Abstract Title Page

Title:

Measuring Social-Emotional Skills to Advance Science and Practice

Authors and Affiliations:

Clark McKown
Nicole Russo-Ponsaran
Jason Johnson

Rush University Medical Center Department of Behavioral Sciences

Abstract Body

Background / Context

The ability to understand and effectively interact with others is a critical determinant of academic, social, and life success (DiPerna & Elliott, 2002). This fact is increasingly recognized in educational policy and practice (Common Core State Standards Initiative, 2010; Nagaoka, Farrington, Ehrlich & Heath, 2015; National Research Council, 2012). Recognizing that these skills are critical for school and life, a growing number of states have adopted standards requiring school systems to assess and address children's social-emotional learning, or SEL (Dusenbury, Zadrazil, Mart, & Weissberg, 2011).

Despite calls to treat SEL as a critical component of academic learning, and despite the expansion of SEL standards, <u>few tools are available for educators to assess SEL skill and to use assessment results to inform educational practice</u>. Appropriate SEL assessment should help educators to: (a) evaluate student and classroom progress towards meeting SEL standards, (b) clarify students' strengths and instructional needs, (c) understand the needs of students who are considered for intensive SEL instruction, and (d) decide what skills to teach to which students.

An area in particular need of scalable, feasible, usable, and scientifically sound assessment tools is <u>social-emotional comprehension</u>, which includes mental processes enlisted to encode, interpret and reason about social and emotional information. Social-emotional comprehension includes the abilities to infer others emotions from nonverbal cues, to take others' perspectives, to solve social problems, and to enlist cognitive strategies involved in self-control. Well developed social-emotional comprehension is associated with academic, social, behavioral, and other important life outcomes (McKown, Allen, Russo-Ponsaran, & Johnson, 2013; McKown, Russo-Ponsaran, Johnson, Russo, & Allen, 2015; McKown, Russo-Ponsaran, Allen, Johnson, & Warren-Khot, 2015). However, there are few tools educators can use to evaluate their students' social-emotional comprehension and use findings to guide instruction.

Purpose / Objective / Research Question / Focus of Study

As defined here, social-emotional comprehension comprises mental skills that may not have straightforward behavioral correlates. An important issue, then, concerns the optimal method to assess these skills. Although teacher report is widely used, because social-emotional comprehension involves mental processes, observers must make a high level of inference, potentially attenuating validity. