschool through 2nd grade
Construct	Effective classroom management, Instructional Support, SEL-Supportive Environments
Rating type	Teacher, Other
Description of measure as related to construct of interest	Population Measure: Developed with the Assessment Profile for Early Childhood Programs: Research Edition I (1992) was originally standardized using 401 preschool classrooms in child care, Head Start, and kindergarten settings. Later revised using a national standardization sample of 2,820 classrooms. Subsequent analyses across the original 87 items were conducted to confirm the factor structures, to estimate reliability, and to recalibrate the IRT properties. Following the analyses, each scale was reduced to 12 items and the Assessment Profile: Research Edition II was published in 1998. Summative Measure: The Assessment Profile for Early Childhood Programs: Research Edition II (APECP) is a global measure of quality used by researchers to evaluate the learning environment and teaching practices in classrooms for young children. The Assessment Profile for Homes with Young Children: Research Version was developed using items on the Assessment Profile for Early Childhood Programs: Research Version. The Family Child Care Homes version (APFCCH) has only been used in the NICHD Early Childhood Research Project and the authors have never established any psychometric properties. Formative Measure: The Assessment Profile for Early Childhood Programs: Preschool, Toddler, Infant, School-Age, and Administration instruments are formative evaluation measures used for program improvement purposes. These measures are more comprehensive than the summative research tool, and provide user-friendly procedures for self-evaluation of early childhood settings. As formative measures, they are supported by software that provides extensive analyses and detailed program improvement recommendations. The Assessment Profile for Early Childhood Programs tool evaluates center- based, classroom and administrative practices while the Assessment Profile for Family Child Care Homes is a companion tool for formative evaluation purposes in the family child care setting.
Administration	Test administration: Data collection requires observation, review of records, and interview with teachers, administrator(s), and/or family child care provider(s). Training required: Training is required to establish inter-rater reliability. Training involves a review of the criteria and data collection methods and on-site practice observation, record review, and interviews. Training generally involves 2-3 days.

Scoring	Summative Measure: The Assessment Profile for Early Childhood Programs: Research Edition II is an observation checklist with dichotomous items and includes five scales with 12 items each to assess Learning Environment, Scheduling, Curriculum Methods, Interacting, and Individualizing. These five scales have met the unidimensionality criteria for Information Response Theory (IRT) creation of scales and have shown a strong fit to a three- parameter IRT model (Abbott-Shim, Neel, & Sibley, 2001). Formative Measure: The Assessment Profile for Early Childhood evaluates the Safety (109 items), Learning Environment (73 items), Scheduling (34 items), Curriculum Methods (49 items), Interacting (61 items), and Individualizing (25 items) practices within classrooms. The number of items for each dimension varies depending upon the age group observed; the maximum number of items is noted in parentheses. Administrative practices are evaluated in terms of Physical Facilities (68 items), Food Service (45 items), Program Management (63 items), Personnel (38 items), and Program Development (31 items).
Reliability	Inter-rater Reliability For both the summative and formative versions of the Assessment Profile, inter-rater reliabilities between a trainer and observers is consistently reported with a mean of 93 to 95 percent agreement with a range of 83 to 99 percent agreement (Abbott-Shim, Lambert, & McCarty, 2000). Numerous other research studies have reported similar inter-rater reliabilities. Internal Consistency The reliability coefficients for the five scales (Learning Environment, Scheduling, Curriculum, Interacting, and Individualizing) range from .79 to .98 for the Kuder-Richardson 20 and from .81 to .98 for the Spearman-Brown corrected split-half. The IRT based reliabilities for the five scales range from .83 to .91 (Abbott-Shim, Neel & Sibley, 1992).
Validity	Criterion Validity Criterion related validity was established by examining the relationship of the Assessment Profile: Research Edition I to the Early Childhood Environment Rating Scale (ECERS) (Harms & Clifford, 1980). In these criterion related validity studies, Wilkes (1989) found a significant correlation ($r = .64$, $p = .001$), and Abbott-Shim (1991) found a significant correlation ($r = .74$, $p = .001$). Construct Validity A second-order factor analysis was used to determine whether the five scales of the Assessment Profile: Research Edition II form a single latent construct of classroom quality. These results indicated that observed measurements using these factor scores stem from a single underlying construct of classroom quality (Abbott-Shim, Lambert, & McCarty, 2000).

	Content Validity Content validity was documented through a review of the instrument by a wide range of early childhood professionals and a cross-reference of the items with the initial NAEYC Accreditation Criteria (National Association for the Education of Young Children, 1998). The cross-reference showed extensive consistency between the two measures with 100% match of the criteria. This has been periodically updated as the accreditation criteria have been modified (Abbott-Shim, Neel, & Sibley, 2001).
Strengths	The measure has also demonstrated to be reliable and valid in assessing the child care environment.
Weaknesses	This assessment does not measure teacher-children interactions, and must be used in combination with other assessments that examine interaction variables.
Publication/Pricing	Summative Measure: o Assessment Profile for Early Childhood Programs: Research Edition II o Assessment Profile for Homes with Young Children: Research Version o Assessment Profile for Early Childhood Programs: Research Edition Technical Manual Publisher: Martha Abbott-Shim. 294 Woodview Drive Decatur, GA 30030 Email: martha.abbottshim@gmail.com Formative Measure: o Assessment Profile for Early Childhood Programs o Assessment Profile for Family Child Care Homes Publisher: Quality Assist, Inc. 17 Executive Park Drive, Suite 150 Atlanta, GA 30329 Phone: 404-325-2225 Website: www.qassist.com Summative Measure: Assessment Profile for Early Childhood Programs: Research Edition II: $18 (3 classrooms), Technical Manual: $25 Formative Measure: Assessment Profile for Early Childhood Programs and Assessment Profile for Family Child Care Homes – pricing is based on the scope and specification of the evaluation plan regarding training, data collection, technology support (PDA), data analysis and reporting.

Abbott-Shim, M., Lambert, R., & McCarty, F. (2000). Structural model of Head Start classroom quality. *Early Childhood Research Quarterly, 15*(1), 115–134.

Abbott-Shim, M., Neel, J., & Sibley, A. (2000). *Assessment profile for early childhood programs: Research technical manual.* Atlanta, GA: Quality Counts, Inc.

Harms, T., & Clifford, R. M. (1980). *Early Childhood Environment Rating Scale*. New York: Teachers College Press.

International Reading Association (1998). *Using multiple methods of beginning reading instruction. A position statement of the International Reading Association*. Newark, DE.

Measure	Early Childhood Environment Rating Scale – Revised Edition (ECERS-R)
Constructs	Effective classroom management, Instructional Support, SEL-Supportive Environments
Age range	Preschool
Rating type	Observational
Description of measure as related to construct of interest	As described by the authors: The Early Childhood Environment Rating Scale (ECERS-R) measures global quality in center-based early childhood programs. The ECERS-R can be used as a tool "to see how well a program is meeting children's needs – to see whether children receive the protection, learning opportunities, and positive relationships they need for successful development" (Cryer, Harms & Riley, 2003, p. x). It can be used by researchers, practitioners, program monitors and early childhood professionals providing technical assistance to programs. The ECERS-R is a revision of the ECERS originally published in 1980, which "… retains the original scale's broad definition of environment, including those spatial, programmatic, and interpersonal features that directly affect the children and adults in an early childhood setting" (Harms, Clifford, & Cryer, 1998, p. 1).
Administration	Test Administration: The ECERS-R book provides questions for each item that can guide the interview. The authors also provide specific instructions for administering the scale and for conducting the observation in a way that minimizes the impact of the observer on the classroom environment. Because of the large number of indicators that need to be scored, the observer should have the ECERS-R book with her/him while in the classroom and should complete scoring before leaving the facility. Training Required: The authors recommend that observers "participate in a training sequence led by an experienced ECERS-R trainer before using the scale formally. The training sequence for observers who will use the scale for monitoring, evaluation, or research should include at least two practice classroom observations with a small group of observers, followed by inter-rater reliability comparison" (Harms et al., 1998, p. 5). Five-day and three-day trainings are offered by the authors of the scale at the University of North Carolina, Chapel Hill. Observers can purchase additional resources including a video training package (available from Teachers College Press) or the All About the ECERS-R book (Cryer, Harms, & Riley, 2003), particularly to develop reliability and be more consistent with the ECERS-R authors.
Scoring	The scale consists of 43 items categorized into seven subscales. Items are scored on a 7-point scale from 1 to 7. Numbered indicators outlining the specific requirements for the item are provided at score points 1 (inadequate), 3 (minimal), 5 (good), and 7 (excellent). The observer begins at level 1 and scores each indicator "yes," "no," or

	"NA." The final score is determined by the number of indicators that have been "passed." All indicators must be passed at each level to score at or above that level. Thus, to score a 7 on an item, all indicators must be passed including all of those included under Level 7. Includes the following scales: Space and Furnishings (8 items); Personal Care Routines (6 items); Language-Reasoning (4 items); Activities (10 items) ; Interaction (5 items); Program Structure (4 items); and Parents and Staff (6 items)
Reliability	<u>Inter-rater reliability</u> "Overall the ECERS-R is reliable at the indicator and the item level, and at the level of the total score. The percentage of agreement across the full 470 indicators in the scale is 86.1%, with no item having an indicator agreement level below 70%. At the item level, the proportion of agreement was 48% for exact agreement and 71% for agreement within one point. For the entire scale, the correlations between the two observers were .92 product moment correlation (Pearson) and .87 rank order (Spearman). The interclass correlation was .92" (Harms et al., 1998, p. 2). <u>Internal Consistency</u> The authors "also examined the internal consistency of the scale at the subscale and total score levels. Subscale internal consistencies range from .71 to .88 with a total scale internal consistency of .92" (Harms et al., 1998, p. 2).
Validity	<u>Predictive Validity</u> The authors note that, since the original ECERS had demonstrated that "quality as measured by the ECERS has good predictive validity (i.e., Peisner-Feinberg & Burchinal, 1997; Whitebook, Howes, & Phillips, 1990), the revised version would be expected to maintain that form of validity" (Harms et al., 1998, p.2). <u>Content Validity</u> When the scale was revised, the authors conducted focus groups with experts in the field who made suggestions for the revision based on how the ECERS had worked in inclusive and culturally diverse settings. The authors also gathered feedback and suggestions from researchers and other ECERS users that informed the content in the ECERS-R.
Strengths	There are many items on the ECERS-R that measure diversity. Inter-rater reliability and internal consistency were also high on the ECERS-R. It has been used extensively in research and practice.
Weaknesses	---

Publication/Pricing	Source: Harms, T., Clifford, R. M., & Cryer, D. (1998). Early Childhood Environment Rating Scale – Revised Edition. New York, NY: Teachers College Press. Harms, T., Clifford, R. M. & Cryer, D. (2005). Early Childhood Environment Rating Scale – Revised Edition. New York, NY: Teachers College Press. (Updated with additional notes and a new expanded scoresheet). Publisher: Teachers College Press 1234 Amsterdam Avenue New York, NY 10027 Cost: All materials are available through Teachers College Press Manuals (ECERS-R, 2005) $17.95 (ECERS-R, 1998) $14.95 Video Training Packages 1999, VHS $59.00 2006, DVD $59.00 Training Workbook 1999 $4.00

Cryer, D., Harms, T. & Riley, C. (2003). All *about the ECERS-R: A detailed guide in words & pictures to be used with the ECERS-R*. PACT House Publishing.

Harms, T., Clifford, R. M., & Cryer, D. (1998). Early Childhood Environment Rating Scale – Revised Edition. New York, NY: Teachers College Press.

Harms, T., Clifford, R. M. & Cryer, D. (2005). Early Childhood Environment Rating Scale – Revised Edition. New York, NY: Teachers College Press. (Updated with additional notes and a new expanded scoresheet).

Peisner-Feinberg, E., & Burchinal, M. (1997). Relations between preschool children's child care experiences and concurrent development: The Cost, Quality and Outcomes Study. *Merrill-Palmer Quarterly, 43*, 451-477.

Whitebook, M., Howes, C., & Phillips, D. (1989). *Who cares? Child care teachers and the quality of care in America. Executive summary of the National Child Care Staffing Study.* Oakland, CA: Child Care Employee Project.

Measure	Preschool Program Quality Assessment, 2nd edition (PQA)
Constructs	Instructional Support, SEL-Supportive Environments
Age range	Preschool
Rating type	Observational
Description of measure as related to construct of interest	The current revision of the PQA includes two notable differences from earlier versions: 1) the number of content areas has increased from four to seven, and 2) the scoring system has been revised to adequately measure the full range of quality along each quality construct. As described by the authors: "The Preschool Program Quality Assessment (PQA), Second Edition, is a rating instrument designed to evaluate the quality of early childhood programs and identify staff training needs. Developed by High/Scope Educational Research Foundation, it is appropriate for use in all center-based settings, not just those using the High/Scope educational approach. The Preschool PQA intentionally reflects "best practices" in early childhood education as a whole. The measure identifies the structural characteristics and dynamic relationships that effectively promote the development of young children, encourage the involvement of families and communities, and create supportive working environments for staff" (High/Scope Educational Research Foundation, 2003, p. 1). The PQA can be used for a variety of purposes including both pre-service and in-service training initiatives, self-assessment and monitoring. The PQA can also be used to conduct observations and provide feedback to staff. In addition, the Preschool PQA can be used as a research tool when administered by trained outside observers to document program practices, compare quality, examine the relationship between quality of care and children's outcomes, and evaluate the effectiveness of staff development initiatives. Finally, the Preschool PQA can be used to explain research-based practices to a variety of individuals and agencies including administrators, policymakers, and support staff in the preschool (High/Scope Educational Research Foundation, 2003).
Administration	Who Administers Measure/Training Required <u>Test Administration</u>: The measure may be administered by independent raters including researchers, program evaluators, outside consultants or agency administrators. In addition, site staff including directors, early childhood specialists, curriculum coordinators, teachers, or parents may also complete it as part of a self-assessment. Students may also use their PQA observations as part of their training to become teachers or caregivers.

	Training Required: Training to acceptable levels of inter-rater reliability on the PQA takes 2 days. The first day is devoted to reviewing and practicing the PQA, using anecdotes and raw-footage videotapes. The second day is used to conduct actual observations and determine inter-rater reliability.
Scoring	The PQA is comprised of seven areas of program quality, three of which are based on classroom observation, and four of which are based on interviews with teachers and/or directors. The first four areas are classroom-specific, while the latter three are program-specific. Each area has between 5 and 13 items, with several indicators per item. Raters score each indicator on a 5-point scale. The administration manual provides a detailed description of the scoring procedures. The areas of program quality and items are summarized below. Classroom Items o Learning Environment (9 items) o Daily Routine (12 items) o Adult-Child Interaction (13 items) o Curriculum Planning and Assessment (5 items) Agency Items o Parent Involvement and Family Services (10 items) o Staff Qualifications and Staff Development (7 items) o Program Management (7 items)
Reliability	The revised PQA was field tested in two research projects: the 2000 cohort of Phase 2 of the Michigan School Readiness Program (MSRP) evaluation with a sample of 19 classrooms and 2,000 children (Smith, Jurkiewicz, & Xiang, 2002), and the Michigan Full-Day Preschool Comparison Study with two cohorts comprising 121 and 132 classrooms (Jurkiewicz, 2003). A broad range of public and private early childhood settings were represented by these samples, permitting rigorous testing of the psychometric properties of the new PQA. The following are exemplary points of information from these studies. Inter-rater Reliability Pairs of raters were sent to 10 classrooms to observe the learning environment, daily routine, and adult-child interaction. Pearson's correlations were calculated to be 0.57 for learning environment ($p<0.10$), 0.75 for daily routine ($p<0.05$), and 0.74 for adult-child interaction ($p<0.05$). Internal Consistency "To assess internal consistency, Cronbach's alpha was calculated on five quality constructs (learning environment, daily routine, adult-child interaction, curriculum planning and assessment) and total PQA scores. There was insufficient data to

	determine internal consistency on the other two constructs (staff qualifications and development, and program management) since these were only rated once at the agency level rather than for each classroom. . .Internal consistency for the new version was calculated with 185 classrooms in three samples. . .and averaged 0.93, with all but two of the results within the acceptable range of 0.70 to 0.90" (High/Scope Educational Research Foundation, 2003, p. 11).
Validity	<u>Concurrent Validity</u> "The validity of quality constructs within sections I through V of the revised PQA was assessed in relationship to the Teacher Beliefs Scale. The PQA was significantly correlated, in the expected positive or negative direction, with appropriate and inappropriate teacher beliefs and practices. With one exception [(the correlation between the learning environment of the PQA and appropriate practices of the Teacher Beliefs Scale, r = 0.16)], all correlations were significant and ranged in magnitude from 0.28 to 0.49" (High/Scope Educational Research Foundation, 2003, p. 12). <u>Predictive Validity</u> PQA scores are significantly related to children's developmental outcomes, both while children are in preschool, and kindergarten, and is associated with established measures of child development (e.g. DIAL-R, High/Scope COR) and teacher ratings. <u>Confirmatory Factor Analysis</u> "A confirmatory factor analysis was conducted with sections I through V using a sample of approximately 150 classrooms. Five factors emerged, accounting for 58% of the variance, and their content aligned with the five corresponding PQA sections: Learning Environment, Daily Routine, Adult-Child Interaction, Curriculum Planning and Assessment, and Parent Involvement and Family Services. Factor loadings ranged from 0.43 to 0.82, with the majority (64%) at 0.60 or higher. However, several daily routine items, notably those related to group times (e.g., small- and large-group time), loaded on the adult-child factor. These items were modified in the final version of the PQA" (High/Scope Educational Research Foundation, 2003, p. 12).
Strengths	The PQA has been shown to be both a reliable and valid measure to assess program quality. One strength of the PQA is that it measures both structural features of the preschool classroom as well as relationships between children and adults.
Weaknesses	There was insufficient data to determine internal consistency on staff qualifications and development and program management since these were only rated once at the agency level rather than for each classroom.
Publication/Pricing	Jurkiewicz, T. (2003). The Revised Preschool PQA: Report on psychometric properties. Instrument evaluation report to the Michigan Department of Education. Ypsilanti, MI: High/Scope Educational Research Foundation, Research Division.

	Smith, C., Jurkiewicz, T., & Xiang, Z. P. (2002). Program quality in Michigan School Readiness Program classrooms: Classroom characteristics, teacher beliefs, and measurement issues. Evaluation report to the Michigan Department of Education. Ypsilanti, MI: High/Scope Educational Research Foundation, Research Division. Cost: The cost of the PQA is $25.95.

High/Scope Educational Research Foundation. (2003). PQA: Preschool Program Quality Assessment. Ypsilanti, MI: Author

Jurkiewicz, T. (2003). *The Revised Preschool PQA: Report on psychometric properties.* Instrument evaluation report to the Michigan Department of Education. Ypsilanti, MI: High/Scope Educational Research Foundation, Research Division.

Smith, C., Jurkiewicz, T., & Xiang, Z. P. (2002). *Program quality in Michigan School Readiness Program classrooms: Classroom characteristics, teacher beliefs, and measurement issues.* Evaluation report to the Michigan Department of Education. Ypsilanti, MI: High/Scope Educational Research Foundation, Research Division.

Measure	Classroom Assessment Scoring System (CLASS)
Constructs	Healthy Relationships, Effective classroom management, Instructional Support, Classroom Emotional Environment
Age range	Pre-K through 12th grade (various versions; the original CLASS is used through 3rd grade)
Rating type	Observational
Description of measure as related to construct of interest	As described by the authors: The Classroom Assessment Scoring System (CLASS) is an observational instrument developed to assess classroom quality in preschool through third grade classrooms. The CLASS dimensions are based on observed interactions among teachers and students in classrooms. The dimensions were derived from a review of constructs assessed in classroom observation instruments used in child care/Elementary school research, literature on effective teaching practices, focus groups, and extensive piloting. The Observational Record of Classroom Environments (ORCE, ECRN, NICHD, 1996) served as a foundation for the development of the CLASS.
Administration	Test Administration: Trained CLASS users observe in classrooms for twenty minute intervals and then score each CLASS dimension. The manual recommends gathering at least four of these twenty minute intervals to assess a classroom. It is also possible to score with the CLASS based on videotaped footage. Although the manual describes a standardized protocol for observation, the procedure can be modified to meet the goals of specific projects. Training Required: Training is required to assure proper use of the instrument for each of its intended uses (i.e., research, professional development, program development and evaluation). All observers must attend training and pass a reliability test. Regular training sessions are available at the University of Virginia and University of North Carolina – Greensboro. Trainers are also available to provide local trainings. In addition, there are several "Train the Trainer" workshops each year at the University of Virginia in which people can become certified CLASS trainers. Refer to website for a schedule of trainings.
Scoring	Scores for each dimension are made on a 7-point scale (low range – 1, 2; mid range – 3, 4, 5; high range – 6, 7) using the descriptions of classrooms that fall within the scoring ranges for each dimension. Each dimension has corresponding indicators and then behavioral descriptions with scale points for each indicator. Domain scores can be calculated as an average of each of the corresponding dimension scores. To generate composite scores across cycles, individual cycle scores for each dimension are averaged across the number of cycles of observations completed (Pianta, LaParo, & Hamre, 2004).

Reliability	Inter-rater Reliability All observers must attend training on the CLASS and take a reliability test. Observers code five 20-minute videotaped classroom sessions. The average inter-rater reliability (within one point of master codes) is reported in the Technical Appendix (p. 9) as 87 percent. Two observers both coded a total of 33 30-minute digital videotapes submitted by teachers in the MyTeachingPartner (MTP) Study. Inter-rater reliability (within 1 point of each other) ranged from 78.8 percent (for Behavior Management and Instructional Learning Formats) to 96.9 percent (for Productivity). Similar levels of reliability have been obtained in live observations (Hamre et al., 2006, p. 9). Internal Consistency Correlations among the CLASS dimensions range from .11 to .79. Correlations for the preschool sample in the MS/SWEEP Studies were generally lower than those for the third grade sample in the 4R's Study. Confirmatory factor analyses were performed on data from each of the studies except for the Induction Study (Hamre et al., 2006). Analyses revealed three factors representing Emotional Support, Classroom Organization, and Instructional Support. Within the MTP sample, which used the most current version of the CLASS, internal consistencies were: Emotional Support (alpha = .89); Classroom Organization (alpha = .77); and Instructional Support (alpha = .83). Stability across Time Stability of ratings across observation cycles was assessed in preschool and 3rd grade classrooms using data from the NCEDL MS Study of preschool and the 4R's Study of 3rd grade classrooms in New York City. For the 3rd grade sample, correlations between the first cycle and the total score are moderate to high, ranging from .68 for Productivity to .87 for Positive Climate. For the preschool sample, correlations between the first 4 cycles and the final score ranged from .84 for Productivity to .91 for Concept Development. By completing two cycles correlations with the final score are uniformly high with almost all correlations above .90 in both preschool and 3rd grade (Hamre et al., 2006, p. 10). Correlations between observations made on two consecutive days suggest a high degree of stability, with correlations between the two days ranging from .73 for Productivity to .85 for Teacher Sensitivity. "There were small but significant mean changes across several of the dimensions with a general trend toward lower quality scores on the second day. Given that there is no reason to expect a systematic difference in quality across two consecutive days these small changes may be due to observer bias in which scores become slightly lower over time. Again, however, although these differences are statistically significant, they are relatively small effects and correlations between the two days are high" (Hamre et al., 2006, p. 13).

	CLASS scores have also been found to be relatively stable across the school year, at least in a large number of preschool classrooms. Analyses also indicate that 7-point rating scales of the classroom are highly stable and not dependent on occasion.
Validity	<u>Criterion Validity</u> The CLASS domains of Emotional Support, Classroom Organization, and Instructional Support are correlated with teacher reports of depression and adult- centered attitudes. Specifically, classrooms with lower scores across the CLASS dimensions had teachers who reported higher levels of depression while those with lower scores on classroom organization and instructional support had teachers who reported more adult-centered attitudes. <u>Concurrent Validity</u> In comparisons of the CLASS with the Early Childhood Environmental Rating Scale (ECERS-R), classrooms with higher CLASS scores were rated higher on the ECERS interactions factor (correlations range from .45 to . 63). Correlations between CLASS ratings and the Furnishings and Materials factor from the ECERS were only moderate, ranging from .33 to .36 (Pianta et al., 2005). The CLASS has also been compared to The Snapshot, a time-sampling method used to assess the percent of time spent on various activities (Pianta et al., 2005). Because the CLASS assesses the quality rather than the quantity of classroom activities, it is not surprising that there were low (but still significant) correlations between the CLASS instructional support domain and time spent in literacy and math according to The Snapshot. Children in classrooms with higher CLASS scores spent more time in elaborated interactions with adults and significantly more time engaged. <u>Predictive Validity</u> Results from the NCEDL multi-state study provide evidence that classroom quality, as assessed by the CLASS, is associated with children's performance at the end of preschool, as well as gains in their performance across the preschool year (Howes et al., in press). These associations were sustained, even after controlling for a variety of covariates, including maternal education, ethnicity, and gender. The most consistent and robust classroom quality dimension for predicting growth across time was the Instructional Support of the classroom as assessed by the CLASS. The CLASS Emotional Support scale was associated with growth in children's expressive and receptive language scores, as well as decreases in teacher-reported behavior problems (Howes et al., in press). <u>Content Validity</u> The CLASS dimensions are based on observed interactions among teachers and students in classrooms. The dimensions were derived from an extensive review of constructs assessed in classroom observation instruments used in child care/Elementary school research, literature on effective teaching practices, focus groups, and piloting.

Strengths	The CLASS has strong reliability and validity supporting its status as an effective measure of classroom interactions support. It has a wide age range, as the CLASS can be used in Pre-K through 12th grade classrooms. It is also aligned with a variety of professional development support to help teachers improve the quality of their interactions with students.
Weaknesses	---
Publication/Pricing	Source: Hamre, B. K., Mashburn, A. J., Pianta, R. C., Locasale-Crouch, J., & LaParo, K. M. (2006). Classroom Assessment Scoring System Technical Appendix. Pianta, R. C., LaParo, K. M., & Hamre, B. K. (2008). Classroom Assessment Scoring System. Brookes Publishing Publisher: Paul H. Brookes Publishing Co. Post Office Box 10624 Baltimore, MD 21285-0624 Phone: 800-638-3775 Website: www.brookespublishing.com Cost: Two-day training at UVA: $600/person Four-day training (Train the trainer): $1,000/person Local Training: $3,000 for up to 15 people (plus travel costs for 1 trainer) Pre-K Manual: $49.95 K-3 Manual: $49.95 Pack of 10 scoring forms: $25 See www.classobservation.com

Curby, T.W., LoCasale-Crouch, J., Konold, T.R., Pianta, R., Howes, C., Burchinal, M. et al (2009). The relations of observed pre-k classrooms quality profiles to children's academic achievement and social competence. *Early Education and Development, 20*, 346-372.

Hamre, B. K., Mashburn, A. J., Pianta, R. C., Lacasle-Crouch, J., & LaParo, K. M. (2006). *Classroom Assessment Scoring System technical appendix*.

Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development, 76,* 949-967.

Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (in press). Ready to learn? Children's pre-academic achievement in pre- kindergarten programs. *Early Childhood Research Quarterly.*

LaParo, K. M., Pianta, R. C., & Stuhlman, M. (2004). The classroom assessment scoring system: Findings from the pre-kindergarten year. *The Elementary School Journal, 104,* 409-426.

Mashburn, A.J., Pianta, R., Hamre, B.K., Downer, J.T., Barbarin, O., Bryant, D., Burchinal, M., Clifford, R., Early, D., & Howes, C. (2008). Measures of classroom quality in pre-kindergarten and children's development of academic, language and social skills. *Child Development, 79,* 732-749.

National Institute of Child Health and Human Development, Early Child Care Research Network (NICHD ECCRN) (2002). The relation of global first-grade classroom environment to structural classroom features and teacher and student behaviors. *The Elementary School Journal, 102*(5), 367-387.

Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science, 9,* 144-159.

Pianta, R., LaParo, K. M., & Hamre, B. K. (2004). *Classroom Assessment Scoring System (CLASS).* Unpublished measure, University of Virginia, 2004.

Measure	Early Childhood Classroom Observation Measure (ECCOM)
Constructs	Effective classroom management, Instructional Support, SEL-Supportive Environment
Age range	4-7 years old
Rating type	Observational
Description of measure as related to construct of interest	Population measure developed with kindergarten and first-grade teachers public and private schools, in urban and rural areas, in the northeast, and on the west coast). Teachers were predominantly female and Caucasian. As described by the authors: "Most extant observation measures of early childhood classroom environments focus predominantly on the social climate and resources of the classroom, with less attention given to the quality of instruction provided by the teacher. The Early Childhood Classroom Observation Measure (ECCOM) was developed to tap the nature and quality of academic instruction as well as the social climate, resources, and other aspects of effective classrooms" (Stipek & Byler, undated coding manual, p. 1). The measure focuses on the approach used for instruction rather than subject matter content. The instrument was developed primarily as a research tool. However, at least one research team (Head Start Quality Research Project) is using the ECCOM as an intervention tool as well as for research (Stipek & Byler, 2004). Thus, the ECCOM may be used for research, as a professional development tool, and/or as a program development and evaluation tool, but the value of the ECCOM for professional development purposes has not yet been systematically assessed.
Administration	Test administration: Observations are recommended to be conducted by a trained observer, on a typical day, beginning at the beginning of the day for full-day programs or at the beginning of the Early Childhood Classroom Observation Measure (ECCOM) program for less-than-full-day programs. Observations occur over a 3-hour period, and should always include observations of both math and literacy instruction. Training Required: All observers should attend two full days of training and pass a reliability test (i.e., demonstrate 80% reliability on coding with the head trainer or previously certified observer).
Scoring	The ECCOM reported on in Stipek and Byler (2004) consists of 32 items (17 constructivist, 15 didactic) rated on a scale of 1 (practices are rarely seen) to 5 (practices predominate). There were parallel items for both constructivist and didactic practices, but there were two additional items in the constructivist scale (relevance of instruction activities and teacher warmth).

	The rating of each item occurs after an observation of the classroom. Scores are based roughly on the percentage of time the described practices were seen during observation. <u>Constructivist Subscales</u> - Instruction. A high score occurs if children are held accountable for completing work and held to a clear standard, lessons are coherent and well-connected to children's previous knowledge, lessons teach identifiable concepts and are focused on understanding, children are active participants in instructional conversations, and specific strategies for math and literacy instruction are implemented. - Management. A high score occurs if teachers provide children with choices in both teacher-planned activities and during free time, rules and routines are clear but flexible, children are given developmentally appropriate responsibilities, and discipline is brief and non-disruptive (often involving explanations or assisting children in their own social problem solving). - Social climate. A high score occurs if teachers are warm, responsive, attentive, and respectful of children. <u>Didactic Subscales</u> - Instruction. A high score occurs if the teacher holds children accountable for completing work and for attaining universal rather than individualized standards, lessons focus on discrete skills, the teacher focuses on facts and procedural knowledge, the teacher controls the classroom conversations, and math and literacy instruction . Early Childhood Classroom Observation Measure (ECCOM) emphasizes learning distinct skills which are not embedded in meaningful contexts and also strongly emphasizes correctness. -Management. A high score occurs if the rules and routines are teacher-determined, children do not select their own activities outside of recess, and the teacher takes responsibility for maintaining order in the classroom, including intervening quickly in social conflict situations. -Social climate. A high score occurs if there are few social interactions among children, little collaborative work among children, and most children work individually or in a teacher-led group. Tasks and expectations are teacher- or curriculum-driven and uniform across all children.
Reliability	<u>Inter-rater Reliability</u> Observers independently rated 26 classrooms in pairs. Intraclass correlations were used to calculate reliability. Reliability was high for all subscales (Stipek & Byler, 2004, p. 387). <u>Internal Consistency</u> Alphas were high for all subscales (Stipek & Byler, 2004, p. 388).

Validity	<u>Predictive Validity</u> Stipek and Byler (2004) found predictable associations between ECCOM scores, and teachers' self-reported practices, teaching goals, relationships with children, and perceptions of children's ability to be self-directed learners. The authors concede that direct observation of child behaviors and skills would be better than relying on teacher report for assessing associations between measures. Additionally, the study found strong correlations between the constructivist and didactic subscales of the ECCOM and teachers' educational level. The less education teachers had, the more they engaged in didactic instruction ($r = -0.29$, $p < 0.001$), and the more education teachers had, the more they engaged in constructivist instruction ($r = 0.27$, $p < 0.001$).
Strengths	The measure addresses diversity by having a checklist of "Representations Related to Diversity" and a "Treatment of Native Language." Additionally, the measure has strong predictive validity.
Weaknesses	The validity information was gathered by teacher report, and would be more accurate if data had been gathered through direct observation of child behaviors and skills.
Publication/Pricing	Source: Stipek, D., & Byler, P. (2004). The early childhood classroom observation measure. *Early Childhood Research Quarterly, 19*, 375-397. Publisher: Unpublished. The measure may be obtained by emailing Deborah Stipek at stipek@stanford.edu Cost: Contact Dr. Deborah Stipek at stipek@stanford.edu

Stipek, D., & Byler, P. (2004). The early childhood classroom observation measure. *Early Childhood Research Quarterly, 19,* 375-397.

Measure	School-Age Care Environment Rating Scale (SACERS)
Constructs	SEL-Supportive Environments
Age range	School-age
Rating type	Observational
Description of measure as related to construct of interest	As described by the authors: The School – Age Care Environment Rating Scale (SACERS) measures environmental quality in school age care settings.
Administration	Test Administration: The instrument may be used by the care giving staff for self-assessment, by directors as a program-quality measure for planning program improvement, by agency staff for monitoring, in teacher training programs, and by parents concerned about quality care for their school-age children. Training Required: Training is required to assure proper use of the instrument for each of its intended uses (i.e., research, program evaluation, and self-evaluation). It is preferable to participate in a training sequence led by an experienced SACERS trainer following the training guide in the SACERS book, pages 38 – 40.
Scoring	Forty-nine items of school-age care environment quality are categorized into seven subscales, each with several items. Items are rated on a 7-point scale from 1 (inadequate) to 7 (excellent). Descriptions are provided at score points 1, 3, 5, and 7. o Space and Furnishings (11 items) o Health and Safety (8 items) o Activities (8 items) o Interactions (9 items) o Program Structure (4 items) o Staff Development (3 items) o Special Needs Supplementary Items (6 items)
Reliability	Inter-rater Reliability Weighted Kappas were calculated for 24 centers, rated independently by two observers. Weighted Kappas for each of the subscales and total score ranged from .79 to .91. Intraclass Correlations were also calculated on 13 centers that were observed by the same two independent observers. These ranged from .87 to .99. Internal Consistency Cronbach's Alphas for each of the subscales and total scores based on 24 classrooms ranged from .67 to .95.

Validity	Validity Information Validity was assessed in two ways: content validity was assessed using expert ratings of each item's importance to their definition of quality; and construct validity was assessed by correlating SACERS total and subscale scores with staff training and staff-to-child ratios. Construct Validity SACERS total and subscale scores were correlated with staff training and staff-to- child ratio. Staff training has moderate positive correlations with Space and Furnishings (r = .31), Interactions (r = .29), Program Structure (r = .40), and Total Score (r = .29). Staff-to-childe ratios have moderate negative correlations with Health and Safety (r = -.40), Activities (r = -.39), Staff Development (r = -.24), and Total Scores (r = -.30). Content Validity Content validity was assessed by asking nine recognized experts from the United States and Canada to rate the importance of each SACERS item to their intuitive definition of high quality on a 5-point scale. (1 = not important to 5 = very important). A mean rating of 4.5 to 5 was found for 91% of the items. The overall mean rating of the items was 4.8. The lowest mean rating assigned to any item was 3.9.
Strengths	The SACERS provides a broad summary of school-age classroom environments that is aligned with a very similar measure to be used in preschool and child care environments (ECERS). It has strong reliability and moderate evidence of validity.
Weaknesses	This assessment does not measure instructional interactions in much detail.
Publication/Pricing	Source: Harms, T., Vineberg Jacobs, E., & Romano White, D. (1996). School – Age Care Environment Rating Scale. New York, NY: Teachers College Press. Publisher: Teachers College Press 1234 Amsterdam Avenue New York, NY 10027 Cost: The cost of a five-day in-depth training is $1225/person. A three-day training costs $825/person. Fees include all materials.

Harms, T., Vineberg Jacobs, E., & Romano White, D. (1996). *School – Age Care Environment Rating Scale*. New York, NY: Teachers College Press

Measure	Student-Teacher Relationship Scale
Constructs	Healthy relationships
Age range	Early childhood and early primary grades (through 3rd grade)
Rating type	Teacher
Description of measure as related to construct of interest	Pianta's scales (Pianta, 1997; Pianta & Nimetz, 1991; Pianta, Steinberg, & Rollins, 1995), yield measures of the child's relationship with his/her teacher, regarding whether the relationship is conflicted, warm, troubled, open, or dependent.
Administration	The scales are quick for teachers to complete, approximately 5 – 10 minutes.
Scoring	Sum Likert ratings across 5-point scales for 28 items, as follows: Closeness = Sum (items 1, 3, 4, 5, 7, 9, 12, 15, 21, 27, 28) Conflict = Sum (items 2, 11, 13, 16, 18, 19, 20, 22, 23, 24, 25, 26) Dependency = Sum (items 6, 8, 10, 14, 17) Total = Sum (Closeness, Recoded/All Items Reversed Conflict, Recoded /All items Reversed Dependency) Reversed/Recoding: 1=5, 2=4, 3=3, 4=2, 5=1 Interpretation of scores is normative, based on percentile range compared to norm groups. Separate norm groups include total population, gender, and three ethnic groups (Caucasian, Hispanic, and African American)
Reliability Validity	Psychometric properties are good. Internal consistencies and test-retest reliability average in the .80s and .90s, except for Dependency scale, which had a lower alpha of .64. Subscale scores are associated with children's classroom and home behaviors (Pianta, Steinberg, & Rollins, 1995). Furthermore, these relationship qualities persist across time and to some extent across teachers. In recent research, STRS scales were negatively related to externalizing behaviors in preschoolers; that is, children with whom teachers report closeness showed less aggression and other out-of-control behavior (Ramos-Marcuse & Arsenio, 2001). Finally, and importantly, scores are moderate predictors of school success through grade 8 (e.g., Pianta, 1997).
Strengths	Relationships with teachers are an important, newly emphasized area of resilience promotion for young children.
Weaknesses	Not necessarily weaknesses, but areas where more attention is needed, include the facts that child age, gender, and ethnicity, as well as teacher-child ethnic match were consistently related to teachers' perceptions.

Publication/Pricing	Published by Psychological Assessment Resources, Inc. Introductory kit w/50 response sheets = $94.00 50 response forms = $60.00

Pianta, R. C. (1997). Adult-child relationship processes and early schooling. *Early Education and Development, 8,* 11-26.

Pianta, R. C., & Nimetz, S. L. (1991). The student-teacher relationship scale: Results of a pilot study. *Journal of Applied Developmental Psychology, 12,* 379-393.

Pianta, R. C., & Steinberg, M. (1992). Teacher-child relationships and the process of adjusting to school. *New Directions for Child Development, 57,* 61-80.

Pianta, R. C., Steinberg, M., & Rollins, K. (1995). The first two years of school: Teacher-child relationships and deflections in children's classroom adjustment. *Development and Psychopathology, 7,* 295-312.

Ramos-Marcuse, F., & Arsenio, W. F. (2001). Young children's emotionally-charged moral narratives: Relations with attachment and behavior problems. *Early Education and Development, 12,* 165-184.

Saft, E. W., & Pianta, R. C. (2001). Teachers' perceptions of their relationships with student: Effects of child age, gender, and ethnicity of teachers and children. *School Psychology Quarterly, 16*, 125-141.

Example Items from the Student-Teacher Relationship Scale

STRS Scale	Example Items
Closeness	I share an affectionate, warm relationship with this child; this child shares information about himself; It is easy to be in tune with what this child is feeling
Conflict	This child and I always seem to be struggling with each other; despite my best efforts, I am uncomfortable with how this child and I have gotten along;
Dependency	This child reacts strong to separation from me; this child is overly dependent on me

Measure	Assessment of Practices in Early Elementary Classrooms (APEEC)
Constructs	Instructional Support
Age range	Elementary
Rating type	Observational, Other
Description of measure as related to construct of interest	"The APEEC was developed to provide a useful tool for both practitioners and researchers who want to understand Elementary school practices (K-3) in general education classrooms serving children with and without disabilities. The APEEC does not measure specific curriculum content or in-depth teacher-child interactions" (Hemmeter, Maxwell, Ault, & Schuster, 2001, p. 1). "The APEEC (Maxwell, McWilliam, Hemmeter, Ault, & Schuster, 2001) measures three domains of classroom practices: physical environment, curriculum and instructional context, and social context."
Administration	According to the authors, these domains can be observed in a typical elementary classroom environment and can be used in classrooms that include children with and without disabilities. Each item has two or more descriptors that characterize the instructional setting and classroom practices (Daly & Dolgos, 2003). Observers are asked to spend a day in the classroom making observation based on the items. A teacher interview with suggested interview questions can be completed to supplement the observations. Administration procedures are not standardized. The APEEC should be administered by individuals knowledgeable about developmentally appropriate practices, early elementary classrooms, and special education practices. Individuals are expected to familiarize themselves with the items and scoring procedures and to read over the administration instructions provided by the authors.
Scoring	The APEEC consists of 16 items covering three broad domains of classroom practices: physical environment, curriculum and instruction, and social context. All items are rated on a seven-point, Likert-type scale. "A score of '1' indicates the classroom is inadequate in terms of developmentally appropriate practices, a score of '3' indicates minimal developmentally appropriate practices, a score of '5' indicates the classroom is good in terms of developmental appropriateness, and a score of '7' indicates excellent developmentally appropriate practices. Intermediate scores of '2', '4', and '6' can also be obtained" (Hemmeter et al., 2001, p. 4). Descriptors are provided at points 1, 3, 5 and 7. After each item is scored, items are summed and then divided by the total number of items administered to generate a total score. Ratings are made using information collected both through classroom observation and teacher interview, with more weight placed on classroom observation.

Reliability	The psychometric report for the measure presents the results from analyses of ratings of 59 classrooms. Interrater agreement at the item level was moderate to high (coefficients ranged from .31 to .81) for exact agreement and ranged from .5 to 1.0 for agreement within 1 point. The median intraclass correlation, an indication of internal consistency, was .86.
Validity	The psychometric report also indicates that the APEEC is moderately correlated with other measures of classroom quality (e.g., the Assessment Profile for Early Childhood Programs), a preliminary indication of validity (Maxwell, et al., 2001). Construct validity was established by comparing the APEEC to several measures of developmentally appropriate practices. Correlations with each scale are presented below: The Assessment Profile for Early Childhood Programs (Abbott-Shim & Sibley, 1988), r = 0.67 • The Teacher Beliefs and Practices Scale (Buchanan, Burts, Bidner, White, & Charlesworth, 1998; Charlesworth, Hart, Burts, Thomasson, Mosley, & Fleege, 1993) • Developmentally appropriate practices, r = 0.55 • Developmentally inappropriate practices, r = -0.28 • The Caregiver Interaction Scale (Arnett, 1989), r = 0.61
Strengths	This measure was designed for use in classrooms serving children with and without disabilities for at least part of the day and thus addresses diversity issues within the classroom. Additionally, classroom diversity is measured through an observation item that has raters assess the degree to which materials and information on diversity are present in the classroom, and the extent to which diversity is discussed or integrated in the classroom and daily activities.
Weaknesses	Analyses did not model the nested data structure despite the inclusion in the data of multiple teachers within some schools (Van Horn & Ramey, 2004). The factor structure of the APEEC has not been assessed (Van Horn & Ramey, 2004). A review in the Fifteenth Buros Mental Measurement Yearbook noted further concerns about this instrument. Reviewers stated that administration procedures for the APEEC are not standardized; test score meaning is unclear, and its psychometric properties are insufficient to recommend it for use (Daly & Dolgos, 2003).
Publication/ Pricing	• Teachers College Press, 1234 Amsterdam Avenue, New York, NY 10027 • Hemmeter, M. L., Maxwell, K. L., Ault M. J., & Schuster J. W. (2001). Assessment of Practices in Early Elementary Classrooms (APEEC). Teachers College Press: New York, NY. The APEEC costs $13.95 (paperback).

Abbott-Shim, M., & Sibley, A. (1988). *Assessment profile for early childhood programs*. Atlanta, GA: Quality Assist.

Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology, 10,* 541-552.

Buchanan, T. K., Burts, D. C., Bidner, J., White, F., & Charlesworth, R. (1998). Predictors of the developmentally appropriateness of the beliefs and practices of first, second, and third grade teachers. *Early Childhood Research Quarterly, 13,* 459-483.

Charlesworth, R., Hart, C. H., Burts, D. C., Thomasson, R. H., Mosley, J., & Fleege, P. O. (1993). Measuring the developmental appropriateness of kindergarten teachers. *Early Childhood Research Quarterly, 8*, 255-276.

Child Trends. (2007, November). *Quality in early childhood care and education settings: A compendium of measures*. Child Trends: Washington, DC.

Chomat-Mooney, L. I., Pianta, R.C., Hamre, B.K., Mashburn, A., Luckner, A.E., Grimm, K.J., Wang, L., Curby, T.W., & Downer, J. (2008). *A practical guide for conducting classroom observations – A summary of issues and evidence for researchers*. Charlottesville: University of Virginia, Center for Advanced Study of Teaching and Learning.

Daly, E. J., & Dolgos, K. A. (2003). Test review of the Assessment of Practices in Early Elementary Classrooms. In B. S. Plake, , J. C. Impara, & R. A. Spies (Eds.), *The fifteenth mental measurements yearbook.* Lincoln, NE: The Buros Institute of Mental Measurements.

Hemmeter, M. L., Maxwell, K. L., Ault M. J., & Schuster J. W. (2001). *Assessment of Practices in Early Elementary Classrooms* (APEEC). Teachers College Press: New York, NY.

Maxwell, K. L., McWilliams, R. A., Hemmeter, M. L, Ault, M. J., & Schuster, J. W. (2001). Predictors of developmentally appropriate classroom practices in kindergarten through third grade. *Early Childhood Research Quarterly, 16*, 431-452.

Van Horn, M. L., & Ramey, S. L. (2004). A new measure for assessing developmentally appropriate practices in early elementary school, A developmentally appropriate practice template. *Early Childhood Research Quarterly, 19,* 569-587.

Section 2: The five SEL Core Competencies (see Table 3):

Constructs in this section include the five SEL core competencies (Zins et al., 2007):

- Self-Awareness
- Self-Management
    - Including both expression and regulation of emotion
        - Internalizing symptoms are included here
    - Including regulation of behavior, but not the entire burgeoning field of self-regulation (e.g., executive function tasks are generally omitted)
    - We realize that internalizing symptoms and externalizing behaviors can also, later in the child's life, be seen as a long-term outcome of lack of SEL
- Social Awareness
    - Including empathy
- Relationship Skills
    - Also including empathy
- Responsible decision-making
    - Including lack of aggression

Some decisions were made about how and where to include various constructs, subconstructs, and their attendant assessment tools, as noted above. We understand that there is some overlap amongst the constructs.

Table 3 summarizes the following measures. Gaps include:

- Fewer responsible decision making scales for the preschool age range – because we elected to omit those that require learning complex coding systems
- Because of children's expressive and cognitive abilities, fewer self and peer ratings of SEL are present for younger age ranges, as well
- For parallel reasons, there are more observational tools at the preschool level

Table 3. Measures, SEL Core Construct Assessed, Age Range, Pre-school or Elementary School, and Rating Type*

Scale Name	SEL Core Competency Assessed					School/Grade Level		Rating Type						
	Self-Awareness	Self-Management	Social Awareness	Relationship Skills	Responsible Decision-Making	Preschool	Elementary School	Teacher	Parent	Peer	Self	Observational	Performance-based	Other (e.g., interview)
Challenging Situations Task (CST)					■	■							■	
Denham's Affect Knowledge Test (AKT)			■			■							■	
The Devereux Early Childhood Assessment (DECA)		■		■		■		■	■					
Minnesota Preschool Affect Checklist		■	■	■		■						■		
Penn Interactive Preschool Play Scales		■		■		■		■	■					
Preschool Self-Regulation Assessment		■				■						■	■	
Self Description Questionnaire for Preschoolers (SDQP)	■					■							■	
Social Competence and Behavior Evaluation		■		■		■		■	■					
Southampton Test Of Empathy For Preschoolers (STEP)			■	■		■					■		■	
Battelle Developmental Inventory (BDI)	■	■		■		■	■	■				■		■

Scale Name	SEL Core Competency Assessed					School/Grade Level		Rating Type						
	Self-Awareness	Self-Management	Social Awareness	Relationship Skills	Responsible Decision-Making	Preschool	Elementary School	Teacher	Parent	Peer	Self	Observational	Performance-based	Other (e.g., interview)
Behavior Assessment System For Children, Second Edition (BASC-2)		■				■	■	■	■		■			
Berkeley Puppet Interview (BPI)	■	■				■	■						■	
Coping with Emotional Situations		■				■	■	■	■					
Emotion Regulation Checklist		■				■	■	■	■					
The Pictorial Scale of Perceived Competence and Social Acceptance for Young Children (PSPCSAYC)	■					■	■		■		■		■	
Positive And Negative Affect Scale (PANAS)		■				■	■		■		■			
Rothbart Temperament Scales – Infant, Early Childhood, Child,		■				■	■	■	■					
Social Skills Rating System, Social Skills Improvement System		■		■	■	■	■	■	■		■			

Scale Name	SEL Core Competency Assessed					School/Grade Level		Rating Type						
	Self-Awareness	Self-Management	Social Awareness	Relationship Skills	Responsible Decision-Making	Preschool	Elementary School	Teacher	Parent	Peer	Self	Observational	Performance-based	Other (e.g., interview)
Sociometric Ratings and Nominations				■		■	■			■	■			
Assessment of Children's Emotional Skills (ACES)			■				■						■	
Behavioral and Emotional Rating Scale-Second Edition: (BERS) Parent Rating Scale (PRS)	■	■	■	■			■	■	■		■			
Bryant Empathy Scale for Children			■	■			■				■			
Child/Teacher/Parent Rating Scale	■			■			■	■	■	■	■			
Children's Emotion Management Scales: Anger And Sadness	■	■					■				■			
Devereux Student Strengths Assessment (DESSA)	■	■	■	■	■		■	■	■					
Emotion Expression Scale For Children (EESC)	■	■					■				■			
Feelings About School (FAS)	■						■				■			

Scale Name	SEL Core Competency Assessed					School/Grade Level		Rating Type						
	Self-Awareness	Self-Management	Social Awareness	Relationship Skills	Responsible Decision-Making	Preschool	Elementary School	Teacher	Parent	Peer	Self	Observational	Performance-based	Other (e.g., interview)
Friendship Quality Questionnaire				■			■			■	■			
How I Feel Scale	■	■					■				■			
Katz-Gottman Regulation Scale		■					■		■					
Kusché Affect Interview -Revised	■	■	■				■							■
Measure Of Prosocial and Aggressive Behavior				■			■	■	■	■	■			
Multidimensional Self-Concept Scale (MSCS)	■						■				■			
Relationship Questionnaire (REL-Q)	■		■	■	■		■				■			
Resiliency Inventory	■	■	■	■			■				■			

* Greyed cells indicate SEL core skills assessed, age level, or rating type

Measure	Challenging Situations Task (CST)
Constructs	Responsible decision-making
Age range	Preschool
Rating type	Performance-based
Description of measure as related to construct of interest	This measure was designed to assess children's affective and behavioral responses to hypothetical peer situations. A challenging situation was defined as one which would elicit affect and test the limits of the child's behavioral abilities within the crucial peer relationship.
Administration	The three situations chosen for inclusion in the CST were: (a) a peer knocking down a tower of blocks which the child was building; (b) being hit by a peer on the playground; and (c) entering a group of peers playing a game. Four categories of affective responses (i.e., happy, sad, angry, and neutral or "just okay"), and four categories of behavioral responses (i.e., prosocial, aggressive, manipulative, and avoidant) were identified for each situation. Prosocial responses included engaging the other person in constructive play, not becoming upset, and discussing the problem. Aggressive responses included yelling, hitting the other person, or destroying the peers' game. Crying and/or pouting were manipulative responses. Avoidant responses were ignoring the other person, withdrawing from the interaction, or waiting on the sidelines. The child was instructed to pretend that he or she was in that situation and to respond to questions as if it were a real situation for them. The tester first presented a 3x4 inch (7.6 x 10.2 cm) picture and verbal description of each challenging situation. Following this presentation of each challenging situation, four pictures of happy, sad, angry, and neutral affect were presented in random order and labeled for the child. Then the child was asked to point to the picture which best described the answer to "How do you feel when [this situation] happens to you?" Next, four pictures of behavioral responses (prosocial, aggressive, manipulation of others' feelings, and avoidant) were presented in random order and the child was asked, "what do you do when you feel that way [in this situation]?"
Scoring	Scores for affective and behavioral responses used were number of times each affect and each behavioral response was chosen by each child, across the three situations.
Reliability	Internal consistency for emotion response and behavioral response in the Warren et al. (2010) sample was $\alpha = .43$ and $\alpha = .54$, respectively. For such a small number of items (3 each), average inter-item correlations can be instructive – for emotion responses, the inter-item average correlation was .20 ($p < .001$) and for behavioral responses, it was .28 ($p < .001$).

Validity	Related to academic competence (Bierman et al, 2008; Warren et al., 2010), behavior problems (Coy et al., 2001). In earlier research, behavioral and affective choices showed relations with emotion knowledge (as measured by Denham's AKT) and with teacher ratings of classroom social behavior. (Zahn-Waxler et al.,, 1994),
Strengths	Portable, developmentally appropriate (e.g., does not ask about ambiguous situations) direct assessment, does not need coding as other measures of social problem solving invariably do, gets at these aspects of responsibility at an earlier age than most other instruments, includes the emotional side of social information processes.
Weaknesses	Needs more research, perhaps more scenarios added for greater internal consistency.
Publisher/Price	Public domain

Bierman, K. L, Domitrovich, C. E., Nix, R, L., Gest, S. D., Welsh, J. A., Greenberg, M. T., Blair, C. Nelson, K., E.& Gill, S. (2008). Promoting academic and social-emotional school readiness: The Head Start REDI program. *Child Development, 79,* 1802-1817.

Coy, K., Speltz, M. L., DeKlyen, M., & Jones, K. (2001). Social-cognitive processes in preschool boys with and without oppositional defiant disorder. *Journal of Abnormal Child Psychology, 29*(2), 107-119.

Denham, S. A., Bouril, B., & Belouad, F. (1994). Preschoolers' affect and cognition about challenging peer situations. *Child Study Journal, 24*, 1-24.

Warren, H. K., Way, E., Kalb, S. C., Denham, S. A., & Bassett, H. H. (2010). *Utilizing Emotion and Behavior for Understanding Preschoolers' Social Information Processing: The Predictive Validity of the Challenging Situations Task.* Submitted for publication.

Zahn-Waxler, C., Cole, P. M., Richardson, D. T., Friedman, R. J., Michel, M. K., & Belouad, F. (1994). Social problem solving in disruptive preschool children: Reactions to hypothetical situations of conflict and distress. *Merrill-Palmer Quarterly, 40*, 98–119.

Measure	Denham's Affect Knowledge Test (AKT)
Constructs	Social Awareness
Age range	30 to 60 months (older limit may be almost an underestimate for low SES children, and is an overestimate for high SES children, who show a ceiling effect around 54 months)
Rating type	Performance-based
Description of measure as related to construct of interest	Denham's Affective Knowledge Test (AKT; 1986) utilizes puppets to measure preschoolers' developmentally appropriate understanding of emotional expressions and situations. Children's understanding of emotion is assessed using puppets with detachable faces that depict happy, sad, angry, and afraid expressions. First, children are asked to both verbally name the emotions depicted on these faces, and then to nonverbally identify them by pointing. This procedure taps into their ability to recognize expressions of emotion. Then, in two subtests of emotion situation knowledge, the puppeteer makes standard facial and vocal expressions of emotions while enacting emotion-laden stories, such as fear during a nightmare, happiness at getting some ice cream, and anger at having a block tower destroyed. Children place on the puppet the face that depicts the puppet's feeling in each situation. In eight situations, the puppet feels emotions that would be common to most people, such as those mentioned above. Finally, children are asked to make inferences of emotions in nonsterotypical, equivocal situations. This subtest measures how well children identify others' feelings in situations where the "other" feels differently than the child. All the situations that the puppeteer depicts during this section of the measure could easily elicit one of two different emotions in different people, as in feeling happy or afraid to get into a swimming pool. Before the assessment, children's parents report, via forced-choice questionnaire, how their children would feel; these responses determine the emotions expressed by the puppet. For example, if the parent reports that the child would be happy to come to preschool, the puppet is depicted feeling sad.
Administration	The AKT is easy to learn and to administer, children enjoy it, and it takes only about 20 minutes to perform; it may be administered across 2 sessions.
Scoring	Children receive 2 points per question for a correct response (using scoring key found in manual); they receive 1 point for getting the valence of the response correct (e.g., if they pick the "sad" rather than "angry" face, since both are negative).
Reliability	Internal consistency and test-retest reliabilities are good (Denham, Caverly, et al., 2002; Denham & Couchoud, 1990a, 1990b), in the .60 - .85 range depending on the specific aggregate of scores created (i.e., total, receptive expression knowledge, expressive expression knowledge, situation knowledge [unequivocal and equivocal]). Dunn, Slomkowski, et al. showed relations of the AKT with later indices of emotion knowledge at age six.

Validity	This measure appears to be especially ecologically valid, as it requires little verbalization and is performed during play. Scores on the AKT are slightly to moderately related to other indices of SEL. For example, researchers have found that children's concurrent AQS attachment ratings are related to scores on the Measure (Denham, Caverly, et al., 2002; Laible & Thompson, 1998); more secure children perform better on the AKT. Moreover, predominantly happier, less angry children also tend to perform better (Denham, 1986; Denham et al., 1990; Denham et al, 2003). Furthermore, AKT scores are related to other indices of SEL, such as moral sensibility and decision-making (Dunn, Brown, & Maguire, 1995), conflicts and interactions with friends (Cutting & Dunn, 1999; Dunn & Herrera, 1997). Finally, AKT scores are both concurrently and longitudinally related to peers' and teachers' evaluations of children's social competence (Denham et al., 1990; Denham et al., 2003). Thus, knowing a child's status on this measure can help investigators not only in knowing about emotion knowledge, but also to prognosticate about skills to which the AKT is related. In fact, it a useful assessment tool to document status and change in emotion knowledge during intervention programming; it has already demonstrated its usefulness in this role (Domitrovich et al., 2002; Shields et al., 2001).
Strengths	Good to excellent psychometrics. Children enjoy it. Emotion knowledge in preschool seems to be pivotal for later social development.
Weaknesses	Not yet standardized but computerized version is planned; needs training. There are other Measures of preschool emotion knowledge, but none seem to have the solid network of research around them.
Publisher/Price	Public domain

Cutting, A.L., & Dunn, J. (1999). Theory of mind, emotion understanding, language, and family background: Individual differences and interrelations. *Child Development, 70,* 853-865.

Denham, S. A. (1986). Social cognition, social behavior, and emotion in preschoolers: Contextual validation. *Child Development, 57,* 194-201.

Denham, S. A., Blair, K. A., DeMulder, E., Levitas, J., Sawyer, K., Auerbach-Major, S. T., & Queenan, P. (2003). Preschool emotional competence: Pathway to social competence? *Child Development, 74,* 238-256.

Denham, S. A., Caverly, S., Schmidt, M., Blair, K., DeMulder, E., Caal, S., Hamada, H., & Mason, T. (2002). Preschool understanding of emotions: Contributions to classroom anger and aggression. *Journal of Child Psychology and Psychiatry, 43,* 901-916.

Denham, S. A., & Couchoud, E. A. (1990a). Young preschoolers' understanding of emotion. *Child Study Journal, 20,* 171-192.

Denham, S. A., & Couchoud, E. A. (1990b). Young preschoolers' understanding of equivocal emotion situations. *Child Study Journal, 20,* 193-202.

Denham, S. A., McKinley, M., Couchoud, E. A., & Holt, R. (1990). Emotional and behavioral predictors of peer status in young preschoolers. *Child Development, 61,* 1145-1152.

Domitrovich, C. E., Cortes, R., & Greenberg, M. (2002, June). Preschool PATHS: Promoting social and emotional competence in young children. Paper presented at the 6th National Head Start Research Conference.

Dunn, J., Brown, J., Beardsall, L. (1991). Family talk about feeling states and children's later understanding of others' emotions. *Developmental Psychology, 27,* 448-455.

Dunn, J., Brown, J., & Maguire, M. (1995). The development of children's moral sensibility: Individual differences and emotion understanding. *Developmental Psychology, 31,* 649-659.

Dunn, J., Brown, J., Slomkowski, Tesla, C., & Youngblade, L. (1991). Young children's understanding of other people's feelings and beliefs: Individual differences and their antecedents. *Child Development, 62,* 1352-1366.

Dunn, J., & Cutting, A. L. (1999). Understanding others, and individual differences in friendship interactions in young children. *Social Development, 8,* 201-219.

Dunn, J., & Herrera, C. (1997). Conflict resolution with friends, siblings, and mothers: A developmental perspective. *Aggressive Behavior, 23,* 343-357.

Dunsmore, J. C., & Karn, M. A. (2001). Mothers' beliefs about feelings and children's emotional understanding. *Early Education and Development, 12*, 117-138.

Laible, D. J., & Thompson, R. A. (1998). Attachment and emotional understanding in preschool children. *Developmental Psychology, 34,* 1038-1045.

Ontai, L. & Thompson, R. A. (2002). Patterns of attachment and maternal discourse effects on children's emotion understanding from 3 to 5 years of age. *Social Development, 11,* 433-450.

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

Note. Please see three other possible emotion knowledge measures for preschoolers/early primary grades by Cassidy, Schultz, and Pons & Harris.

Measure	The Devereux Early Childhood Assessment (DECA)
Constructs	Self-management, Relationship skills, Responsible decision-making, Lack of disruptive behaviors and internalizing symptoms
Age range	24 to 72 months
Rating type	Teacher, Parent
Description of measure as related to construct of interest	The DECA is a newly developed standardized, norm-referenced measure of resilience, completed by parents and teachers in a collaborative and supportive partnership. Subscales include: initiative, attachment, self-control, and behavioral concerns. We like the DECA's subscale demarcation, closely mirroring as it does our notions of SEL.
Administration	Each child is rated one at a time by teachers, parents. Takes 5 – 10 minutes to complete.
Scoring	Items are rated on a 5-point scale varying from "never" to "very frequently." Scoring is shown on answer sheet, as follows: Initiative = Sum (items 2, 3, 7, 12, 16, 19, 20, 24, 28, 32, 36) Self-Control = Sum (items 4, 5, 13, 21, 25, 30, 33, 34) Attachment = Sum (items 1, 6, 10, 17, 22, 29, 31, 37) [Behavioral Concerns = Sum (items 8, 9, 11, 14, 15, 18, 23, 26, 27, 35)] Online administration, scoring and reporting available.
Reliability and Validity	This measure is theoretically and psychometrically sound. Furthermore, its utility is being demonstrated. For example, the total resilience score (i.e., Initiative + Attachment + Self-control) is related to school readiness, as assessed by the Learning Accomplishment Profile-D cognitive and language scales (Devereux Early Childhood Initiative, 2001c). The DECA is already being used to document pre- and post-programming change (Devereux Early Childhood Initiative, 2001a, 2001b).
Any modifications	The Devereux Early Childhood Assessment Scale-Clinical Form (DECA-C) is now available.
Strengths	Use in both applications and research. Resilience viewpoint is a strength, as is the inclusion of attachment. Available in English and Spanish.
Weaknesses	None seen
Publisher/Price	Published by Kaplan Early Learning Co.

	Kaplan prices include $199.95 for a full kit, and 40 record forms for $40.00

Devereux Foundation (2001a) Devereux Early Childhood Initiative Research Bulletin #3. Retrieved from http://www.devereux.org/site/PageServer?pagename=deci_research_bulletins

Devereux Foundation (2001b) Devereux Early Childhood Initiative Research Bulletin #5. Retrieved from http://www.devereux.org/site/PageServer?pagename=deci_research_bulletins

Devereux Foundation (2001c). Research Report #5, Pilot Study Year 2. Retrieved from http://www.devereux.org/site/PageServer?pagename=deci_research_bulletins

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

Example Items from the Devereux Early Childhood Assessment

DECA Subscales	Sample Items
Initiative	Try or asks to try new things or activities
Attachment	Trust familiar adults and believe what they say; seek help from children/adults when necessary
Self-Control	Keep trying when unsuccessful (act persistent); calm herself/himself down when upset
Behavioral Concerns	Destroy or damage property, fight with other children (in DECA-C only)

Note. Item content quoted with permission of the publisher, the Devereux Foundation.

Measure	Minnesota Preschool Affect Checklist (MPAC)
Constructs	Social Awareness, Self-Management, Relationship Skills, (Lack of Disruptive Behaviors and Internalizing Symptoms)
Age range	30-72 months
Rating type	Observational
Description of measure as related to construct of interest	53 items are organized into "mega"-scales for positive and negative affect, inappropriate affect, positive and negative involvement with activities and environment, impulsivity (negative reactions to frustration), positive reactions to frustration, aggression/unusual behaviors, social isolation, peer skills, and empathy/prosocial behavior. Thus, many elements of emotional competence, as well as some elements of social problem solving (e.g., deals with frustration by verbalizing the problem), and numerous relationship skills, are tapped by the MPAC Has been recently shortened by Denham and computerized version is in the works.
Administration	Trained observers watch children's behaviors for 5-minute intervals, noting the presence of items. In previous research, 20 minutes of observation per child across a several-month time period resulted in valid and reliable measurement.
Scoring Reliability	Upon finishing observation periods, observers sum each item across periods, for each "mega"-scale. Previous research has shown good interobserver reliability for "mega"-scales, and concurrent validity with other indices of young children's SEL (Denham, Zahn-Waxler, et al., 1991; Sroufe et al., 1984).
Validity	Several of these scales, notably skills in peer leading and joining, showed change across pre-program to post-program periods, with those showing pre-measure deficits especially benefiting from the program (Denham & Burton, 1996).
Any modifications	Possibly but would need to pilot – maybe fewer items, although these have not been daunting in earlier work.
Strengths	Excellent, detailed profiles of children's social-emotional competence via direct observation.
Weaknesses	Timing for training and obtaining observer reliability is approximately 12 hours. Denham and colleagues (i.e., Denham & Burton, 1996; Denham et al., 1991) have standardized training materials.
Publisher/Price	Public domain.

Denham, S. A., & Burton, R. (1996). A social-emotional intervention for at-risk 4-year-olds. *Journal of School Psychology, 34*, 225-245.

Denham, S. A., Zahn-Waxler, C., Cummings, E. M., & Iannotti, R. J. (1991). Social-competence in young children's peer relationships: Patterns of development and change. *Child Psychiatry and Human Development, 22,* 29-43.

Sroufe, L. A., Schork, E., Motti, F., Lawroski, N., & LaFreniere, P. (1984). The role of affect in social competence. In C. E. Izard, J. Kagan, & R. B. Zajonc (Eds.), Emotions, cognition, & behavior (pp. 289-319). Cambridge: Cambridge University Press.

Table. Items from the Minnesota Preschool Affect Checklist (MPAC)

MPAC "Mega"-Scales	Exemplars of behaviors observed
Expression and regulation of positive affect	Displays positive affect in any manner—facial, vocal, bodily; shows ongoing high enjoyment (30 sec. or more)
Expression and regulation of negative affect	Uses negative affect to initiate contact, to begin a social interaction with someone; uses face or voice very expressively to show negative affect
Inappropriate affect	Expresses negative affect to another child in response to the other's neutral or positive overture; takes pleasure in another's distress
Productive involvement in purposeful activity	Engrossed, absorbed, intensely involved in activity; independent—involved in an activity that the child organizes for himself
Unproductive, unfocused use of personal energy	Wandering; listless; tension bursts
Lapses in impulse control (negative reactions to frustration)	Context-related, physical, interpersonal aggression; inability to stop ongoing behavior; becomes withdrawn
Positive reactions to frustration	Promptly expresses, in words, feelings arising from problem situation, then moves on; shows ability to tolerate frustration well even if does not verbalize
Skills in peer leading and joining	Successful leadership; inept attempts at leadership; smoothly approaches an already ongoing activity
Isolation	No social interaction continuously for 3 minutes or more
Hostility/Aggression	Unprovoked, physical, interpersonal aggression; hazing, teasing, or other provocation or threat
Prosocial response to needs of others	Interpersonal awareness—behavior reflecting knowledge or awareness about another person; helping behavior

Note. General item content adapted from Denham, Zahn-Waxler, et al. (1991), and Sroufe et al. (1984).

Measure	Penn Interactive Preschool Play Scales
Constructs	Self-Management, Relationship Skills
Age range	Preschool
Rating type	Teacher, Parent
Description of measure as related to construct of interest	"Play is an important vehicle for children's social, emotional, and cognitive development, as well as a reflection of their development" (Bredekamp & Copple, 1997 p. 6). Derives information on young children's social competence, in context (Fantuzzo, Sutton-Smith, Coolahan, Manz, Canning, & Debnam, 1995; McWayne, Sekino, Hampton, & Fantuzzo, 2002).
Administration	Informants, whether teachers/caregivers or parents, report on the rate of occurrence of developmentally appropriate behaviors within concrete, observable contexts in which preschoolers are actively engaged in their various play environments. Takes approximately 15 minutes to complete.
Scoring	The PIPPS yields three overarching scales: (1) Play Interaction—i.e., how creative, cooperative, and helpful children are during play; (2) Play Disruption—i.e., how aggressively and antisocially they behave during play; and (3) Play Disconnection—how withdrawn and avoidant children are in contexts where engaged play is more normative. Likert scale items are summed as per manual which is ordered.
Reliability	These scales are internally consistent for both teachers and parents, and appear equally appropriate for low-income children of varying ethnicities, including African American and Hispanic (Fantuzzo et al., 1998; Fantuzzo & McWayne, 2002).
Validity	In terms of validity, parents' PIPPS scales are related to teacher PIPPS scales. As well, positive learning styles, the Social Skills Rating System, conduct problems, emotion regulation, and sociometric acceptance are also related in theoretically expected ways to the scales.
Strengths	The PIPPS offers an advantage to both teachers and parents: because young children's play is so salient a part of their daily activities, informants have ample opportunities to observe it, and are likely to have the skills to understand and reliably complete a measure grounded in this phenomenon. Thus, informants are not required to list or describe behaviors—processes that are open to social desirability and other errors, both systematic and nonsystematic.
Weaknesses	None noted
Publisher/Price	Public domain as far as we know

Bredekamp. S., & Copple, C. (1997). (Eds.). Developmentally appropriate practice in early childhood programs (rev. Ed.) Washington, DC: National Association for the Education of Young Children.

Fantuzzo, J., Coolahan, K., Mendez, J., McDermott, P. & Sutton-Smith, B. (1998). Contextually-relevant validation of peer play constructs with African American Head Start children: Penn Interactive Peer Play Scale. *Early Childhood Research Quarterly, 13*, 411-431.

Fantuzzo, J. W., & McWayne, C. (2002). The relationship between peer-play interactions in the family context and dimensions of school readiness for low-income preschool children. *Journal of Educational Psychology. 94,* 79-87

Fantuzzo, J. W., Sutton-Smith, B., Coolahan, K. C., Manz, P., Canning, S. & Debnam, D. (1995). Assessment of play interaction behaviors in young low-income children: Penn Interactive Peer Play Scale. *Early Childhood Research Quarterly, 10*, 105-120.

McWayne, C., Sekino, V., Hampton, G., & Fantuzzo, J. (2002). Manual: Penn Interactive Peer Play Scale. Teacher and parent rating scales for preschool and kindergarten children. Philadelphia: University of Pennsylvania.

Measure	Preschool Self-Regulation Assessment (PSRA)
Constructs	Self-Management
Age range	Preschool
Rating type	Performance-based, observational
Description of measure as related to construct of interest	Several observational tasks that show good value as denoting children's ability to regulate emotions have been identified by Kochanska and colleagues. Radiah Smith-Donald and Cybele Raver are now piloting a very clear, detailed use of such tasks, with specific measurement of the child's performance both quantitatively and qualitatively.
Administration	Direct assessment. Over several short (e.g., ~ 5 minute) tasks for each child individually. Coding is done by trained observer as the testing proceeds.
Scoring	See Raver's "Emotion Matters" coding sheets and script
Reliability	Shown to be good by Kochanska and colleagues, as well as Smith-Donald and colleagues (2007). (see below).
Validity	
Any modifications	Picking the ones deemed most valuable.
Strengths	Actual observation of children; tested methodologies
Weaknesses	Takes time (but no videotape); requires training
Publisher/Price	Public domain

Kochanska, G., Murray, K. T., & Harlan, E. T. (2000). Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. *Developmental Psychology, 36*, 220-232.

Kochanska, G., Murray, K., Jacques, T. Y., Koenig, A. L., & Vandegeest, K. A. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Development, 67*, 490-507.

McClelland, M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology, 43*, 947-959.

Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989). Delay of gratification in children. *Science, 244*, 933-938.

Murray, K. T., & Kochanska, G. (2002). Effortful control: Factor structure and relation to externalizing and internalizing behaviors. *Journal of Abnormal Child Psychology, 30*, 503-514.

Smith-Donald, R., & Raver, C. Emotion Matters Protocol. Unpublished manuscript, University of Chicago.

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

Measure	Self-Description Questionnaire for Preschoolers (SDQP)
Constructs	Self-Awareness
Age range	Preschool
Rating type	Performance-based
Description of measure as related to construct of interest	The SDQP was designed to assess two areas of academic self-concept (Verbal and Math) and four areas of nonacademic self-concept (Physical, Appearance, Peers, and Parents). for preschool children (preschool children do not clearly understand General self-concept items; see also Harter & Pike, 1984) This measure is a downward extension of the SDQ–I (Marsh, Craven, & Debus, 1991), an excellent instrument in terms of psychometric properties and construct validation (see, e.g., Byrne, 1996).
Administration	Some researchers promote the use of nonverbal methods (e.g., puppets, pictures) for assessing young children's self-concept (Eder, 1990; Harter & Pike, 1984). Children can also respond to simple, direct questions, to provide reliable information on the self (Marsh et al., 1991; Marsh, Craven, & Debus, 1998). Administration of the SDQP thus uses only verbal presentations (see table for items). Items are worded in question format (e.g., "Can you run fast?") rather than the declarative format (e.g., "I can run fast") to reduce the linguistic complexity that young children face when they are required to verify declarative statements.
Scoring	38 items were developed to represent six self-concept scales on the SDQP. Scoring follows from answers to probes as follows: (1) No always; (2) No sometimes; (3) Yes sometimes; and (4) Yes always (see also Harter & Pike, 1984).
Reliability	Internal consistency evaluated via coefficient ω range from .75 for Math to .89 for Physical and Appearance scales. The authors suggest these are reasonable based on the brevity of the scale and age of participants.
Validity	There is support for construct validity the academic self-concept and for the separation of academic and nonacademic components of self-concept; young children do appear to distinguish amongst these aspects of self. In terms of concurrent validity, mathematics achievement (via WRAT scores) was significantly and positively correlated with Math and Verbal self-concept scores. However, verbal achievement was positively but nonsignificantly correlated with these same self-concept scores. Nonetheless, verbal achievement was significantly more highly correlated with the academic self-concept factors than with the

	nonacademic self-concept factors.
Strengths	Because children have difficulty responding appropriately to negatively worded items (Marsh, 1986), all of the SDQP items were worded positively. Covers theoretically important areas of self-perceived competence/self-concept with promising psychometric properties
Weaknesses	Further research needed to replicate the factors and substantiate not discriminating between math and verbal constructs with young children. Further, it would not appear as yet a useful measure for making decisions about individual children, although it can describe groups of children well.
Publisher/Price	Public domain, as far as we know

Scales and Sample Items from the SDQP

Self-Concept Area	**Sample Questions**
Physical Ability (6 items)	Can you run fast? Do you enjoy sports and games?
Appearance (6 items)	Are you good looking? Do you like the size and shape of your body?
Peer Relations (6 items)	Do you have lots of friends? Do you play with lots of kids at preschool?
Parent Relations (8 items)	Do your parents like you? Do you enjoy doing things with your parents?
Verbal (6 items)	Do you know lots of letters of the alphabet? Do you know lots of different words?
Math (6 items)	Do you like playing number games? Are you good at counting?

Byrne, B. (1996). *Measuring self-concept across the life span: Issues and instrumentation.* Washington, DC: American Psychological Association.

Eder, R. A. (1990). Uncovering young children's psychological selves: Individual and developmental differences. *Child Development, 61,* 849–863.

Harter, S. & Pike, R. (1984). The pictorial scale of perceived competence and social acceptance for young children. *Child Development, 55*, 1969-1982.

Marsh, H. W. (1986). Negative item in rating scales for preadolescent children: A cognitive-developmental phenomenon. *Developmental Psychology, 22,* 37–49.

Marsh, H. W., Craven, R. G., & Debus, R. L. (1991). Self-concepts of young children aged 5 to 8: Their measurement and multidimensional structure. *Journal of Educational Psychology, 83,* 377–392.

Marsh, H. W., Craven, R. G., & Debus, R. L. (1998). Structure, stability, and development of young children's self-concepts: A multicohort–multioccasion study. *Child Development, 69,* 1030–1053.

Marsh, H. W., Ellis, L. A., & Craven, R. G. (2002). How do preschool children feel about themselves? Unraveling measurement and multidimensional self-concept structure. *Developmental Psychology, 38*, 376-393.

Measure	Social Competence and Behavior Evaluation - 30
Constructs	Self-Management, Relationship Skills, Responsible decision-making, (Lack of Disruptive Behaviors, Internalizing Symptoms)
Age range	30 to 78 months
Rating type	Teacher, Parent
Description of measure as related to construct of interest	"….scale developed to assess patterns of social competence, emotion regulation and expression, and adjustment difficulties in children ages 30 to 78 months" (LaFreniere & Dumas, 1996, p. 369) The short form, derived from the 80-item form now published by Western Psychological Services, has been widely used in research. The short form of the SCBE is intended to preserve several important strengths of the original while reducing its overall length. Both have been extensively normed with stratified samples of French Canadian and American preschoolers. The measure also has been translated into Spanish (Dumas et al., 1998a, 1998b). Main subscales include Anger/Aggression, Anxiety/Withdrawal, and Cooperation/ Sensitivity.
Administration	This teacher-report measure has both a long and a short 30-item version, as well as a parent short version (Kotler & McMahon, 2002). The informant gives each child a rating from 1 to 6 on items like those shown below. The item content allows evaluations to be completed by anyone who knows the child well (i.e., informants are able to complete the questionnaire after they have known the child for at least two months). Responses of experienced teachers tend to be distributed differently from inexperienced teachers, which is an issue to take into account generally when using data from teacher reports. The standardization samples are not large but considered adequate. The measure was not developed strictly for clinical use, though it correlates with the Child Behavior Check List. Teachers and parents can complete the short version in 10 minutes.
Scoring	Sum ratings for the three scales: Anger/Aggression = Item1 + Item 4 + Item 7 + … + Item 28 Cooperation/Sensitivity = Item 2 + Item 5 + Item 8 + … + Item 29 Anxiety/Withdrawal = Item 3 + Item 6 + Item 9 + … + Item 30 In past work (e.g., Denham et al., 2003) has also shown that a standardized aggregate (i.e., $z_{sensitive/cooperative} - z_{angry/aggressive} - z_{anxious/withdrawn}$) can be used reliably for an overall measure of social competence.

Reliability	This questionnaire has been widely used, and its reliability has been established within diverse cultures (LaFreniere & Dumas, 1996; LaFreniere, et al. 2002). Interrater agreement across teachers is uniformly high, with high internal consistency of the scales, with test-retest reliability across 2 weeks very high and 6 months slightly lower. The parent form scales have excellent internal consistency as well. AA and AW scales are relatively orthogonal, with social competence negatively correlated with both other scales, for both reporters.
Validity	Factor structure of the short version is very clear. 10-item scales from the original SCBE standardization sample were computed and correlate highly with the original scales. For 517 children in Indiana, teacher ratings of conduct disorder and anxiety-withdrawal were obtained from the Revised Behavior Problem Checklist; the concomitant scales of the SCBE-30 were highly correlated with this measure. Kotler and McMahon (2002) show that the three parent scales differentiate passive noncompliant, simple noncompliant, and negotiated noncompliant behaviors in preschool children.
Strengths	Several advantages over broadband assessments of externalizing and internalizing behavior obtain here, including: (a) the orthogonality of AA and AW scales; (b) inclusion of strength-based scale; and (c) clearer to social-emotional developmental tasks in item content.
Weaknesses	None noted
Publisher/Price	LONG version published by Western Psychological Services They request option to "approve" use of short form, which is only 30 items and well validated (LaFreniere & Dumas, 1996) LONG version prices are $82.50 for a kit with manual and 25 scoring forms LONG version price for 25 scoring forms is $39.95

Example Items from the Social Competence/Behavior Evaluation (SCBE)

SCBE Scale	Sample Items
Aggression	Gets into conflicts with other children; opposes the teacher
Withdrawal	Doesn't talk or interacting during group activities; avoids new situations
Cooperation/Sensitivity	Negotiates solutions to conflicts (note social problem-solving content); cooperates with other children

Denham, S. A., Blair, K. A., DeMulder, E., Levitas, J., Sawyer, K. S., Auerbach-Major, S. T., & Queenan, P. L. (2003). Preschoolers' emotional competence: Pathway to mental health? *Child Development, 74*, 238-256.

Dumas, J. E., Martinez, A., & LaFreniere, P. J. (1998a). The Spanish version of the Social Competence and Behavior Evaluation (SCBE)—Preschool edition: Translation and field testing. *Hispanic Journal of Behavioral Sciences. 20,* 255-269.

Dumas, J. E., Martinez, A., LaFreniere, P. J., & Dolz, L. (1998b). La versión española del Cuestionario "Evaluación de la Conducta y la Competencia Social" para preescolares (Social Competence and Behavior Evaluation SCBE): Adaptación y validación. / Spanish Version of the Social Competence and Behavior Evaluation—Preschool Edition (SCBE): Adaptation and validation. *Psicologica. 19,* 107-121

Kotler, J. C., & McMahon, R. J. (2002). Differentiating anxious, aggressive, and socially competent preschool version: Validation of the Social Competence and Behavior Evaluation-30 (parent version). *Behaviour Research and Therapy, 40,* 947-959.

LaFreniere, P. J., & Dumas, J. E. (1996). Social competence and behavioral evaluation in children ages 3 to 6 years: The short form (SCBE-30). *Psychological Assessment, 8*, 369-377.

LaFreniere, P, Masataka, N., Butovskaya, M., Chen, Q., Dessen, M. A., Atwanger, K., Schreiner, S., Montirosso, R., & Frigerio, A., (2002). Cross-cultural analysis of social competence and behavior problems in preschoolers. *Early Education and Development*, *13, Special issue: Cultural perspectives on social competence in early childhood,* 201-219.

Measure	Southampton Test of Empathy for Preschoolers (STEP)
Constructs	Social Awareness, Relationship Skills
Age range	Preschool
Rating type	Self-rating, performance-based
Description of measure as related to construct of interest	In STEP, the role of affective and cognitive perspective taking in empathy is considered. Assesses a child's ability to understand and share in the experience of another person across a number of hierarchically organized, emotion judgment contexts linked to facial expression, situational cues, verbal cues, and desires. STEP incorporates computer–presented, videotaped vignettes of children in emotional scenarios that focus on four emotional outcomes (angry, happy sad, fearful). It asks children to indicate their reactions to vignettes by selecting a picture of the relevant facial expression. The test incorporated 8 video vignettes of children in emotional scenarios, assessing a child's ability to understand (STEP-UND) and share (STEP-SHA) in the emotional experience of a story protagonist. Each vignette included 4 emotions (angry, happy, fearful, sad) that reflected emotion judgments based on the protagonist's facial expression, situation, verbal cues, and desire. The test incorporates eight video vignettes of children in emotionally evocative scenarios to assess a child's ability to understand (STEP-UND) and share (STEP-SHA) in the emotional experience of a story protagonist. Each child watched one practice story and eight test stories. The videos used continuous movement and child actors. Four stories had a male protagonist, and four had a female protagonist. Each story was made up of seven consecutive parts. In Parts 1, 2, 4, and 6 children were asked to judge how the protagonist (STEP-UND) and they themselves (STEP-SHA) would feel, on the basis of the protagonist's facial expression, a situation cue, a verbal cue, and the protagonist's desire. Each story part related to one of four emotional outcomes (angry, happy, sad and fearful), such that emotion judgments (related to facial expression, situation cues, verbal cues, and desire) were linked twice with each outcome. Desires were represented by a thought bubble (Wellman, Hollander, & Schult, 1996). Two further story parts (3 and 5) contained check questions that were designed to assess concentration and understanding. The story ending made up the final part (see Table 1). Each story was accompanied by an 85–90 word narrative and lasted approximately 120 s. Children indicated their emotion judgments by

	clicking on one of five schematic emotion faces (happy, sad, angry, fearful, OK/neutral) displayed at the bottom of the computer screen. After children watched the practice story, a computer screen appeared, featuring the eight story protagonists waving. Children had to click on successive characters in order to hear each story, and stickers were given between stories. This process was repeated until children had completed all eight vignettes, and a goodbye screen was presented.
Administration	Video vignettes and computerized responses
Scoring	The program generated two scores that represented the children's ability to identify the protagonists' emotions (STEP-UND; score 0, 1 or 2) and the children's tendency to share in these emotional experiences (STEP-SHA; score 0, 1 or 2). A score of 2 represented an accurate judgment of a protagonist's emotion (STEP-UND) or a match between the character's emotion and the child's judgment of his or her own emotion (STEP-SHA). A score of 1 was given for sad, angry, and fearful judgments when children responded with any other negative emotion. Zero represented an inaccurate judgment of a character's emotion or no match between the character's emotion and a child's judgment of his or her own emotion. Both STEP-UND and STEP-SHA had a minimum score of 0 (not empathic) and a maximum score of 64 (highly empathic; 8 stories, each with 4 emotion judgments).
Reliability	The results showed good internal consistency; Cronbach's alpha reliability coefficients for STEP-UND and STEP-SHA were .70 and .86, respectively. They also highlighted moderate concurrent validity with parent-rated empathy, a measure of facial indices, and construct validity with teacher-rated prosocial behavior.
Validity	Considering concurrent validity, children's ability to understand the perspective of others (STEP-UND) and to respond to them appropriately (STEP-SHA) was positively associated with parent report dispositional empathy.
Strengths	Links with parent-report empathy and facial responding support the proposition that STEP taps empathic responsiveness in preschool children.
Weaknesses	The small sample size and homogeneity of the sample characteristics limits the generalization of these findings. Concurrent validity was only moderate with existing measures of empathy, highlighting the difficulty in developing a valid and reliable instrument to measure emotional responsiveness in this age group. Furthermore, the absence of test–retest reliability limits our understanding of the stability of emotional perspective taking and empathy in this age group.
Publisher/Price	Public domain

Howe, A., Pit-ten Cate, I. M., Brown, A. & Hadwin, J. A. (2008). Empathy in preschool children: The development of the Southampton Test of Empathy for Preschoolers (STEP). *Psychological Assessment 20*, 305–309.

Wellman, H. M., Hollander, M., & Schult, C. A. (1996). Young children's understanding of thought bubbles and of thoughts. *Child Development, 67*, 768-788.

Measure	Battelle Developmental Inventory (BDI), 2nd Edition
Constructs	Self-Awareness, Self-Management, Relationship Skills
Age range	Birth to 7 years, 11 months
Rating type	Teacher, Observational, Other
Description of measure as related to construct of interest	Items on adult interaction, expression of feelings/affect, self-concept, peer interaction, coping, and social role are included. For example, topic areas include showing appropriate affection toward people, pets, or possessions, using adults appropriately to help resolve peer conflict, recognizing the feelings of others, and recognizing the basic similarities of all children.
Administration	The BDI is designed to be used by teachers, diagnosticians, and multidisciplinary teams. Its authors consider it useful for screening and or for more in-depth assessment of specific nonhandicapped or handicapped children's strengths and weaknesses for programming, as well as to help demonstrate the effects of programming. Use of a transdisciplinary assessment format also is possible. If administering only the Personal-Social domain, from which the aforementioned items are extracted would take about 15 minutes to complete – mixture of interview, observation, etc.
Scoring	Because some items can be scored via interview or observation methods, or structured format items can be corroborated via these methods, it behooves the examiner to gather all possibly relevant data on Personal-Social items before scoring. Items are scored from each subscale noted above, and summed. Raw scores can be converted to age norms, percentiles, or T-scores, for example.
Reliability	Excellent reliability data for the last version (before the current revision) are reported, with very small standard errors of measurement and high test-retest reliabilities.
Validity	As yet not reported for revision? Substantial for earlier versions.
Strengths	The BDI's comprehensiveness, standardized test scores, empirically based age placement of its approximately 130 items, behaviorally anchored item descriptions, and improved, easier administration and scoring, all maximize its usefulness. Each skill item chosen for the new revision for the Personal-Social (and all other domains) has gone through a rigorous process of judgment on how critical or important it is to a child's development.
Weaknesses	Care needs to be taken regarding observation, structured items. The examiner also must apply, when there are disagreements among these sources of data, a standard set of decision rules. Although these are not necessarily weaknesses, they are complications.

Publisher/Price	Published by Riverside Publishing Full kit with manipulatives and software = $932.00 (this includes other domains than just personal-social)

Newborg, J. (2005). *Battelle Developmental Inventory, Second Edition.* Itasca, IL: Riverside Publishing.

Measure	Behavior Assessment System for Children, Second Edition (BASC-2)
Constructs	Self-Management, Responsible decision-making, (Lack of Risky, Disruptive/Externalizing Behaviors and Internalizing Symptoms)
Age range	Four to 18 years of agec
Rating type	Teacher, Parent, Self
Description of measure as related to construct of interest	This measure rates adaptive and problem behavior. Divided into four composites (Externalizing Problems, Internalizing Problems, School Problems, and Adaptive Skills) and 15 subscales. BASC-2 TRS Externalizing Problems composite reflects child's overall disruptive behavior symptoms and includes: Aggression, Conduct Problems, and Hyperactivity
Administration	Teachers, parents, observers, and students rate student's behavior; self-report scales go down to age 6-7 now.
Scoring	139 items; rating of adaptive and problem behavior by frequency: Ranging from 0 (Never) to 3 (Almost always) Five components (Structured Developmental History, Parent Rating Scale, Teacher Rating Scale, Self-Report of Personality and Student Observation System, which can be used separately or in combination. Designed to assist in the differential diagnosis of emotional and behavioral problems, educational, classification, treatment planning and evaluation and for use in research (Reynolds and Kamphaus, 1992, 1998).
Reliability	Test-retest reliability estimates (8-59 days between ratings) indicate strong score stability over time. Internal consistency estimates were above .90 across scales. Interrater reliability estimates indicate moderate consistency in ratings between teachers.
Validity	The BASC has been the test of choice to substantiate the convergent validity of a number of scales, with findings generally in the expected directions, for example such that: BASC TRS Correlated with The Revised Class Play (RCP; Masten et al., 1985), a peer assessment of social reputation (Realmuto et al., 1997). The BASC scales and composites (PRS and TRS) were correlated with the Social Skills and Problem Behaviours scales on the Social Skills Rating Scales (SSRS; Gresham and Elliot, 1990) for the purpose of establishing convergent validity for both (Flanagan et

	al., 1996). When the SSRS Social Skills scale was used as the criterion, convergent validity evidence was identified for the parent form of the BASC Social Skills scale, but the same was not found for the teacher form. Correlations between the BASC Adaptive Skills Composite and the SSRS Social Skills scales were in the expected directions, as were the correlations between the problem scales of both the BASC and SSRS. The authors concluded that, generally, these results provided preliminary convergent validity.
Strengths	Advantages of the BASC over the CBCL include, among others, the existence of separate forms for primary and adolescent age groups and the separation of the Anxiety and Depression scales. Its strength-based scales are vastly better (see note).
Weaknesses	Some authors, for example, Flanagan (1995), have seen the absence of norms for minority groups, as a limitation. Adams and Drabman (1994) caution against the use of clinical norms for female children or children presenting with internalizing difficulties as both were underrepresented in the clinical norm group. Vaughn et al. (1997) found that neither the CBCL nor the BASC was advantageous with respect to differentiating children with ADHD from those who do not meet criteria for ADHD, except for the BASC TRS which had better predictive ability for differentiating children who do not meet ADHD criteria.
Publisher/Price	Pearson Assessment, http://www.pearsonassessments.com/basc.aspx $500+ for starter kit in English and Spanish

Note. There are now Anger Control, Bullying, Executive Function, Emotional Self-Control, Negative Emotionality Scales, and Resiliency Scales.

Adams, C.D. & Drabman, R.S. (1994) 'BASC: A Critical Review', *Child Assessment News, 4,* 1–5.

Bergeron, R., Floyd, R. G., McCormack, A. C., & Farmer, W. L. (2008) The generalizability of externalizing behavior composites and subscale scores across time, rater, and instrument. *School Psychology Review, 37*(1), 91-108.

Flanagan, R. (1995) A review of the Behaviour Assessment System for Children (BASC): Assessment consistent with the requirements of the Individuals with Disabilities Education Act (IDEA), *Journal of School Psychology, 33,* 177–86.

Flanagan, D.P., Alpfonso, V.C., Primavera, L.H., Povall, L. & Higgins, D. (1996) Convergent validity of the BASC and SSRS: Implications for Social Skills Assessment, *Psychology in the Schools, 33*, 13–23.

Greenspoon, P.J., & Saklofske, D.H. (1997) Validity and reliability of the Multidimensional Students' Life Satisfaction Scale with Canadian children, *Journal of Psychoeducational Assessment 15,* 138–55.

Huebner, E.S. (1994) 'Preliminary development and validation of a Multidimensional Life Satisfaction Scale for Children', *Psychological Assessment, 6,* 149–58.

Masten, A., Morison, P. and Pellegrini, D. (1985) 'A Revised Class Play Method of Peer Assessment' *Developmental Psychology, 21,* 523–33.

Realmuto, G.M., August, G.J., Sieler, J.D. & Pessoa-Brandao, L. (1997) 'Peer assessment of social reputation in community samples of disruptive and nondisruptive children: Utility of the revised class play method', *Journal of Clinical Child Psychology, 26,* 67–76.

Reynolds, C.R. & Kamphaus, R.W. (1992) Behaviour Assessment System for Children Manual. Circle Pines, MN: American Guidance Service.

Reynolds, C.R. & Kamphaus, R.W. (1998) Behaviour Assessment System for Children Manual – Second Edition. Circle Pines, MN: American Guidance Service.

Vaughn, M., Riccio, C.A., Hynd, G.W. & Hall, J. (1997) 'Diagnosing ADHD (Predominately Inattentive and Combined Type Subtypes): Discriminant validity of the Behaviour Assessment System for Children and the Achenbach Parent and Teacher Rating Scales', *Journal of Clinical Child Psychology, 26*, 349–57.

Measure	Berkeley Puppet Interview (BPI)
Constructs	Self-awareness
Age range	Preschool through early elementary
Rating type	Performance-based
Description of measure as related to construct of interest	The BPI was developed to address the absence of standardized methodologies appropriate for measuring young children's perceptions of themselves and their environments. It blends structured and clinical interviewing methods, using the puppets, Iggy and Ziggy. For example: Iggy: I have lots of friends. Ziggy: I don't have lots of friends. Iggy: How about you? Ziggy: My parents' fights are about me. Iggy: My parents' fights are not about me. Ziggy: How about your parents? Use of puppet play allows children to respond naturally and comfortably.
Administration	Direct Assessment. Most of the BPI subscales consist of 4 to 8 items. Most successful interviewing lasts no longer than 20 minutes before a break; 2 such periods can be used in one day.
Scoring	Self-Perception Scales include Academic Competence, Achievement Motivation, Social Competence, Peer Acceptance, Depression/Anxiety, and Aggression/Hostility. The authors ask that each scale be used in its entirety (Measelle et al., 1998). The methodology is flexible, and allows for other scales, including Teacher Closeness, Teacher Conflict, and School Engagement, and parent-child subscales such as Warmth and Enjoyment, Anger & Hostility, Responsiveness, Emotional Availability, Limit-Setting, and Autonomy Granting & Control (separate scales for mother and father).
Reliability	Excellent
Validity	The BPI's utility has been tested on socioeconomically, culturally, and clinically diverse samples. A Spanish language version is being tested in Chile. Children understand the questions and become engaged in dialogue with the puppets. Agreement between young children and adult informants is as strong if not stronger than that between pairs of adult informants. These data are important in light of the field's tendency to view young children's perception as less valid.

Strengths	Allows for verbal or nonverbal, elaborated or limited responses. Coding manual helps coder to decipher figures of speech, reasoning processes, and conditional responses that reflect ambivalent self-perceptions or uncertainty due to lack of experience. May also be used to examine areas of self-awareness parallel to adult "Big Five" dimensions of personality (Measelle et al., 2005).
Weaknesses	Requires 2 ½ day training workshop and certification, which also takes time. Price of workshop from University of Oregon personnel, including training and reliability checks, equals $900 per trainee.

Measelle, J., R., Ablow, J. C., Cowan, P. A., & Cowan, C. P. (1998). Assessing young children's self-perceptions of their academic, social, and emotional lives: An evaluation of the Berkeley Puppet Interview. *Child Development, 69,* 1556-1576.

Measelle, J. R., John, O. P., Ablow, J. C., Cowan C. P. Cowan, P. A. (2005). Can children provide coherent, stable, and valid self-reports on the Big Five dimensions? A longitudinal study from ages 5 to 7. *Journal of Personality and Social Psychology, 89*, 90-106

Measure	Coping with Emotional Situations
Constructs	Self-management
Age range	Preschool through Middle Childhood
Rating type	Teacher, Parent
Description of measure as related to construct of interest	Children's coping behavior when faced with emotional situations with peers can be assessed with items developed by Eisenberg, Fabes, Nyman, Bernzweig, and Pinuelas (1994). Item content reflects: Instrumental Coping (e.g., taking action to improve a situation), Instrumental Aggression (e.g., hitting), Emotional Intervention (e.g., crying to elicit help), Avoidance (e.g., leaving a problem), Distraction (e.g., keeping busy), Venting (e.g., crying to release frustration), Emotional Aggression (e.g., aggressing to release frustration), Cognitive Restructuring (e.g., saying "I don't care"), Cognitive Avoidance (e.g., not thinking about the problem), Instrumental Intervention (e.g., getting help), Instrumental Support (e.g., talking to someone about the problem), and Denial (e.g., saying nothing happened). Based on the work of Eisenberg and colleagues (1994), data can be reduced to three summary scales: Emotional Venting (e.g., cries to release feelings/get help, solves problems/releases feelings through aggression). Constructive Strategies (e.g., getting emotional support or pragmatic assistance with the problem; solving the problem) Avoidant Strategies (e.g., using distraction, denying the problem)
Administration	Informants indicate on a 7-point scale, with 1 indicating "never" and 7 indicating "usually," how often the child would engage in each of 12 general types of coping behavior when confronted with a problem situation. Takes less than 10 minutes.
Scoring	Sum 1 – 7 rating across rational subscales, as follows: Constructive coping = Σ items 1, 7, 10, 12, 12 Venting coping = Σ items 2, 3, 8, 9 Passive coping = Σ items 4, 5, 6, 13 It is, however, possible that factors may differ across ages.
Reliability	Good test-retest and internal consistency reliability have been reported by the authors and Denham. Specifically, α's for these scales in Denham's research have been good to excellent, especially considering the number of items per subscale.
Validity	Research suggests that parent's completion of these scales is related to teacher's evaluations of young children's social competence.

Strengths	These scales are very quick and easy for either teachers or parents to complete. They yield a snapshot of the child's emotional coping strategies, and would thus be useful for individualizing programming as well as pre- and post-programming measurement.
Weaknesses	Coping may not be considered the "same as" emotion regulation.
Publisher/Price	Public domain as far as we know

Eisenberg, N., Fabes, R. A., Nyman, M., Bernzweig, J., & Pinuelas, A. (1994). The relation of emotionality and regulation to preschoolers' anger-related reactions. *Child Development, 65*, 1352-1366.

Measure	Emotion Regulation Checklist
Constructs	Self-Management
Age range	Preschool and early primary
Rating type	Teacher, Parent
Description of measure as related to construct of interest	The 24-item Emotion Regulation Checklist taps both prevalent emotional expressiveness and the product aspect of emotion regulation: that is, it targets processes central to emotionality and regulation, including affect lability, intensity, valence, flexibility, and contextual appropriateness of expressiveness (Shields & Cicchetti, 1997; Shields et al., 2001). The Lability/Negativity subscale is comprised of items representing a tack of flexibility, mood lability, and dysregulated negative affect; sample items include "Exhibits wide mood swings," and "Is prone to angry outbursts?' The Emotion Regulation subscale includes items describing situationally appropriate affective displays, empathy, and emotional self-awareness; sample items include "Is empathic toward others," and "Can say when s/he is feeling sad, angry or mad, fearful or afraid."
Administration	Other-report questionnaire (teachers, could be adapted for parents); takes about 10 minutes or less to complete.
Scoring	Sum 4-point Likert ratings for scales Lability/Negativity Scale= Σ items 2, 6, 8, 10, 13, 14, 17, 19, 20, 22, 24, 4R, 5R, 9R, 11R Emotion Regulation Scale = Σ items 1, 3, 7, 15, 21, 23, 16R, 18R R= recode, which in this case denotes subtracting that score from the total Item 12 is not included in either scale as it did not load on either in early validation studies.
Reliability	Internal consistency for the emotion regulation and lability/negativity subscales is excellent; in Shields & Cicchetti (1997) alphas were.96 for lability/negativity and .83 for Emotion Regulation.
Validity	In terms of validity, the measure distinguishes well regulated from dysregulated children. More specifically, overall emotion regulation at the start of the preschool year was associated with school adjustment at year's end, whereas early emotional lability/negativity predicted poorer outcomes.
Any modifications	No
Strengths	Good content validity, important constructs assessed

Weaknesses	At times it is concerning that expression and regulation are so confounded in this measure – this is an endemic problem in the literature as a whole. However, the ability/negativity subscale is at least separate.
Publisher/Price	Public domain as far as we know

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

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

Example items from the Emotion Regulation Checklist

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

Measure	The Pictorial Scale of Perceived Competence and Social Acceptance for Young Children (PSPCSAYC) Also: The Perceived Competence Scale for Children (PCSC)
Constructs	Self-awareness
Age range	Preschoolers and Kindergarteners for one scale, 1st and 2nd graders for another. PCSC is for 3rd through 9th grades
Rating type	Performance-based, self
Description of measure as related to construct of interest	PSPCSAYC is a pictorial scale of perceived competence and social acceptance for young children, tapping 4 domains with 6 items each: cognitive competence, physical competence, peer acceptance, and maternal acceptance. Items differ somewhat for some scales across the two age levels. For For the PSPCSAYC, factor analyses reveal a 2-factor solution. The first factor, general competence, is defined by the cognitive and physical competence subscales. The second factor, social acceptance, comprises the peer and maternal acceptance subscales. It is urged that this instrument not be viewed as a general self-concept scale but be treated as a measure of 2 separate constructs: perceived competence and social acceptance. However, the PCSC has factors of cognitive, social, physical, and general self-worth.
Administration	Direct Assessment for PSPCSAYC, self-report for PCSC <u>PSPCSAYC</u> Picture plates —The pictures accompanying each version are bound separately, as are sets for boys and girls. Sample item — The child is first read a brief statement about each child depicted. For the sample item, the female subject would be told that the girl on the child's left is good at puzzles but the child on the right is not very good at puzzles. The child's first task is to indicate which of the two girls she is most like. After making that decision, the child is then asked to think only about the picture on that side and indicate whether she is a lot like that girl (the big circle) or just a little bit like that girl (the smaller circle). For each item there are more specific descriptive questions that accompany each circle, such as "Are you just pretty good at puzzles [small circle] or really good [large circle]?" <u>PCSC</u> Similar items with four point scale – two choices on one side of the item for “true for me – really or sort of” and two choices on the other side of the item for “true for me –really or sort of.”

Scoring	Each item is scored on a four point scale, where a score of 4 would be the most competent or accepted and a score of 1 would designate the least competent or accepted. Thus, for the PSPCSAYC sample item, the child who indicates that she is a lot like the girl who is good at puzzles would receive a score of 4. If she chose the smaller circle on the left, she would get a 3. If she indicates that she is a little like the girl who is not very good at puzzles, she would receive a 2. And if she is a lot like that girl, she would get a score of 1. Item scores are averaged across the six items for a given subscale, and these four means provide the child's profile of perceived competence and social acceptance. Teacher rating scale - A teacher rating scale parallels the PSPCSAYC. On this scale, teachers are given a brief verbal description of each item (e.g., good at puzzles) and then rate how true that statement is on a four-point scale (really true, pretty true, only sort of true, and not very true). Thus, these scores can be compared with the child's scores, depending on one's purposes. Similar scoring and parallel forms are available for the PCSC.
Reliability	Internal consistency for the PSPCSAYC subscales ranged from .50 to 85. When combined according to their designated factors, these reliabilities increase substantially, falling within a range of .75-.89. Attenuated variability somewhat spuriously attenuates internal consistency reliability somewhat. Internal consistency also adequate for the PCSC.
Validity	Convergent, discriminant, and predictive validity are adequate for both measures. On the PSPCSAYC, for pupils whose ratings are either congruent or lower than the teachers', their perceived cognitive competence is predictive of their actual behavior. However, the presence of over raters in the sample attenuates the predictive validity of this subscale.
Strengths	Direct Assessment
Weaknesses	Some psychometric properties are slightly weak.
Publisher/Price	Must contact Susan Harter at University of Denver, sharter@du.edu

Harter, S. (1982). The Perceived Competence Scale for Children. *Child Development, 53*, 87-97.

Harter, S. & Pike, R. (1984). The pictorial scale of perceived competence and social acceptance for young children. *Child Development, 55*, 1969-1982.

Measure	Positive and Negative Affect Scale (PANAS) Positive and Negative Affect Scale, Child Version (PANAS-C)
Constructs	Self-management
Age range	PANAS: preschool and elementary school PANAS-C Elementary School (grades 4-8)
Rating type	Parent, Self
Description of measure as related to construct of interest	The PANAS assesses how the child/adult "feels on average," for 12 negative emotions (e.g., sad, angry) and 3 positive emotions (e.g., excited, enthusiastic). The PANAS has been modified for parent report for preschoolers and early elementary grades; examines overall emotional expressiveness. The PANAS-C (Laurent et al., 1999) is a brief, useful measure of emotional experience that can be used to differentiate anxiety from depression in youngsters. It has been developed mainly as an instrument for use in a general school population.
Administration	Very simple and quick; less than 10 minutes The PANAS-C consists of 27 items. Children are instructed to indicate how often they have felt a specific way (e.g., interested, sad, etc.) during the "past few weeks" or "past 2 weeks" on a 5-point Likert scale (1 = very slightly or not at all, 5 = extremely).
Scoring	Sum Likert ratings for subscales – i.e., all negative emotions' ratings are summed (NA), and all positive emotions' ratings are summed (PA).
Reliability	PANAS Watson, Clark, and Tellegen (1988) reported alpha coefficients of .87 for both the negative affect (NA) and positive affect (PA) scales; importantly, the time instructions used in this study were similar to those used for validation of the PANAS-C. PANAS-C: The 27-item PANAS-C demonstrated psychometric properties much like those of the PANAS: alpha coefficients for the NA scale were .94 and .92 for the scale development and replication samples, respectively; the alpha coefficients for the PA scale were .90 for the scale development sample and .89 for the replication sample. Also, a two-factor solution best described the structure of the PANAS-C, consistent with its parent measure. The scales also demonstrated good convergent and discriminant validity.
Validity	Used in an increasing number of studies related emotional expressiveness to numerous social and psychological outcomes.

Any modifications?	No
Strengths	Short, good psychometrics
Weaknesses	--
Publisher/Price	Public domain as far as we know

Laurent, J., Catanzaro, S.J., Joiner, T.E., Rudolph, K.D., Potter, K.I., Lambert, S., et al. (1999). A Measure of Positive and Negative Affect for Children: Scale Development and Preliminary Validation. *Psychological Assessment, 11*, 326-338.

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.

Measure	Rothbart Temperament Scales – Infant, Early Childhood, Child
Constructs	Self-Management
Age range	Infancy through childhood Note. The Rothbart scales include several versions based on the age of the ratee. For this compendium's purpose, the following are applicable: Early Childhood Behavior Questionnaire (ECBQ), Children's Behavior Questionnaire (CBQ)
Rating type	Teacher, Parent
Description of measure as related to construct of interest	Three higher-order temperament factors pertinent to the assessment of emotional expressiveness and regulation have been isolated: (a) negative affectivity, (b) surgency; and (c) effortful control (Rothbart et al, 1994). Taken together, they comprise a child's constitutional, individual pattern of self-regulation and reactivity, relatively enduring biological predispositions that are influenced over time by maturation and experience. Negative affectivity items involve discomfort experienced in over-stimulating situations, frustration, anger, and inability to soothe oneself, fearfulness, and sadness. The Surgency dimension includes active, approach, pleasure, and smiling scales. Use of Rothbart Temperament Questionnaires can add to knowledge of children's expressiveness across many everyday contexts. Many children high on the temperament dimension of negative affectivity are easily angered in many situations. Others high on this dimension are anxious, fearful in new situations, and easily saddened. Thus, this factor can be divided into "externalizing negative emotions" and "internalizing negative emotions." It is easy to see how this potent combination could make interacting with both peers and adults problematic. Effortful control, also assessed by the CBQ, is associated with sensitivity to the emotional experiences of peers, which can lead to empathic and other prosocial responses, as well as to inhibition of aggressive impulses (Kochanska, 1993; Rothbart et al., 1994). More specifically, regulatory abilities in attention, in particular the ability to focus and shift attention voluntarily, and the ability to disengage attention from one's own perspective to attend to another's, are hallmarks of prosocial development (Kochanska, 1993). We would expect children higher on the effortful control dimension to be seen by teachers, observers, and peers alike as more socially competent. Effortful control encompasses scales measuring inhibitory control; maintenance of attentional focus during tasks; pleasure experienced during low intensity situations (e.g., looking at picture books); and perceptual sensitivity and awareness of external cues. Thus, the CBQ's scales related to emotion regulation, or internally consistent abbreviations thereof, could be useful.

	For regulation, four scales are used, as follows: (a) attention focusing ("will move from one task to another without completing them" (reversed); (b) attention shifting ("can easily shift from one activity to another"; (c) inhibition control (e.g., "can lower her voice when asked to do so; and (e) impulsivity ("rushes into new situations"). Surgency is an aspect of temperament associated with extraversion, approach to novel stimuli, positive emotional expressiveness, activity, and high level pleasure. Hence, a child high on this dimension of temperament might be a lot of fun to be around-eagerly initiating contact with others, finding interesting things to do, sharing positive affect. On the other hand, there could be "too much of a good thing," with children high on such a dimension possibly seen as irritatingly active and boisterous, risk-taking, and impulsive.
Administration	As an example, the Child Behavior Questionnaire (all scales scored in the same manner): The CBQ is an upper extension of Rothbart's Infant and Toddler Behavior Questionnaires (which is also recommended), with similarly excellent reliability and validity. It is an instrument that assesses temperamental characteristics of children aged 3-8 years. Raters score, on seven-point scales, how "true" 195 specific descriptive behaviors have been of the person being rated, over the past six months. The option of indicating that any item is "not applicable" to the child is also available for infant/child measures.
Scoring	The instructions are generally clear and useful to raters, although the scales include many items and take some parents over an hour to complete (short versions may be recommended, provided that scales of interest are still included). Very clear scoring instructions; facilitated by SPSS routines.
Reliability	All reported internal consistency reliabilities and test-retest reliabilities for all versions for all scales are moderate to excellent. CBQ exhibits substantial interparental agreement.
Validity	Concurrent and predictive validity established in many studies
Any modifications	Use short forms, or if deemed more appropriate, very short forms
Strengths	Extremely well constructed and validated, with theoretical foundations in brain and emotional development.
Weaknesses	Despite some disagreement in the literature, we would choose parental report of temperament, because: (1) parents see a wide range of behavior; (2) recent measurement advances allow their reports even greater objective validity; and (3) most importantly, the social relationship aspects of child temperament are best captured in parental reports Rothbart and Bates (1998).
Publisher/Pricing	The following is a quote from Dr. Rothbart's website, http://darkwing.uoregon.edu/~maryroth/: "We do not charge researchers to use our temperament measures. Dr. Rothbart believes the free exchange of scientific information is essential to research improvement. Access permission to the questionnaires can be repaid by sharing with us the results of your studies.

	To request access to the questionnaires, you can click on the website links, which send an email request to Dr. Rothbart's secretary. Please describe your intended study, the age ranges, and which questionnaire(s) you plan to review or use. Our email reply will provide the access codes and information on how to obtain copies of Dr. Rothbart's publications (some of which are linked and available for download). You may also phone (541-346-5534), or write to Rothbart Temperament Lab, Attention: Cheré DiValerio, 1227 University of Oregon, Eugene, OR 97403-1227."

Capaldi, D. M., & Rothbart, M. K. (1992). Development and validation of an early adolescent temperament Measure. *Journal of Early Adolescence, 12*, 163-173.

Kochanska, G. (1993). Toward a synthesis of parental socialization and child temperament in early development of conscience. *Child Development, 64*, 325-347.

Rothbart, M. K., Ahadi, S. A., & Hershey, K. L. (1994). Temperament and social behavior in childhood. *Merrill-Palmer Quarterly, 40*, 21-39.

Rothbart, M.K. & Bates, J.E. (1998). Temperament. In W. Damon (Ed.), Handbook of child psychology: Vol. 3 Social, emotional, and personality development (5th ed., pp. 105-176). New York: Wiley.

For more specific reading on temperament constructs and measures, see the following:

Posner, M. I. & Rothbart, M. K. (2000). Developing mechanisms of self-regulation. *Development and Psychopathology, 12,* 427-441.

Putnam, S. P., Ellis, L. K., & Rothbart, M. K. (2001). The structure of temperament from infancy through adolescence. In A. Eliasz & A. Angleitner (Eds.), Advances in research on temperament (pp. 165-182). Germany: Pabst Science.

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*, (5), 1394-1408.

Rothbart, M. K., Chew, K., & Gartstein, M. A. (2001). Assessment of temperament in early development. In L. Singer & P. S. Zeskind (Eds.), Biobehavioral assessment of the infant (pp. 190-208). New York: Guilford.

Rothbart, M. K., Derryberry, D., & Hershey, K. (2000). Stability of temperament in childhood: Laboratory infant assessment to parent report at seven years. In V. J. Molfese & D. L. Molfese (Eds.), Temperament and personality development across the life span, (pp. 85-119). Hillsdale, NJ: Erlbaum.

Rothbart, M. K., & Hwang, J. (2002). Measuring infant temperament. *Infant Behavior & Development, 25*(1), 113-116.

Measure	Social Skills Rating System (SSRS)
Constructs	Self-Management, Relationship Skills, Responsible decision-making, (Lack of Risky, Disruptive/Externalizing Behaviors and Internalizing Symptoms), Academic Competence
Age range	Preschool through secondary school
Rating type	Teacher, Parent, Self
Description of measure as related to construct of interest	"To assist professions in screening and classifying children suspected of having significant social behavior problems and aid in the development of appropriate interventions for identified children" (Gresham & Eliot, 1990, p. 1) Scales include: Social Skills: Cooperation, Assertion, Responsibility, Empathy, and Self-Control Problem Behaviors: Externalizing, Bullying, Hyperactivity/Inattention, Internalizing, Academic Competence: Reading Achievement, Math Achievement, Motivation to Learn Norm-referenced rating scale comprised of three different rating forms for teachers, parents, and students. Intended for use with preschool, elementary, and secondary students (with separate forms for raters at each level). Number of items ranges from 40-57 for the adult raters, and 34-39 for students in Elementary school and higher. All include cooperation, assertion, and self-control subscales. Parents also complete a responsibility scale, and students also complete an empathy subscale.
Administration	All items evaluated on a 3-point Likert scale. Each rater can complete their task in 20 minutes. Materials are user-friendly and readable, with clear, comprehensive examples.
Scoring	Hand scoring as described in manual and on individual item booklets takes about 5 minutes. Manual presents raw scores, standard scores, percentile ranks, confidence bands, and descriptive "behavior levels." These "behavior levels" garner some criticism, to the point where some advise only using raw scores.
Reliability	Manual includes detailed information on reliability. Teacher form internal consistency is excellent, parent and student internal consistencies adequate overall. Test-retest excellent for teacher form and parent (social skills scale; we would not use the attendant behavior problems scale for this construct). Test-retest reliability for the student form is somewhat limited, as might be, at least in part, expected with developing children. Interrater reliabilities are rather low, but slightly better than many other cross-informant, cross-context reliability in the child behavioral testing literature.

Validity	Evidence is presented in the manual for content, criterion, and construct validity of the SSRS (the exception is that the criterion-related validity evidence so far for the student form is limited to adequate).
Strengths	Reliable and valid measure capturing important aspects of social functioning, by multiple informants, across many years of childhood/adolescence.
Weaknesses	Because the SSRS was developed to assist in identifying children with difficulties in this area, outcome studies may be influenced by a ceiling effect (i.e., global and subscale scores are negatively skewed); thus the SSRS may form a better estimate of social skills deficits than well-developed social skills. However, no extant standardized measure covering such a wide age range is strength-based in this way. The student forms might be used more cautiously because of relatively lower reliabilities. Some evaluators urge use of total score only, rather than four subscales. A stronger standardization sample is needed that includes children from families with lower incomes. Despite some reviewers' naysaying, I believe this measure is useful. Demaray et al. (1995) found the SSRS to be a laudable tool.
Publisher/Price	Published by American Guidance Services. SSRS Preschool/Elementary Starter Set : $180.99 SSRS Secondary Starter Set: $162.00 SSRS All Levels, All informants Questionnaires, Scannable (25): $42.99

Demaray, M. K., Ruffalo, S. L., Carlson, J., Busse, R. T., Olson, A. E., McManus, S. M., & Leventhal, A. (1995). Social skills assessment: A comparative evaluation of six published rating scales. *School Psychology Review, 24,* 648-671.

Gresham, F. M., & Elliott, S. N. (1990). The Social Skills Rating System. Circle Pines, MN: American Guidance Service.

Measure	Social Skills Improvement System (SSIS)
Constructs	Self-Management, Relationship Skills, Responsible decision-making, (Lack of Risky, Disruptive/Externalizing Behaviors and Internalizing Symptoms), Academic Competence
Age range	Ages 3-18
Rating type	Teacher, Parent, Self
Description of measure as related to construct of interest	• Purpose of the Assessment: o Screen for problem behaviors and identify students at risk for social behavior difficulties and poor academic performance. o Identify specific social behavior acquisition and performance deficits that can be addressed with skill-building school and home interventions. o Identify social skills strengths. o Compare students to national norms to identify individuals functioning below normative expectations (candidates for intervention services). o Provide a baseline for post-intervention progress evaluation. o Track progress. o Gather longitudinal research data.
Administration	• Used by: School, after-school, social services, mental health professionals. • Versions: Teacher, parent and student versions of rating scale available (Student self-report available for 8-18). • Rating procedure: o Raters should be familiar with the SSIS and the rating forms. Teachers need to establish rapport with the student. o Each form takes 15 – 20 minutes to complete.
Scoring	• Scale Contents: o Number of items differs based on form and age of child. Approximately 140 items per form. o Teachers and parents indicate frequency of behaviors exhibited by the child on a 4 point scale from “Never” to “Almost Always” o Students indicate how true a statement about themselves is on a 4 point scale from “Not True” to “Very True” o Constructs assessed ▪ Social Skills – Includes the following sub-constructs: Communication, Cooperation, Assertion, Responsibility, Empathy, Engagement, Self-Control ▪ Problem Behaviors (only on the Teacher and Parent forms) – Includes the following sub-constructs: Externalizing, Bullying, Hyperactivity, Internalizing, Autism Spectrum

	▪ Academic Competence (only on the Teacher form) – Includes the following sub-constructs: Reading and Math Performance, Motivation, Parental Support, General Cognitive Functioning • Scoring procedure: o Scoring procedures – can be hand scored or computer scored. o Has procedures for interpreting the reports. o Has procedures for reporting when there are multiple raters. o Provides a case example of scoring and reporting the results. • Normative sample: o Normative sample included 4700 students aged 3 through 18. 385 teachers provided ratings. 2800 parents provided ratings. o Conducted a national standardization sample aligned with demographic results of the 2006 US Census. Norm sample consisted of approximately 5,000 teachers, parents and students. o Normative scores by gender, grade, and ethnicity for the 3 forms
Reliability	Alpha coefficients for scales range from .70 to upper .90s. Test retest reliability ranges from .72 - .87 for the 3 forms. Inter-rater reliability ranges from .36 - .69.
Validity	Manual reports modest support for the convergent and discriminate validity of the SSIS rating subscales. Correlations between SSIS and other rating scales (e.g. Home and Community Social Behavior Scale (HCSBS) and the Behavioral Assessment System for Children (BASC-2)) demonstrated low or moderate correlations and these correlations varied by age level. Validity was acceptable for testing on special populations (Student's with Autism, ADHD, Learning Disabilities and Speech/Language Impairments).
Strengths	Has teacher, parent and child versions. All strengths of SSRS plus changes from SSRS: Spanish versions available. Four new subscales (Communication, Engagement, Bullying, and Autism Spectrum) National norms for pre-school. Standardization based on a nationwide sample matched to the US population estimates for race, region, and SES. Combined norms and separate-sex norms Computerized scoring and reporting software available. Results link directly to interventions. Forms can be completed within 15-20 minutes. Part of a larger system that includes a universal screening tool and to social skills building programs.
Weaknesses	Manual states that interpretation of scores and reports should be done by a professional who is familiar with test construction and interpretation. Manual does not cite a theoretical basis for the positive behavior items. No analysis of the dimensionality of the assessment was conducted. Cost is high.

Publisher/Price	SSIS Hand scored starter kit (Includes manual, package of each set of forms): $241.00 SSIS Computer scored starter kit (Includes manual, package of each set of forms, ASSIST Scoring Software): $502.20 SSIS Manual : $101.00 SSIS Hand Scored Forms (Teacher, Parent and Student forms sold separately): $42.75 SSIS Computer Scored Forms (Teacher, Parent and Student forms sold separately): $52.00

Measure	Sociometric Ratings and Nominations
Constructs	Relationship skills, (Lack of Risky, Disruptive/Externalizing Behaviors and Internalizing Symptoms)
Age range	Preschool to adolescence
Rating type	Peer, Self
Description of measure as related to construct of interest	Sociometric measures capture overall social status in each group; participants' classmates are asked to name an unlimited number of children who they "like a lot" and who they "don't like very much." As well, to identify aggression, they will be asked to name children who "start fights," "yell and call other kids mean names," "hit and push other kids." To pinpoint relational aggression that can be such a roadblock to forgiveness (Crick & Grotpeter, 1995, 1996), they will name children who "get even by keeping kids they are mad at out of their group of friends," "tell their friends that they will stop liking them unless they do what they say," "try to keep certain kids from being in their group during activity or play time," and "ignore or stop talking to kids they are mad at."
Administration	For preschoolers, use photographs of classmates and ratings (3-point for preschoolers; see Denham & Holt, 1993). For elementary school and older, use nominations. The administration takes 20 minutes or less, although usually < 10 minutes; at the elementary school level, instructions take a few minutes, and the measure can be lengthened to 20 minutes by the inclusion of numerous nominations. Preschoolers require individual interview administration, with "props" that ease explanation of the task; older children can be administered sociometric nomination measures in group settings. Schonert-Reichl also included assessments of prosocial behaviors ("cooperates in a group," "someone you can trust," "kind to others," "helpful," "understand other kids' point of view," "fair"), antisocial/aggressive behaviors ("starts fights," "disrupts things in a group," "can't take teasing," "talks behind other peoples' backs"), and socially withdrawn behaviors ("shy," "easy to push around") in sociometric measures (Schonert-Reichl, 1999).
Scoring	For ratings, find a weighted average (e.g., dislike weight = 1, neutral weight = 2, like weight = 3; for kindergartners, use 5-point scale) For nominations, frequencies of each nomination will be calculated for each classroom group and standardized within group. Social preference, social impact, overt aggression, and relational aggression scores for each participant will be calculated (see Dodge & Coie, 1987).

Reliability & Validity	Many developmental studies in the last two decades support the reliability and validity of these procedures.
Strengths	Gets view of these constructs from actual social partners. It is important to note that although teachers can tell us who is well accepted in a group of peers, they are not good reporters on more problematic aspects of peer reputation (Landau et al., 1984).
Weaknesses	Difficulties sometimes exist in convincing parents and school systems that these measures to not pose harm to children. They speculate that there may be risk in involved, in that: Sociometrics implicitly sanction making negative statements about other people, and may lead children to view rejected children more negatively than they already do, increase negative interactions with unpopular peers, or increase salience of social ostracism in the peer group and thus increase children's unhappiness. We can happily report that there is no support for these worries – in appears that participation in studies including sociometric measures involve no more risks than everyday social life. Researchers (e.g., Bell-Dolan et al., 1989a, 1989b, 1992) have found, following administration of sociometrics measures, NO increase in negative interactions with unpopular peers, NO increase of socially withdrawal in less accepted children, and NO expression of unhappiness or loneliness after participation in studies with sociometric measures. Most children appear to enjoy considering such issues, which are paramount in the minds of elementary students in any case, do not change behavior, and may in fact benefit from discussion of such issues with researchers It is possible to obtain parallel social impact/social information more affirmatively by asking children who they like most and also using sociometric ratings (see Asher & Dodge, 1986). This measure is the 'gold standard' of peer competence measurement during the middle childhood period. The concerns of school systems and parents should be treated respectfully; but much research (as well as the common sense developmental notion that children during middle childhood and early adolescence are constantly making the judgments subsumed within sociometric measurement, just more informally than presented in psychological measurement). Even if only subset of data sites can collect these data, however, we would highly recommend gathering as much sociometric data as possible.

Asher, S. R., & Dodge, K. A. (1986). Identifying children who are rejected by their peers. *Developmental Psychology, 22*, 444-449.

Bell-Dolan, D. J., Foster, S. L., & Christopher, J. M. (1992). Children's reactions to participating in a peer relations study: Child, parent, and teacher reports. *Child Study Journal, 22*, 136-156.

Bell-Dolan, D. J., Foster, S. L., & Sikora, D. M. (1989a). Effects of sociometric testing on children's behavior and loneliness in school. *Developmental Psychology, 25*, 306-311.

Bell-Dolan, D. J., Foster, S. L., & Tishelman, A. C. (1989b). An alternative to negative nomination sociometric procedures. *Journal of Clinical Child Psychology, 18*, 153-157.

Crick, N. R., & Grotpeter, J. K. (1995). Relational aggression, gender, and social-psychological adjustment. *Child Development, 66,* 710-722.

Crick, N., & Grotpeter (1996). Children's treatment by peers: Victims of relational and overt aggression. *Development and Psychopathology, 8,* 367-380.

Denham, S. A., & Holt, R. (1993). Preschoolers' peer status: A cause or consequence of behavior? *Developmental Psychology, 29,* 271-275.

Dodge, K. A., & Coie, J. D. (1987). Social-information processing factors in reactive and proactive aggression in children's peer groups. *Journal of Personality and Social Psychology, 53*, 1146-1158.

Hayvren, M., & Hymel, S. (1984). Ethical issues in sociometric testing: The impact of sociometric Measures on interaction behavior. *Developmental Psychology, 20*, 844-849.

Landau, S., Milich, R., & Whitten, P. (1984). A comparison of teacher and peer assessment of social status. *Journal of Clinical Child Psychology, 13*, 44-49.

Parkhurst, J. T., & Asher, S. R. (1992). Peer rejection in middle school: Subgroup differences in behavior, loneliness, and interpersonal concerns. *Developmental Psychology, 28(2),* 231-241.

Ratiner, C., Weissberg, R., & Caplan, M. (1986, August). Ethical considerations in sociometric testing: The reactions of preadolescent subjects. Paper presented at the 94th annual meeting of the American Psychological Association, Washington, DC.

Schonert-Reichl, K. A. (1999). Relations of peer acceptance, friendship, adjustment, and social behavior to moral reasoning during early adolescence. *Journal of Early Adolescence, 19(2),* 249-279.

Wentzel, K. R., & Erdley, C. A. (1993). Strategies for making friends: Relations to social behavior and peer acceptance in early adolescence. *Developmental Psychology, 29(5),* 819-826.

Measure	Assessment of Children's Emotion Skills (ACES)
Constructs	Social Awareness
Age range	Elementary
Rating type	Performance-based
Description of measure as related to construct of interest	Assesses children's emotion attribution accuracy and emotion biases in threesections that cover social behaviours, social situations, and facial expressions. ACES includes sections concerning social behaviors, social situations, and facial expressions.
Administration	ACES consists of three subscales: Facial Expressions, Social Situations, and Social Behaviors. The Facial Expressions subscale assesses the ability to understand emotions conveyed through other children's facial expressions. The social behaviors and social situations sections each contain 15 one- to three sentence items. In response to each item, children label the protagonist's feeling by choosing happy, sad, mad, scared, or no feeling. Items describe behaviors or situations associated with emotion arousal. For example, the emotion behavior items related to sadness reflect lethargic and withdrawn somatic responses, and., the emotion situation items related to anger describe events in which a person is blocked from a desired goal. An additional three vignettes in each subscale are ambiguous, describing behaviors or situations that do not depict prototypic emotion behaviors or situations. To elicit children's attribution biases, each section also includes three additional items that describe social behaviors or social situations not associated exclusively with one discrete emotion. The facial expressions section includes 26 photographs of elementary-aged children posing facial expressions. Four photographs each contain happy, sad, mad, and afraid faces. For the purposes of eliciting children's biases, an additional 10 photographs are included that contain a mixture of emotion signals and did not receive consensus as depicting a particular emotion during pilot testing. For all sections of ACES, items are randomized within blocks containing happy, sad, angry, afraid, and ambiguous items.
Scoring	The emotion attribution accuracy score reflected how many items a child answered correctly on the 40 items for joy, sadness, anger, and fear across the three sections ~M= 29.7, SD 5 4.2). ACES' three sections also jointly produce emotion attribution bias scores. The emotion knowledge score was calculated by determining the number of

	correct responses to happiness, sadness, anger, and fear items within each subscale. Then, the raw subscale scores were standardized based on the present sample so that each subscale standard score had a mean of 0 and a standard deviation of 1. Finally, each child's subscale standard scores were added to attain the total emotion knowledge score.
Reliability	Internal consistency is adequate across studies (e.g., Mavroveli et al., 2009; Schultz et al., 2004). In the current study, this scale had moderate internal reliability (a = .71). These items cohered moderately well (Cronbach's alpha = .68).
Validity	Mavroeli et al. (2009) found that the ACES correlated well with trait emotional intelligence.
Strengths	Taps emotion knowledge in theoretical context of discrete emotions theory, for an age range that needs a good measurement tool. Psychometrics appear adequate, and addition of assessment of emotions attribution biases is a real plus.
Weaknesses	Not used as yet in much published research.
Publisher/Price	Public domain, as far as we know

Mavroveli, S., Petrides, K. V., Sangareau, Y., & Furnham, A. (2009). Exploring the relationships between trait emotional intelligence and objective socio-emotional outcomes in childhood. *British Journal of Educational Psychology, 79*, 25-272.

Schultz, D., Izard, C. E., & Bear, G. (2004). Children's emotion processing: Relations to emotionality and aggression. *Development and Psychopathology, 16*, 371-387.

Measure	Behavioral and Emotional Rating Scale-Second Edition: (BERS), Parent Rating Scale (PRS), Youth Rating Scale (YRS)
Constructs	Self-Awareness, Social Awareness, Self-Management, Relationship Management
Age range	Elementary and middle school. (PreK through 12th grade for teacher and parent rating forms. Students 11 years old and up (5th grade and older) can take the self-report scale
Rating type	Teacher, Parent, Self
Description of measure as related to construct of interest	Standard scores are calculated for an overall strength index and five subscales derived from factor analysis: (1) interpersonal strength, which measures ability to control emotions and behavior in social situations; (2) family involvement, which measures participation and relations with the family; (3) intrapersonal strength, which assesses the child's perception of competence and accomplishment; (4) school functioning, which addresses competence in school and classroom tasks; and (5) affective strength, which focuses on the ability to give and receive affect. Used as an evaluation measure, for planning interventions and services, and as an outcome measure.
Administration	Rapid-assessment instrument (RAI) which is a tool that can be completed as a self-report or by others (e.g., classroom teachers, psychologists, social workers, parents). Typically, they take less than 15 minutes to finish, are written in clear concise language, and do not require special training to administer or score RAIs are time efficient, nonintrusive, and useful as planning and outcome measures. Rater should have had regular, daily contact with the child for at least a few months before responding to the rating scale. Rater should understand the theoretical basis for the assessment, be conversant with the construction of the scale, be proficient in administering the BERS, have working knowledge of interpreting the results.
Scoring	52 items on teacher scale; 57 items on student and parent scale. 8 open-ended questions on each version that ask about child academic, social, athletic, family, and community strengths. 5 subscales that correspond to each of the 5 constructs listed below. All items are rated on a scale of 0 to 3 (0 = not at all like the child; 1 = not much like the child; 2 = like the child; 3 = very much like the child). Raters rate how often engaged in each behavior over the past four weeks. Assesses behavioral and emotional strengths of children. Defined as the measurement of emotional and behavioral skills that create a sense of personal accomplishment, contribute to satisfying relationships with families, peers, and adults, enhance one's ability to deal with adversity and stress, promote one's personal, social, and academic development.

	Constructs assessed (note that no definition were given for these constructs): • Interpersonal Strength • Family Involvement • Intrapersonal Strength • School Functioning • Affective Strength • Career Strength Scores on all items are added by hand to create an overall raw score. The rating scale sheet allows for hand scoring of the items. The BERS-2 Strength index converts raw subscale scores into a standard score with a mean of 100 and a standard deviation of 15. Normative tables are provided to convert the subscale raw scores to percentile ranks and scaled scores.
Reliability	High test-retest reliability (at or above .80) over a one week period. The scale's reliability has been determined in the areas of interrater reliability (that is, the consistency with which different individuals rate the same behavior) Teacher-teacher inter-rater reliability was .83. Parent and teacher inter-rater reliability was .54. Reliability findings enhance the confidence with which practitioners can use the measure for screening decisions in which the data is reported for individual students. Alpha coefficients for scales ranged from .80 to .95.
Validity	Convergent validity was assessed as parents completed the PRS and a second measure of perceived child functioning. In the second study, the six BERS-2 PRS subscales and overall strength index were correlated with composite scores from the parent forms of the Social Skills Rating System. Correlations were generally highly positive. PRS was also correlated with problem scales of Achenbach's Child Behavior Checklist. The BERS consistently demonstrated moderate to high correlations with competence- oriented scales and moderate to high negative correlations with deficit-oriented scales across different age ranges (Epstein, Nordness, Nelson, & Hertzog, 2002). To determine content validity, a significant item development, selection, and validation process was followed (see Epstein & Sharma, 1998, for a more detailed description). Parents and professionals developed a list of objective statements that were tested and found able to discriminate between children with developmentally appropriate emotional and behavioral strengths (that is, children without emotional

	and behavioral disorders [EBD]) and those with less well-developed emotional and behavioral strengths (that is, children with EBD). In terms of discriminant validity, the BERS adequately discriminates between students with emotional and behavioral disorders and learning disabilities and those without such disabilities. • Teacher version of BERS has expected correlations with the Walker-McConnell Scale of Social Competence and school adjustment and other similar scales. • Parent version of the BERS has expected correlations with the Social Skills Rating System and the Child Behavior Checklist. • Student version of the BERS has expected correlations with the Social Skills Rating System.
Strengths	The BERS-2 PRS appears to be a psychometrically sound, strength-based rapid-assessment instrument. The technical adequacy has been determined with respect to the instrument's reliability and validity. Focuses on students' strengths
Weaknesses	Does not align with the 5 SEL Core Competencies. Items/construct alignment not 100% convincing. Factor analyses have low eigenvalues for factors 2 through 5, and confirmatory factor analysis were not conducted for the teacher rating scale. No provisions for how to coordinate results from parent, teacher, and student scales. Does not have examples of how to use assessment results for an individual student assessment.
Publisher/Price	Cost: BERS2 prices include $198.00 for an introductory kit, $70.00 for an examiner's manual, $36.00 for teacher rating scales, $36.00 for parent rating scales, $36.00 for youth rating scales, and $36.000 for summary forms. http://www3.parinc.com/products/product.aspx?Productid=BERS-2#

Epstein, M. H., & Cullinan, D. (1998). *Scale for assessing emotional disturbance*. Austin: PRO-ED.

Epstein, M. H., Nordness, P. D., Nelson, J. R., & Hertzog, M. (2002). Convergent validity of the behavioral and emotion rating scale with primary grade-level students. *Topics in Early Childhood Special Education, 22*, 114-121.

Epstein, M. H., & Sharma, H. M. (1998). Behavioral and Emotional Rating Scale: A strength based approach to assessment. Austin, TX: PRO-ED.

Gresham, F.M., & Elliott, S.N. (1990). The Social Skills Rating System. Circle Pines, MN: American Guidance Service.

Mooney, P., Epstein, M. H., Ryser, G., & Pierce, C. D. (2005). Reliability and validity of the behavioral and emotional rating scale-second edition: Parent rating scale. *Children & Schools, 27,* 147-155.

Measure	Bryant Empathy Scale for Children
Constructs	Social Awareness, Relationship Skills
Age range	1st grade through end of junior high
Rating type	Self
Description of measure as related to construct of interest	This measure of children's dispositional sympathetic tendencies contains 16 items (e.g., "I feel sorry for people who don't have the things I have" "It makes me sad to see a kid with no one to play with"). Other sample items include: "People who kiss and hug in public are silly," "Seeing a boy who is crying makes me feel like crying," and "I get upset when I see an animal being hurt."
Administration	Self-report questionnaire, takes about 5 – 10 minutes. Designed for use with children in grades 1-7, this instrument can be administered using one of three formats: (a) first graders (5-6 year olds) place cards (one empathy item per card) in a "me" or "not me" box; (b) older children (8-9 year olds) circle "yes" or "no" in response to each item; and (c) seventh graders use either the yes/no format or Mehrabian and Epstein's 9-point scale ranging from "very strong disagreement" to "very strong agreement. "
Scoring	Responses are scored so that higher scores reflect greater empathy; sum across items, with italicized item numbers weighted negatively 1, *2, 3*, 4, 5, 6, 7, 8, *9, 10*, 11, 12, 13, 14, *15, 16, 17, 18*, 19, 20, *21, 22*
Reliability	Internal consistency is adequate.
Validity	As for concurrent validity, sympathy has been associated with high levels of regulation, teacher-reported positive emotionality and general emotional intensity, and especially for boys, high social functioning and low levels of negative emotionality (Eisenberg et al., 1996).
Strengths	Centrality of this trait to social functioning
Publisher/Price	Public domain

Bryant, B. K. (1982). An index of empathy for children and adolescents. *Child Development, 53,* 413-425.

Measure	Child/Teacher/Parent Rating Scale
Constructs	Self-Management, Relationship Skills, Lack of Disruptive Behaviors and Internalizing Symptoms
Age range	Elementary and Intermediate School
Rating type	Teacher, Parent, Peer, Self
Description of measure as related to construct of interest	Assesses problem behaviors (acting-out, shy/anxious behavior, learning problems), and social competencies (frustration tolerance, peer social skills, assertive social skills, task orientation) in the school context.
Administration	38 items for adult versions, 24 for children; takes less than 20 minutes to complete (Hightower et al., 1986, 1987).
Scoring	Sum items for each scale. Child, peer, teacher, and parent forms are available. <u>For parent form:</u> Acting Out Scale = Σ items 1, 4, 7, 10, 13,16 Shy-Anxious = Σ items 2, 5, 8, 11, 14, 17 Learning Problems = Σ items 3, 6, 9, 12, 15,18 Frustration Tolerance = Σ items 19, 23, 27, 31, 35 Assertive Social Skills = Σ items 20, 24, 28, 32, 36 Task Orientation = Σ items 21, 25, 29, 33, 37 Peer Social Skills = Σ items 22, 26, 30, 34,38 <u>For child/peer form</u>: Acting Out Scale = Σ items 1, 5, 9R, 13, 17R, 21 Social Anxiety Scale = Σ items 2, 6, 10, 14, 18, 22 Peer Social Scale = Σ items 3, 7R, 11R, 15, 19R, 23 School Interest Scale = Σ items 4, 8R, 12, 16, 20R, 24R <u>For both forms</u>, scores 3 = YES 1 = NO R= recode the reverse 1=3 2=2 3=1

Reliability Validity	This measure has excellent internal and test-retest reliability, as well as demonstrated validity, including discrimination of referred from non referred children, and convergent/divergent validity.
Any modifications	No, unless some subscales are not needed, such as problem behaviors and task orientation.
Strengths	Excellent theoretical basis, good psychometrics, quickly completed.
Weaknesses	May be redundant with SSRS/SSIS
Publisher/Price	Public domain as far as we know; however, Primary Mental Health Project, Inc. requests that one obtain permission for use; see http://www.childrensinstitute.net/ obtained from http://vinst.umdnj.edu/VAID/TestReport.asp?Code=CRS

Hightower, A. D. (1986). The teacher-child rating scale: A brief objective measure of elementary children's school problem behaviors and competencies. *School Psychology Review, 15*, 393-409.

Hightower, A. D., Cowen, E. L., Spinell, A. P., & Lotyczewski, B. S. (1987). The Child Rating Scale: The development of a socioemotional self-rating scale for elementary school children. *School Psychology Review, 16,* 239-255.

Measure	Children's Emotion Management Scales: Anger and Sadness
Constructs	Self-Awareness, Self-Management
Age range	Elementary School
Rating type	Self
Description of measure as related to construct of interest	Measures: Emotion regulation The 12-item Children's Sadness Management Scale (CSMS) was developed to assess children's inhibition, dysregulated expression, and coping with sadness experience and expression. (Zeman et. al, 2001). Subscales include: Coping (e.g., "I stay calm and don't let sad things get to me") Inhibition: (e.g., "I hold my sadness in") Dysregulated-Expression: (e.g., "I whine/fuss about what's making me sad")
Administration	Self-report, paper and pencil
Scoring	Children respond to items on a 3–point scale (1 - hardly ever, 2 - sometimes, 3 - often).
Reliability	A three-factor solution was supported with strong internal consistency for the Inhibition scale and moderately strong internal consistency for the Emotion Regulation Coping and Dysregulated-Expression scales. (Zeman et. al, 2001).
Validity	Validity was demonstrated for each of the three scales by providing indices of convergent and discriminant validity with measures of emotion, psychopathology, and social functioning as well as examination of gender differences. (Zeman et. al, 2001).
Strengths	Findings indicate that the CSMS provides a reliable and valid measure of normative sadness management. (Zeman et. al, 2001). The CSMS will add to the literature by enabling both researchers and clinicians to assess efficiently children's self-reported management of normative sadness experience and expression. (Zeman et. al, 2001).
Weaknesses	Although the CSMS represents an important first step in developing a more comprehensive protocol of emotional competence measures, it has several limitations. First, the data were collected from a community sample that likely resulted in a restricted range of emotional functioning and symptoms of emotional disturbance. Second, the range of ages used was somewhat limited, although research indicates that this measure can be used successfully with children from ages 6 to 14 years (e.g., Shipman et al., 2000). Third, one of the practical strengths of this

	instrument is the brevity that was obtained, however, at the expense of more robust reliability indices. Fourth, the scope of this instrument is somewhat narrow and is not intended as a global measure of emotional competence. Future research should build and expand upon the concepts used in this measure when developing a more comprehensive battery of instruments to assess emotional competence. (Zeman et. al, 2001).

Penza, S., Zeman, J., & Shipman, K. (1998, March). Validation of the emotion dysregulation scale for children (EDS). Poster presented at the Conference on Human Development, Mobile, AL.

Shipman, K., Zeman, J., Penza, S., & Champion, K. (2000). Emotion management skills in sexually maltreated and nonmaltreated girls: A developmental psychopathology perspective. *Development and Psychopathology 12*, 47-62.

Zeman, J., & Shipman, K. (1996). Children's expression of negative affect: Reasons and methods. *Developmental Psychology, 32*, 842-849.

Zeman, J., Shipman, K., Penza-Clyve, S. (2001). Development and initial validation of the children's sadness management scale. *Journal of Nonverbal Behavior, 25*, 187-205.

Measure	Devereux Student Strengths Assessment (DESSA)
Constructs	Self-Awareness, Self-Management, Social Awareness, Relationship Skills, Responsible decision-making
Age range	Kindergarten – 8th Grade
Rating type	Teacher, Parent
Description of measure as related to construct of interest	• Used by: school personnel, social service staff, mental health professionals, after-school program staff. • Versions: Teacher and parent versions of observational rating scale of the student. no student self-report • Purpose of the Assessment: o provide measure of social-emotional competence o Identify children at risk of developing social and emotional problems o Identifying the unique strengths of children who have social and emotional concerns o Provide meaningful information on child strengths for individual service plans; facilitate parent-professional collaboration o Inform selection of interventions needing universal, targeted, or indicated support; Evaluate impact of programs on outcomes at child and classroom/program levels o Facilitate progress monitoring of children, evaluating change over time o Serve as a research tool to advance science and support policy development
Administration	• Scale Contents: o 72 items, 5 pt scale Items are rated on a 5-point scale varying from "never" to "very frequently." o Raters how often engaged in each behavior over the past four weeks. o Social and Emotional Competencies – ability of children to interact with other children and adults in a way that demonstrates an awareness of, ability to manage emotions, in an age and context appropriate manner o Items were grouped based on CASEL 5 SEL core competencies and additional deliberation among developers. o Constructs assessed ▪ Self-awareness (7 items), a child's realistic understanding of her/his strengths and limitations and consistent desire for self-improvement ▪ Social awareness (9 items) , a child's capacity to interact with others in a way that shows respect for their ideas and behaviors, recognizes her/his impact on them, and uses cooperation and tolerance in social situations

	▪ Self-management (11 items), a child's success in controlling his or her emotions and behaviors, to complete a task or succeed in a new or challenging situation ▪ Goal directed behavior (10 items) a child's initiation of and persistence in completing tasks of varying difficulty ▪ Relationships skills (9 items) a child's consistent performance of socially acceptable actions that promote and maintain positive connections with others ▪ Personal responsibility (10 items) a child's tendency to be careful and reliable in her/his actions and in contribution to group efforts ▪ Decision making (8 items) a child's approach to problem solving that involves learning from others and from her/his own pervious experiences, using values to guide action, and accepting responsibility for decisions. ▪ Optimistic thinking (7 items: a child's attitude of confidence, hopefulness, and positive thinking regarding herself/himself and her/his life situations in the past, present and future. • Rating procedure: o Raters need to be qualified – have sufficient exposure to child for four weeks, 2 or more hours of three days per week for four week period. o Parents, teachers or school staff who are in regular contact with student. o Each child is rated one at a time by teachers, parents.
Scoring	• Scoring procedure: o Eight scale scores and a composite score. o Scoring procedures – Hand scored, and paper survey has procedure for calculating the standardized score. Online administration, scoring and reporting available. Interpretation of the DESSA at the individual level: review the total scale score, examine the individual scale scores, and examine the profile of scale scores. Use the t-scores to determine above or below expected score for that scale.
Reliability	Alpha coefficients for scales ranged from .82 to .98. Test retest reliability .79-.94 for parent and teacher rating.
Validity	Mean scores between regular education and students classified as seriously emotionally disturbed were different. Scores on the DESSA correlated with scores on the Behavioral and Emotional Rating Scale (BERS) and the Behavioral Assessment System for Children (BASC-2). (Nickerson & Fishman, 2009).

Strengths	• It utilizes only strength-based items, whereas most rating scales have a larger focus on clinically oriented items. • The DESSA also lends itself to be used with the general population in Tier 1 (i.e., universal screening) to promote the socialemotional well-being of all youth as well as for more targeted and intensive interventions. • Another advantage the DESSA has over both the BASC–2 and the BERS–2 is that it is shorter and simpler because the same form can be used with both parents and teachers. This also enables parent and teacher ratings to be directly compared on the same set of behaviors. • Available in English and Spanish. • A brief, universal screener, the DESSA-mini is also available.
Weaknesses	--
Publisher/Price	Kaplan prices include $115.95 for a full kit and $39.95 for 25 record forms.

Lebuffe, P., Shapiro, V., B., & Naglieri, J. (2008). Devereux Student Strengths Assessment (***DESSA***). Lewisville, NC: Kaplan Early Learning Co.

Nickerson, A. B., & Fishman, C. (2009) Convergent and divergent validity of the Devereux Student Strengths Assessment. School Psychology Quarterly, 24, 48-59.

Measure	Emotion Expression Scale for Children (EESC)
Constructs	Self-Awareness, Self-Management
Age range	Elementary to middle school/adolescence
Rating type	Self
Description of measure as related to construct of interest	Measures: Emotion awareness Construct: Emotion awareness. Assesses children's awareness of their own emotions and their ability to express negative emotions. Self-report scale designed to examine 2 aspects of deficient emotion expression: lack of emotion awareness and lack of motivation to express negative emotion (Penza-Clyve & Zeman 2002): Poor Awareness Scale (e.g., "I have feelings that I can't figure out") Expressive Reluctance (e.g., "I prefer to keep my feelings to myself) Initially adapted from the 30-item Toronto Alexithymia Scale for Adults, a measure of impoverished ability to express emotion. Items were adapted for children by group consensus of clinical-child graduate students and psychologists.
Administration	Self-report, pencil and paper
Scoring	Children respond to items using a 5-point Likert scale with scores of 1 (not at all true), 2 (a little true), 4 (somewhat true), and 5 (extremely true) to indicate how well each item describes their experience with these expressive difficulties. Higher scores indicate poorer emotion awareness and greater reluctance to express emotion (Penza-Clyve & Zeman 2002).
Reliability	Used a community sample of 208 4th and 5th grade children – high internal consistency and moderate test-retest reliability of the EESC (Penza-Clyve & Zeman 2002). There were no significant sex differences on the mean scale scores or the total score (Penza-Clyve & Zeman 2002).
Validity	The EESC emotion awareness scale was positively related to the inhibition and dysregulated expression of sadness and anger, and negatively related to constructive coping with sadness and anger (i.e., the more children viewed themselves as lacking emotion awareness, the more they were likely to report inhibiting emotional expressivity, expressing emotions in non-constructive ways, and coping in maladaptive ways when experiencing sadness and anger; Penza-Clyve & Zeman 2002).

	The expressive reluctance scale correlated positively with the Inhibition of Emotion scales and Dysregulated Expression scales, indicating that the less willing children were to express emotion, the more they reported expressing these emotions in exaggerated, dysregulated ways. Unlike the emotion awareness scale, the expressive reluctance scale was not correlated with sadness and anger regulation scales, but was, however, positively correlated with emotion regulation questions for anger, sadness, and pain (Penza-Clyve & Zeman 2002). No significant correlations were found between either EESC scale with peer reports of aggression and withdrawn behavior. (Penza-Clyve & Zeman 2002). Both EESC scales correlated positively with internalizing symptom measures, including the CDI, the STAIC, and the CSI (Penza-Clyve & Zeman 2002).
Strengths	Results provide initial support for concurrent validity for the EESC scales evidenced by relations with measures of emotion management (Penza-Clyve & Zeman 2002). The EESC is an important addition to the emotion and child clinical literatures as it may assist in the assessment of children with emotion regulation skill deficits, who consequently may be at risk for developing symptoms of psychopathology. Thus, the EESC represents an important first step in developing a more comprehensive protocol of emotion regulation measures. Upon further validation, the EESC may be clinically useful when incorporated into a psychological assessment to lend more specific information regarding emotional functioning of school-age children (Penza-Clyve & Zeman 2002).
Weaknesses	Lack of association with peer ratings of aggression and withdrawn behavior.
Price	Public domain

Penza-Clyve, S. & Zeman, J. (2002). Initial validation of the emotion expression scale for children (EESC). *Journal of Clinical Child and Adolescent Psychology, 31*(4), 540-547.

Note. For more assessments on emotional expression, see:

Izard, C. E., Libero, D. Z., Putnam, P., & Haynes, O. M. (1995). Stability of emotion experiences and their relations to traits of personality. *Journal of Personality and Social Psychology, 64*, 847-860.

Kring, A. M., Smith, D. A., & Neale, J. M. (1994). Individual differences in dispositional expressiveness: Development and validation of the emotional expressivity scale. *Journal of Personality and Social Psychology, 66*, 934-949.

Tangney, J. P., Wagner, P. E., Burggraf, S. A., Gramzow, R., & Fletcher, C. (1990). The Test of Self-Conscious Affect for Children (TOSCA-C). Fairfax, VA: George Mason University.

Measure	Feelings about School (FAS)
Constructs	Self-awareness, Attachment to School and Teacher
Age range	Children in kindergarten and first grade. Could probably be used with older children
Rating type	Self
Description of measure as related to construct of interest	In this measure, FAS is used to assess children's: Perceptions of Their Competence in Math and Literacy -- These items reflect the way a child feels about their abilities with numbers, letters, and reading (i.e., "how good they are with", "how much they know about", and "how good they are at learning something new"). Feelings About Their Teachers -- These items reflect how a child perceives their teacher to feel about them and how they feel about their teacher (i.e., doesn't like at all to likes a lot). General Attitudes about School -- These items reflect (i.e., "how they feel about going to school," "how fun the things they do in school are," "how they feel when they are in school."
Administration	Individual assessment
Scoring	Scores for each reaction style are averaged across items, as follows: 1-5 Likert-type scales
Reliability	For kindergartners, values were .68 for math, .61 for literacy, .52 for attitudes toward school, .74 for relationship with teacher. For first graders, values were .63 for math, .74 for literacy, .59 for attitudes toward school, .79 for relationship with teacher.
Validity	Four factors were extracted: Perceived competence in math Perceived competence in literacy Children's feelings about their relationship with the teacher Children's general attitudes toward school

Strengths	Demonstrated reliability and validity of the FAS scale with self-systems theory.
Weaknesses	None noted; Perhaps modify to suit early childhood
Publisher/Price	Public domain

Valeski, T. & Stipek, D. (2001). Young children's feelings about school. *Child Development, 72*, 1198-1213.

Measure	Friendship Quality Questionnaire
Constructs	Relationship skills
Age range	Middle childhood
Rating type	Peer, Self
Description of measure as related to construct of interest	Dyadic relationships, as well as overall peer group acceptance, are crucial during middle childhood, continuing into adolescence. On this measure (FQQ), children report on 6 features of their friendships: (a) companionship/recreation; (b) help/guidance; (c) validation/caring; (d) intimate exchange; (e) conflict/betrayal; and (f) conflict resolution.
Administration	Individual or group; takes about 20 – 25 minutes
Scoring	Sum ratings for each subscale, as follows: Validation Scale = Σ items 4,5,6,8,10,12,13,15,30,41 Conflict Resolution Scale = Σ items 11, 26, 35 Conflict and Betrayal Scale = Σ items 3, 9, 20, 21, 27, 31, 37 Help and Guidance Scale = Σ items 17, 18, 24, 28, 32, 33, 34, 36, 39 Companionship and recreation Scale = Σ items 2, 7, 19, 22, 23 Intimate Exchange Scale = Σ items 14, 16, 25, 29, 38, 40 Note. These are factor structure found by Parker & Asher (1993) – Asher and Rose (1999a) raise the possibility of using only two scales, positive and negative.
Reliability	Mean α = .83, with support for the six subscales and good test-retest reliability.
Validity	A number of indicators of validity have been identified–e.g., partners' perceptions of friendship are related, and accepted children describe friendships more positively.
Any modifications for NCS?	No
Strengths	Well constructed, good psychometrics, interesting for children themselves
Weaknesses	Equivocal factor structure could be seen as a weakness.
Publisher/Price	Public domain as far as we know

Parker, J. G., & Asher, S. R. (1993). Friendship and friendship quality in middle childhood: Links with peer group acceptance and feelings of loneliness and social dissatisfaction. *Developmental Psychology, 29*, 611-621

Rose, Amanda J. & Asher, Steven R. (1999). Children's goals and strategies in response to conflicts within a friendship. *Developmental Psychology. 35,* 69-79.

Measure	How I Feel Scale
Constructs	Self-Awareness, Self-Management
Age range	Middle Childhood
Rating type	Self
Description of Measure as related to construct of interest	30-item self-report measure of emotion for 8- to 12-year-old children—the How I Feel (HIF). Item generation and selection occurred via 2 pilot administrations (ns = 250 and 378, respectively). The HIF can be useful in understanding the interplay between arousal and control in social-emotional adjustment in school-age children. Three scales are included: Positive Emotions (e.g., "I feel happy very often") Negative Emotion (e.g., "When I felt sad, my sad feelings are very strong") Emotion Control (e.g., : "I was in control of how often I felt mad")
Administration	Self-report questionnaire, takes about 15 - 20 minutes.
Scoring	Sum likert ratings for scales.
Reliability	Results showed moderate longitudinal stability for 120 children over 2 years.
Validity	Ten experts provided data on content validity. Exploratory factor analysis and subsequent confirmatory factor analysis with samples of 406, 524, 349, and 349 3rd- through 6th-grade children supported a 3-factor model, including the frequency and intensity of (a) positive emotion, (b) negative emotion, and (c) positive and negative emotion control. Concurrent validity was established.
Strengths	Excellent psychometric study
Weaknesses	None noted
Publisher/Price	Public domain as far as we know

Walden, T. A. Harris, V. S., Catron, T. F. (2003). How I Feel: A self-report measure of emotional arousal and regulation for children. *Psychological Assessment. 15,* 399-412.

Measure	Katz- Gottman Regulation Scale
Constructs	Self-Management
Age range	Middle Childhood
Rating type	Parent
Description of measure as related to construct of interest	Parents complete this 45-item questionnaire, which asks them to report the degree to which the child requires external regulation of emotions and behavior by adults. Questions refer to "Up Regulation" (e.g., "Encourage him/her to be adventurous") and "Down Regulation" (e.g., "Be still," "Help him/her calm down after a scary movie").
Administration	Questionnaire; takes approximately 20 minutes
Scoring	Sum Likert ratings for subscales, as follows: Up Regulation = Σ 17, 25, 26, 30 , 32, 35, 36, 37, 38, 39, 41, 44, 45 Down Regulation = Σ 1, 3, 4, 5, 6, 8, 9, 10, 14, 15, 18, 19, 2 9, 21, 22, 23, 24, 27, 28, 29, 31, 34, 40, 42 Total Regulation = Σ all scores
Reliability Validity	Katz reports that psychometrics are adequate to good; in Gottman and Katz (2002), alpha for the scale was .74. In terms of validity, regulation scores at age 8 were predicted by preschool-aged physiological indices, and mediated by children's concurrent ability to maintain calm during a stressful parent-child interaction.
Strengths	Theoretically sound, examines up-regulation as is needed in really understanding emotion regulation and is hardly ever done.
Weaknesses	Relatively little psychometric data as yet
Publisher/Price	Public domain as far as we know

Gottman, J. M., & Katz, L. F. (2002). Children's emotional reactions to stressful parent-child interactions: The link between emotion regulation and vagal tone. *Marriage & Family Review, 34*, 265-283.

Measure	Kusché Affect Interview – Revised
Constructs	Self-and Social Awareness
Age range	Elementary School
Rating type	Other
Description of measure as related to construct of interest	This measure assesses not only the depth of earlier-attained emotion understanding, but also extends this understanding to more complicated emotions (e.g., pride, guilt, jealousy, and anxiety), and taps new aspects of emotion understanding, such as emotional experience, cues for emotion, ambivalent feelings, display rules, temporal aspects of emotions, and the universality/normative nature of emotion.
Administration	Individually administered interview; takes about a maximum of 30 minutes to 1 hour for the entire interview, audiotaped, for the oldest children.
Scoring	See attached scoring sheet.
Reliability	Cook et al. (1994) and Greenberg et al. (1995) have reported excellent reliability and validity; e.g., results from the KAI are predictive of intervention success, and of behavioral problems over and above intellectual ability. Test- retest and internal consistency reliabilities are adequate to good (Greenberg, personal communication).
Validity	Cook et al. (1994) and Greenberg et al. (1995) have reported excellent validity; e.g., results from the KAI are predictive of intervention success, and of behavioral problems over and above intellectual ability.
Any modifications	Selected subtests could be used if determined most important.
Strengths	Excellent coverage
Weaknesses	Needs to be coded, which means some training is needed, and transcripts/tapes must be used.
Publisher/Price	Public domain as far as we know; check with Mark Greenberg at Penn State.

Cook, E. T., Greenberg, M. T., & Kusché, C. A. (1994). The relations between emotional understanding, intellectual functioning, and disruptive behavior problems in elementary-school-aged children. *Journal of Abnormal Child Psychology, 22,* 205-219.

Greenberg, M. T., Kusché, C. A., Cook, E. T., & Quamma, J. P. (1995). Promoting emotional competence in school-aged children: The effects of the PATHS curriculum. *Development & Psychopathology, 7,* 117-136.

Note. The Kusché Emotional Inventory (KEI; .Kusché, 1984) can be used to measure children's emotional knowledge without coding issues. It assesses children's ability to recognize emotion language, concepts, and visual cues. In group or one-on-one interviews, the assessor displays a page containing four drawings of children experiencing different emotions as indicated by facial expression, body posture, and situation. The assessor then asks "which picture shows a child who feels" a particular emotion (i.e., "love," "fear," etc). The assessor repeats this process 30 times for a variety of emotions. Children receive one point for correct identification of an emotion and one point for correct identification of its valence. Children's KEI scores are represented by the percentage of correct responses. The KEI takes approximately 20 minutes to complete, and has been used in a variety of studies involving preschool and kindergarten children (Cortes, 2002; Kusché, 1984). Reliability of the KEI was good (test-retest .82-.85) and content as well as concurrent validity with other measures was sufficient (.36 to .49).

Kusché, C. A. (1984). *The understanding of emotional concepts by deaf children: An assessment of an affective curriculum*. Unpublished dissertation, University of Washington.

Measure	Measure of Prosocial and Aggressive Behavior
Constructs	Relationship Skills, Responsible decision-making, (Risky, Disruptive/Externalizing Behaviors)
Age range	Elementary school
Rating type	Teacher, Parent, Peer, Self
Description of measure as related to construct of interest	Physical and verbal aggression is a 20-item scale (e.g.," 1 threaten others", "I get into fights) with five control items. Target items offer a description of a child's behavior aimed at hurting others physically and verbally (Pastorelli et. al. 1997). Prosocial behavior was assessed by 10 items in terms of helpfulness, sharing, kindness, and cooperativeness (e.g., "I try to help others ") (Caprara & Pastorelli, 1993).
.Administration	Self, teacher, parent, and peer report Teachers rate children for physical and verbal aggression and prosocial behavior using the scales administered to the children, but shortened to six items each and cast in a third-person format. Mothers also rated the frequency with which their children exhibited prosocial and aggressive forms of behavior, using the same set of scales as administered to the children. (Bandura et. al. 1996; Caprara & Pastorelli 1993).
Scoring	Answer format: often, 3; sometimes, 2; never, 1, summed for each scale Control items that do not contribute to the total score (Pastorelli et. al. 1997)
Reliability	For all reporters' scores, except for an alpha of .61 for peer ratings of prosocial behavior, the reliability coefficients for the four sources of data (self, parents, teachers, peers) across the two educational levels for aggressiveness and prosocialness were virtually all in the .80s and .90s. (Bandura et. al. 1996; Caprara & Pastorelli 1993).
Validity	Convergent and discriminant validity have been demonstrated in a multinational study. (Pastorelli et. al. 1997). Concurrent validity of these measures has been corroborated in studies relating children's ratings of their behavior to level of prosocialness and aggressiveness as rated by parents and teachers and by peers' sociometric nominations (Bandura et. al. 1996), and by correlations with "moral disengagement."
Strengths	Excellent psychometric properties Short scales

	Appears to be an excellent way to get converging information on distal outcomes
Weaknesses	None noted
Publisher/Price	Public domain

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996a). Mechanisms of moral disengagement in the exercise of moral agency. *Journal of Personality and Social Psychology, 71*, 364–374.

Caprara, G. V. & Pastorelli, C. (1993). Early emotional instability, prosocial behavior and aggression: Some methodological aspects. *European Journal of Personality, 7,* 19-36.

Pastorelli, C., Barbaranelli, C., Cermak, I., Rozsa, S., & Caprar, G. V. (1997). Measuring emotional instability, prosocial behavior and aggression in pre-adolescents: A cross-national study. *Personality and Individual Differences, 23*(4)), 6911-703.

Measure	Multidimensional Self-Concept Scale (MSCS)
Constructs	Self-Perception
Age range	9 to 19 years (Wilson, 1998, suggests can be extended downward to 3rd and 4th graders)
Rating type	Self
Description of measure as related to construct of interest	The MSCS is designed to measure multiple context-dependent dimensions of self-concept. Self-concept here is viewed as a multidimensional behavioral construct (Degulach, 1992), and the MSCS measures six contextual domains in six scales of 25 items each: Social competence related to interactions with others Success/failure in attainment of goals Recognition of affective behaviors Academic achievement and competence in other school-related activities Competence related to interactions with family members Physical attractiveness and prowess
Administration	150 items, 4-point Likert scale format; typical time for completion is 30 minutes. Can be administered in groups or individually.
Scoring	Summed Likert scales for each component as on answer sheet; manual reflects age norms.
Reliability	Internal consistency ranges from .97 to .99 for full scale and .85 to .97 for subscales.
Validity	Evidence of content and construct validity (Bracken, 1992; Degulach, 1992). Concurrent validity with four other self-concept measures between .73 and .83. Predictive/concurrent validity with sociometric groups also found (Jackson & Bracken, 1998).
Strengths	Rigorous and extensive psychometric testing. Useful for both clinical and research applications. Subscales solidly linked to a well-established theoretical framework.
Weaknesses	None noted.
Publisher/Price	Published by Pro-Ed, Inc. Complete Kit $97.00. 50 extra scoring sheets $51.00

Bracken, B. (1992). Multidimensional Self-Concept Scale. Austin, TX: Pro-Ed.

Delugach, R R. (1992). Self-concept: Multidimensional construct exploration. *Psychology in the Schools, 29*, 213-223.

Jackson, L. D. Bracken, B. A. (1998). Relationship between students' social status and global domain-specific self-concepts. *Journal of School Psychology, 36,* 233-246.

Wilson, P. L (1998). Multidimensional Self-Concept Scale: An examination of grade, race, and gender differences in third through sixth grade students' self-concepts. *Psychology in the Schools, 35*, 317-326.

Measure	Relationship Questionnaire (Rel-Q)
Constructs	Self-Awareness, Social Awareness, Relationship Skills, Responsible decision- making
Age range	In Schultz, Selman & LaRusso's 2003 study, 4th -12th graders; could be used with somewhat younger children
Rating type	Self
Description of measure as related to construct of interest	Designed for the evaluation of school-based character education programs Multiple choice measure of psychosocial maturity derived from developmental theory that identifies the capacity to differentiate and coordinate the social perspectives of self and other to be central to treating other people in respectful and ethical ways. Items in the measure pose dilemmas common social situations with peers or adults. Uses a model of relationship awareness that includes 3 social-cognitive constructs which are: 1) interpersonal understanding or social reasoning 2) interpersonal negotiation strategies or conflict resolution in thought and action and 3) awareness of the personal meaning of relationships. Relationship framework, in describing how social-cognitive capacities underlie social interaction and development, addresses a common critique of cognitive-developmental approaches to sociomoral research: that there is little relation between the development of social cognition and real-world action.
Administration	Group administration
Scoring	24 questions comprising 5 scales: understanding of interpersonal relationships (6 questions), perspective-taking (4 questions), hypothetical interpersonal negotiation (4 questions), real-life interpersonal negotiation (4 questions), and awareness of personal meaning (6 questions) 2 Rel-Q scales represent interpersonal understanding (understanding and perspective-taking, two scales represent interpersonal skills (hypothetical and real-life interpersonal negotiation) and the fifth Rel-Q scale represents the third competency, (inter)personal meaning awareness.

	The questions on the measure each have four multiple choice responses, which represent points in the continuum on four theoretical levels in the coordination of social perspectives, ranging from egocentric (Level 0) to unilateral (Level 1) to reciprocal (Level 2) to mutual (level 3). For example, in the hypothetical negotiation questions, a situation involving a conflict between a particular protagonist and another person (either an adult or peer) in presented, followed by four actions the protagonist could take to respond to the disequilibrium. Subjects rate each multiple-choice response on a four point Likert scale (poor, average or "OK", good, excellent), then choose the "best" response of the four responses. This method yields two Rel-Q scores from each item (and for each scale): a "response rating "score, based on students' separate ratings of the four multiple choice responses, and a "best response" score based on which response they choose as "best." "Best response" scores for each of the five Rel-Q subscales (social perspective coordination, interpersonal understanding, hypothetical and real-life negotiation, and personal meaning awareness) are computed by averaging the "best response" scores for each question in that domain. Similarly, the response rating subscale scores are computed by averaging the response rating scores for each question in that subscale. The overall best response and response rating scores are computed by averaging the five subscale scores. Because both the best response and response rating scores have similar developmental level metrics, they can be averaged into a composite score for each subscale, and then into one overall psychosocial maturity score.
Reliability	The relatively low internal consistency reliabilities at the Rel-Q subscale level in the Schultz, Selman & LaRusso study (2003) seem to speak more to the contextual nature of the constructs being measured than to the Rel-Q's measurement of the constructs. The high reliability of the overall score reflects the deep structure of social perspective coordination upon which the developmental levels of each of the psychosocial competencies (i.e., the subscales) are based.
Validity	In the Schultz, Selman & LaRusso study (2003), the moderate effect size of the correlation of level or moral judgment with all five psychosocial competency scales suggests that the measures share a developmental orientation but assess separate constructs, supporting the Rel-Q's validity.

Strengths	This evaluation measure can assess mature psychosocial choice due to either development or socialization (including school climate) and can help clarify the extent to which character education programs not only teach social awareness and interpersonal skills but also engage students in personally meaningful ways. Shows sensitivity to group differences based on differential socialization, with significant differences in mean psychosocial maturity between schools. The Rel-Q can assess transformational (formal) change, or qualitative change in form, pattern, or organization, whereas other social skill instruments measure only variation (functional) change, or the extent that what is changing varies from an assumed standard in a quantitative sense. The Rel-Q's assessment of transformational change in social maturity provides a unique tool for evaluating programs that aim to promote character development rather than "better" behavior. The multiple-choice methodology of the Rel-Q seems to be a valid assessment of psychosocial maturity in individuals and groups (if subjects attend to the task carefully).
Weaknesses	Difficult conceptualization and potentially difficult scoring
Publisher/Price	Public domain as far as we know

Schultz, L. H., Barr, D. J., & Selman, R. L. (2001).The value of a developmental approach to evaluating character development programmes: An outcome study of Facing History and Ourselves. *Journal of Moral Education, 30*, 3-27.

Schultz, L.H. & Selman, R.L. (2000) The meaning and measurement of social competence from a developmental perspective, Working Paper 153 (New York, Russell Sage Foundation).

Schultz, L. H. & Selman, R. L. (2004). The development of psychosocial maturity in young children: A Measure for evaluating character education programs. *Journal of Research in Character Education, 2,* 19-87.

Schultz, L.H., Selman, R.L., & LaRusso, M.D. (2003). The assessment of psychosocial maturity in children and adolescents: Implications for the evaluation of school-based character education programs. *Journal of Research in Character Education, 1,* 67-87.

Measure	Resiliency Inventory
Constructs	Self-Awareness, Self-Management, Social Awareness, Relationship Skills
Age range	Adolescence, although it has been used successfully with children as young as 4th grade
Rating type	Self
Description of measure as related to construct of interest	Adapted from the Resilience Inventory (Noam & Goldstein, 1998), a 44-item measure of adolescence resilience. (Song 2003) Taps various domains of resilience: • The Optimism (OP) subscale concerns the respondent's positive perspective on the world and the future. • The Self-Efficacy (SE) subscale concerns confidence in the respondent's ability to deal with situations or things effectively. One who scores high on this subscale can think of a situation from different perspectives and evaluate alternative actions to deal with it more efficiently. • The Relationships with Adults (RA) subscale concerns support from and perceptions of adults. Research suggests that even one supportive figure outside of the family can make children and adolescents feel they are special and important, thereby making them more capable of handling stressors. • The Peer Relationships (PR) subscale centers on the respondent's relationship with friends. A person who scores high on this factor is likely to be popular among friends and have fun with them, thus having an active social life. Quality of peer relationships is reported to play an important role in competence in childhood and adolescence. • The Interpersonal Sensitivity (IS) subscale concerns the respondent's initiative to help others and improve one's surroundings. Resilient children are empathic of the needs of others and willing to care for them. A person who scores high on this subscale takes the initiative to help others and is sensitive to others' feelings. • The Emotional Control (EC) subscale concerns the respondent's ability to control himself or herself emotionally. A person who scores high on this factor is not likely to be agitated by a triggering situation but is likely to endure and be patient.
Administration	Participants are asked to indicated how well each item describes them, using a 5-point Likert-type response scale (1 = always false, 2 = usually false, 3 = half the time true, half the time false, 4 = usually true, 5 = always true). At the end, 11 open-ended questions are included with the aim of providing responses that would permit further refinement of the measure. (Song 2003)
Scoring	1- to 5-point scale, "always false" to "always true"

Reliability	Estimates of the internal consistency reliability for the six subscales ranged from .61 to .81, with test-retest reliability estimates ranging from .57 to .79. (Song 2003).
Validity	Construct validity/discriminant validity was supported by (a) positive correlations with measures of self-esteem, locus of control, social support and (b) a negative correlation with a measure of hopelessness. (Song 2003). Criterion-related validity was supported by (a) positive correlations with teacher ratings of students' resilience; and (b) negative correlations with self-report symptom outcomes. (Song 2003)
Strengths	The original RI was developed with the aim of becoming a cross-culturally sensitive and psychometrically sound measure of adolescence resilience. As a whole, the results from the present study successfully demonstrated that this adaptation is a reliable and valid measure in the Korean context (Song 2003). Although there are important cultural differences in item response patterns across cultural groups, factorial invariance was also found in Korean and American young adolescent samples.
Weaknesses	Still in development but very promising to get at specific distal outcomes of interest.
Price	Public domain

Noam, G. G., & Goldstein, L. S. (1998). *The Resiliency Inventory*. Unpublished Manuscript.

Song, M. (2003). Two studies on the resilience inventory (RI): Toward the goal of creating a culturally sensitive measure of adolescence resilience. Unpublished Doctorial dissertations, Harvard University.

Section 3: Academic-related SEL Competencies
This section includes measures on the following short-term student outcomes (see Table 4):

- Lack of internalizing behavior
- Lack of externalizing behavior
- **School engagement** (e.g., social connections in the classroom, classroom participation, academic motivation), as captured in the following
    - **Feelings about School**
- **Academic Competence** (e.g., GPA, achievement test scores)

Table 4 summarizes the following measures. Measures included are excellent exemplars of their core construct; although some assessment tools may be omitted, these were judged to focus on the areas of SEL outcomes that require focused attention.

Table 4. Short-term Student Outcomes: Behavior Problems, Feelings About School, and Academic Competencies *

Scale Name	Short-term Student Outcomes			School/Grade Level		Rating Type						
	Externalizing/Internalizing Behavior Problems	Feelings about School	Academic Competencies	Preschool	Elementary School	Teacher	Parent	Peer	Self	Observational	Performance-based	Other (e.g., interview)
The Devereux Early Childhood Assessment (DECA, see also Table 3 for placement in compendium)	■			■		■	■					
Preschool Learning Behaviors Scale (PLBS)			■	■		■						
Social Competence and Behavior Evaluation (see also Table 3 for placement in compendium)	■			■		■	■					
Behavior Assessment System For Children, Second Edition (BASC-2, see also Table 3 for placement in compendium)	■		■	■	■	■	■		■			
Child Behavior Checklist (CBCL) and Teacher Report Form (TRF)	■			■	■	■	■					
Sociometric Ratings and Nominations (see also Table 3 for placement in compendium)	■			■	■			■	■			
Social Skills Rating System, Social Skills Improvement System (see also Table 3 for placement in compendium)			■	■	■	■	■		■			

* Greyed cells indicate SEL core competency assessed, age level, or rating type.

Scale Name	Short-term Student Outcomes			School/Grade Level		Rating Type						
	Externalizing/Internalizing Behavior Problems	Feelings about School	Academic Competencies	Preschool	Elementary School	Teacher	Parent	Peer	Self	Observational	Performance-based	Other (e.g., interview)
The Teacher Rating Scale of School Adjustment (TRSSA)		■	■	■	■	■						
Child/Teacher/Parent Rating Scale (see also Table 3 for placement in compendium)	■		■		■	■	■	■	■			
Feelings About School (FAS) (see also Table 3 for placement in compendium)		■			■				■			
Learning Behaviors Scale			■		■	■						
Revised Children's Manifest Anxiety Scale (RCMAS)	■				■				■			
Prosocial and Aggressive Behavior (see also Table 3 for placement in compendium)	■				■	■	■	■	■			
Sense Of Classroom as a Community Scale - "Feelings About My Classroom"		■			■				■			

Measure	Preschool Learning Behaviors Scale (PLBS)
Constructs	Academic Competencies
Age range	Preschool
Rating type	Teacher
Description of measure as related to construct of interest	Teachers rate children's approaches to learning using the PLBS (PLBS; McDermott, Leigh, & Perry, 2002). In general, content focuses on attentiveness, responses to novelty and correction, observed problem-solving strategy, flexibility, reflectivity, initiative, self-direction, and cooperative learning.
Administration	29-item teacher behavior rating instrument assessing preschool children's approaches to learning.
Scoring	Teacher rated children's specific observable behaviors that occurred during classroom learning activities over the previous two months on a 3-point Likert scale. The instrument yields three reliable learning behavior dimensions: (a) competence motivation (i.e., reluctant to tackle a new activity); (b) attention/persistence (i.e., tries hard, but concentration soon fades and performance deteriorates); and (c) attitudes toward learning (i.e., doesn't achieve anything constructive when in a sulky mood).
Reliability	High internal consistency estimates from a national standardization sample were found for the three learning behavior dimensions (α = .87, .88, and .78, respectively). In the current study, the PLBS demonstrated adequate internal consistency (α = .79 to .89).
Validity	Multi-method, multi-source validity analyses further substantiated the PLBS dimensions for preschool children, and reliability estimates were similar for both White and non-White portions of the sample (Fantuzzo, Perry & McDermott, 2004).
Strengths	• Lack of observer effects indicate that PLBS assessments are grounded primarily in distinguishing child features rather than idiosyncratic observer responses. • The PLBS offers a structured and reliable approach to understanding early learning behaviors. • It also is convenient from the teacher's perspective and unobtrusive from the child's.

Weaknesses	---
Publisher/Price	Please contact Dr. Paul McDermott, Graduate School of Education, University of Pennsylvania, 3700 Walnut Street, Philadelphia, PA 19104, andreamc@voicenet.com

Fantuzzo, J., Perry, M. A., & McDermott, P. (2004). Preschool Approaches to Learning and Their Relationship to Other Relevant Classroom Competencies for Low-Income Children. *School Psychology Quarterly, 19*(3), 212-230.

McDermott, P. A., Leigh, N. M., & Perry, M. A. (2002). Development and validation of the Preschool Learning Behaviors Scale. *Psychology in the Schools, 39*, 353-365.

Measure	Child Behavior Checklist (CBCL) and Teacher Report Form (TRF)
Constructs	Self-Management, Responsible decision-making, (Lack of Risky, Disruptive/Externalizing and Internalizing Symptoms)
Age range	Ages 4-18
Rating type	Teacher, Parent
Description of measure as related to construct of interest	The CBCL is a 100-item parent-report measure used to assess problem behaviors in young children. The CBCL provides a total problem behavior score, two broadband scores for internalizing and externalizing behaviors, and seven specific syndrome scores (emotionally reactive, anxious–depressed, somatic complaints, withdrawn, sleep problems, attention problems, aggressive behavior). The C-TRF is also a 100-item measure and is completed by teachers or caregivers. Like the CBCL, the C-TRF is used to assess problem behaviors and yields a total problem behavior score, two broadband scores for internalizing and externalizing behaviors, and specific syndrome scores. The C-TRF excludes sleep problems, to include only six specific syndrome scores. The clinical scales contain a Total Problems score, two broadband dimensions (Internalizing Problems and Externalizing Problems), and eight empirically validated syndromes (Aggressive Behavior, Delinquent Behavior, Withdrawn, Somatic Complaints, Anxious/Depressed, Attention Problems, Social Problems, and Thought Problems).
Administration	Parent Report, Teacher Report
Scoring	Asks raters to make ratings from 0 to 2 depending on the extent to which a particular statement describes their youth: 0 = not true, 1 = somewhat or sometimes true, and 2 = very true or often true. The CBCL and TRF have been standardized to obtain normative points (i.e., what is typically reported by such informants for normative samples of youth; see Achenbach & Edelbrock, 1983, 1986; Edelbrock & Achenbach, 1984). As such, standard T scores quantify a youth's standing in relation to other youth and determine whether elevated scores on a particular scale falls in a clinical range. Recently, six scales were developed to correspond to DSM-IV disorders often present in youth and adolescents (Achenbach et al., 2003): affective, anxiety, somatic, ADH, OD, and conduct problems.
Reliability	Test–retest reliability coefficients ranged from .68 to .92 for the CBCL and .57 to .91 for the C-TRF (Griffith et. al. 2008). CFA of the CBQ and CBCL yielded three meaningful constructs (negative emotionality, internalizing behavior, and externalizing behavior) with satisfactory

	internal consistencies (.68 < □□< .85). (Paulussen-Hoogeboom et. al. 2008)
Validity	Achenbach and Rescorla (2001) found that the eight-factor correlated model had acceptable fit on the basis of the root-mean-square error of approximation (RMSEA) of .06. In one study (Griffith et. al. 2008), the three total scores, internalizing, externalizing, and total problems, of both the CBCL and the C-TRF were used. The CBCL and the C-TRF were selected for convergent validity analyses because of their extensive use in both clinical and research settings and their previous demonstrations with sound psychometric properties (i.e., content, construct, and convergent validity and interrater, test–retest, and long-term reliability coefficients; see Achenbach & Rescorla, 2000, for previous psychometric information).
Strengths	The CBCL and the C-TRF have been well researched and demonstrated to be psychometrically sound measures. Both the CBCL-A and the TRF-A were found to have high internal consistency and to adequately discriminate between youth with and without diagnosed anxiety disorders. Construct validity of the scales were supported by high correlations with other reliable anxiety measures (e.g., MASC, RCMAS). Additionally, both scales displayed sensitivity to treatment effects.
Weaknesses	Whereas the CBCL and C-TRF are both widely used and researched measures for identifying children with EBD, neither are intended for use as universal screeners, and are thus not particularly efficient for use with large numbers of children.
Publisher/Price	PAR, Inc. $435 and up.

Achenbach, T. M., & Edelbrock, C. (1983). Manual for the Child Behavior Checklist and Revised Child Behavior Profile. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M., & Edelbrock, C. S. (1986). *Manual for the Child Behavior. Checklist.* Burlington: University of Vermont

Achenbach, T. M., McConaughy, S. H. (2003). The Achenbach System of Empirically Based Assessment. In C. R. Reynolds & R. W. Kamphaus. (Eds.); *Handbook of psychological and educational assessment of children: Personality, behavior, and context (2nd Ed.),* pp. 406-430. New York, NY, US: Guilford Press, 2003.

Achenbach, T. M., & Rescorla, L. A. (2000). Manual for the ASEBA Preschool Forms & Profiles. Burlington: University of Vermont, Department of Psychiatry.

Achenbach, T. M., & Rescorla, L. A. (2001). Manual for the ASEBA School-Age Forms & Profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Edelbrock, C. S. & Achenbach, T. M. (1984). The teacher version of the child behavior profile: Boys aged 6-11. *Journal of Counseling and Clinical Psychology, 52,* 207-217.

Kendall, P. C., Puliafico, A. C., Barmish, A., Choudhury, M. S., Henin, A., & Treadwell, K. S. (2006) Assessing anxiety with the Child Behavior Checklist and the Teacher Report Form. *Journal of Anxiety Disorders, 21,* 1004-1015.

Griffith, A. K., Nelson, J. R., Epstein, M. H., & Pederson, B. (2008) Convergent Validity of the Early Childhood Behavior Problem Screening Scale. *Journal of Early Intervention, 30,* 282-294.

Paulussen-Hoogeboom, M. C., Stams, G. J. J., Hermanns, J. M. A., Peetsma, T. T. D., Van Den Wittenboer, G. L. H. (2008) Parenting style as a mediator between children's negative emotionality and problematic behavior in early childhood. *Journal of Genetic Psychology, 169*, 209–226.

See also:

Crowley, S. L., & Merrell, K. W. (2003)The structure of the School Social Behavior Scales: A confirmatory factor analysis. *Assessment for Effective Intervention, 28(2),* 41-55.

Emerson, E. N., Crowley, S. L, & Merrell, K. W. (1994) Convergent validity of the School Social Behavior Scales with the Child Behavior Checklist and Teacher's Report Form. *Journal of Psychoeducational Assessment, 12,* 372-380.

Measure	The Teacher Rating Scale of School Adjustment (TRSSA)
Constructs	Academic Competencies, Feelings about School, Teacher-Child Relationship
Age range	Preschool (has been used with preschool teachers) and Primary Grades
Rating type	Teacher
Description of measure as related to construct of interest	The Teacher Rating Scale of School Adjustment (TRSSA) was developed to tap several constructs that are reflective of young children's behavioral and relational adjustment to school or classroom settings. These constructs (subscales) have been labeled: (1) independent participation, (2) cooperative participation, (3) teacher's perception of children's school liking, (4) teacher's perception of children's school avoidance, and (5) teacher's perception of children's interest/comfort with the teacher. Independent participation is the degree to which children display autonomous, self-reliant behavior toward classroom activities and learning tasks. From a motivational perspective, this form of participation may reflect an intrinsic motivational, autonomous, or learning-goal orientation. Cooperative participation is defined as children's willingness to adhere to the social rules and role expectations of the classroom and, from a motivational perspective, may reflect the presence of cooperative and compliant goals (Wentzel, 1991). Children high on this dimension conduct themselves in a cooperative and responsible manner in response to teacher and classroom demands.
Administration	Teacher Ratings
Scoring	Later TRSSA Subscales: Cooperative Participation (7 items), Independent Participation (4 items), School Liking (5 items); School Avoidance (5 items): Teachers rate children from 0-2 on the following scales where: 0 = Doesn't apply 1 = Applies sometimes 2 = Certainly applies
Reliability	Excellent.
Validity	Excellent; relations with school achievement and social relations.
Strengths	Assesses a number of important aspects of children's school lives in one instrument. Excellent psychometrics in general
Weaknesses	Needs for second version, and shortened instrument (Betts & Rotenberg, 2007) reflected a few weakness in terms of delineating the scales clearly.

Publisher/Price	Public Doman

Betts, L. R., & Rotenberg, K. J. (2007). A short form of the Teacher Rating Scale of School Adjustment. *Journal of Psychoeducational Assessment, 25(2),* 150-164.

For research using the TRSSA, see:

Birch, S. H. & Ladd, G. W. (1998). Children's interpersonal behaviors and the teacher-child relationship. *Developmental Psychology, 34,* 934-946.

Buhs, E. & Ladd, G. W. (2001). Peer rejection in kindergarten: Relational processes mediating academic and emotional outcomes. *Developmental Psychology, 37*, 550-560.

Kochenderfer, B. J. & Ladd, G. W. (1996). Peer victimization: Manifestation and relations to school adjustment. *Journal of School Psychology, 34,* 267-284.

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

Ladd, G. W. & Burgess, K. B. (1999). Charting the relationship trajectories of aggressive, withdrawn, and aggressive/withdrawn children during early grade school. *Child Development, 70*, 910-929.

Ladd, G. W. & Burgess, K. B. (2001). Do relational risks and protective factors moderate the linkages between childhood aggression and early psychological and school adjustment? *Child Development, 72*, 1579-1601.

Ladd, G. W. & Kochenderfer, B. J. & Coleman, C. C. (1997). Classroom peer acceptance, friendship, and victimization: Distinct relational systems that contribute uniquely to children's school adjustment? *Child Development, 68,* 1181-1197.

Wentzel, K. R. (1991). Social competence at school: Relation between social responsibility and academic achievement. *Review of Educational Research, 61*, 1-24.

Note. There is also a child report version on school liking and avoidance created by Gary Ladd, see, for example:

Ladd, G. W., Buhs, E. & Seid, M. (2000). Children's initial sentiments about kindergarten: Is school liking an antecedent of early classroom participation and achievement? *Merrill-Palmer Quarterly, 46*, 255-279.

Measure	Learning Behaviors Scale (LBS)
Constructs	Academic Competencies
Age range	Students aged 5-17 years
Rating type	Teacher
Description of measure as related to construct of interest	A measure of differential patterns of classroom learning for students aged 5–17 years. (McDermott, 1999).
Administration	29-item observation device completed by a child's teacher along dimensions of student competence motivation, attention/persistence, strategy/flexibility, and attitudes toward learning.
Scoring	It is composed of 29 positively and negatively worded items (each of which presents a specific learning-related behavior) to reduce response sets and is rated on a 3-point scale (0 = does not apply, 1 = sometimes applies, 2 = most often applies; McDermott, 1999). The valence (positive or negative of item wording is varied as a measure to reduce response sets. (McDermott 1999) Of the 29 items, 25 are used to produce a total score and the four subscales include Competence Motivation (CM), Attention and Learning Attitudes (AL), Attention/Persistence (AP), and Strategy/Flexibility (SF). Five items crossloaded on two factors during creation of the measure, and are thus included on two LBS scales. Total and subscale raw scores are converted to normalized T scores (M = 50, SD = 10) based on the nationally representative standardization sample of 1,500 students aged 5 to 17.
Reliability	Alpha coefficients were computed for the dimensions across the national sample and within pertinent subsamples. All exceeded .70 (McDermott 1999) Stability coefficients were substantial (*M* = .92, range .93- .91) and statistically significant at *p*< .0001. (McDermott 1999) Further, rater reliability is excellent.
Validity	*Convergent and divergent validity is excellent.* *Incremental validity is excellent, in that* overlap with IQ is modest (about 12% in McDermott, 1999).

Strengths	The items are uniquely behavioral and require no inferences concerning mediating thoughts or feelings. Evidence has accumulated that the LBS provides a normative approach assessing basic learning behaviors in a way that is uniform across variations in age, gender, ethnicity, social class, and family/community structures. (McDermott 1999)
Weaknesses	---
Publisher/Price	Please contact Dr. Paul McDermott, Graduate School of Education, University of Pennsylvania, 3700 Walnut Street, Philadelphia, PA 19104, andreamc@voicenet.com

Buchanan, H. H., McDermott, P. A., & Schaefer, B. A. (1998). Agreement among classroom observers of children's stylistic learning behaviors. *Psychology in the Schools, 35*, 355-362

McDermott, P. A. (1999) National scales of differential learning behaviors among American children and adolescents. *School Psychology Review, 28*, 280-291.

Measure	Revised Children's Manifest Anxiety Scale (RCMAS)
Constructs	Internalizing Symptoms
Age range	Children and adolescents ages 5-19 or- 6 to 17 years (O'Toole et al., 2006)
Rating type	Self
Description of measure as related to construct of interest	Self-report measure completed by youth as a measure of manifest anxiety. Designed to assess the presence of generalized and nonspecific or nonsituational anxiety (i.e., trait anxiety). (Varela & Biggs, 2006). Measures the level and nature of anxiety symptoms in children and adolescents aged 6 to 19 years.
Administration	Self-report
Scoring	37 items that are answered in yes/no format (e.g., "I worry a lot of the time"). (Healy& Rucklidge, 2006) A youth's raw score on the RCMAS is converted to provide five T scores—total score, physiological subscale score, worry subscale score, social subscale score, and a lie subscale score. Yields four scores: total anxiety score, Physiological Anxiety, Worry/ Oversensitivity, and Concentration Anxiety. (Varela & Biggs, 2006) In addition, a Lie Scale score is calculated. (O'Toole et al., 2006) The raw score may be converted into a standard score based on the respondent's age, ethnic origin, or both. For the total anxiety score, the T score scale was used, which has a mean of 50 and standard deviation of 10. Scores range from 0 to 28 (raw scores) or 0 to 100 (T scores). There are no clinical cutoffs available for the RCMAS, although it is suggested that "greater significance" should be attached to T scores greater than 60. Three subscales have been derived from the instrument, including: (a) Physiological Anxiety, (b) Worry/Oversensitivity, and (c) Social Concerns/Concentration. In addition, nine items concern a Lie scale.
Reliability	Moderate retest reliability (r = .68; Reynolds & Richmond, 1985). Cronbach's alphas for the student, parent, and teacher ratings respectively were total, .74, .84, .64; Worry/Oversensitivity, .82, .82, .80; Physiological Anxiety, .74, .69, .70; and Concentration Anxiety, .69, .66, .64. Total Fear Score both the FSSC–R and the RCMAS have good evidence of reliability and validity (Varela et al., 2006). Test–retest reliability is .94 after three weeks and .68 after nine months. (Saldinger et. al. 2004).
Validity	Possesses high internal consistency (a = .83, Reynolds & Richmond, 1985)

	The RCMAS was standardized on White, Black, Mexican American, Nigerian, high-IQ, and intellectually handicapped children. Coefficients of congruence across ethnic and gender groups range from .96 to .99, showing consistency of the structure of the RCMAS across nominal variables. Good evidence of validity (Varela et al., 2006) Good convergent and divergent validity. (O'Toole et al., 2006) The RCMAS correlates highly with other measures of trait anxiety, such as the State-Trait Anxiety Inventory for Children Trait Scale, providing considerable support for the construct validity of the instrument (Mather & Cartwright-Hatton, 2004). Concurrent validity of the RCMAS has been supported by its correlation with many anxiety measures, particularly the State-Trait Anxiety Inventory for Children (Healey & Rucklidge, 2006).
Strengths	High Validity Moderate-High Reliability Widely used as a measure of anxiety in this age group Good psychometric properties
Weaknesses	None noted
Publisher/Price	$99.50 for kit with 25 forms (there's also a 5-minute short form). Publisher is Western Psych. http://portal.wpspublish.com/portal/page?_pageid=53,234661&_dad=portal&_schema=PORTAL

Healey, D., & Rucklidge, J. (2006). An investigation into the psychosocial functioning of creative children: The impact of ADHD symptomatology. *Journal of Creative Behavior, 40*, 243-264.

Mather, A., & Cartwright-Hatton, S. (2004). Cognitive predictors of Obsessive-Compulsive symptoms in adolescence: A preliminary investigation. *Journal of Clinical Child and Adolescent Psychology 33*, 743-749.

O'Toole, K., Borden, K. A., & Miller C. (2006). Long-term psychosocial and adaptive outcomes in children with arteriovenous malformations. *Rehabilitation Psychology, 51,* 60–68.

Saldinger, A., Porterfield, K. & Cain, A. C. (2004). Meeting the needs of parentally bereaved children: A framework for child–centered parenting. *Psychiatry, 67,* 331-352.

Varela, R. E., & Biggs, B. K. (2006). Reliability and validity of the Revised Children's Manifest Anxiety Scale (RCMAS) across samples of Mexican, Mexican American, and European American children: A preliminary investigation. *Anxiety, Stress, and Coping, 19*, 67-80.

Measure	Sense of Classroom as a Community Scale – "Feelings about My Classroom"
Constructs	Feelings About School
Age range	Elementary School
Rating type	Self
Description of measure as related to construct of interest	Measures: Students' feelings of caring/supportiveness and autonomy and influence in their classroom and school Self-report measure of students' sense of the classroom as a community (Battistich et. al. 1997). Initial measure included items representing two elements of community: (a) students' perceptions that they and their classmates cared about and were supportive of one another (7 items: e.g., "students at this school work together to solve problems," "the students in this class really care about one another," "my class is like a family"); and (b) students' perceptions that they had an active and important role in classroom norm setting and decision making (10 items: e.g., "Teachers and students plan things together at this school.," " in my class the teacher and students decide together what the rules will be"). (Battistich et. al. 1997).
Administration	Self-report. Response scale: 1 = disagree a lot, 2 = disagree a little, 3 = neither agree nor disagree, 4 = agree a little, 5 = agree a lot. [R] indicates reverse-scored item.
Scoring	Reverse scored item: 3
Reliability	The internal consistency (Cronbach's alpha) of this Measure averaged .78 across the 3 years. (Battistich et. al. 1997).
Validity	Students' classroom level (N = 28) scores were related to conceptually related measures of classroom characteristics derived from classroom observations. The aggregated student scores were found to be strongly correlated with observational measures of student supportive and friendly behavior ($r = .60$, $p < .001$), student spontaneous prosocial behavior ($r = .61$, $p < .001$), the frequency of cooperative activities ($r = .55$, $p < .01$), and opportunities for student autonomy and influence ($r = .51$, $p < .01$) (Battistich et. al. 1997).
Strengths	Asks questions in a reliable manner that access children's feelings about their school as a community, an important aspect of attachment to school
Weaknesses	None noted

Battistich, V., Solomon, D., Watson, M., & Schaps, E. (1997). Caring school communities. *Education Psychologist, 32,* 137-151.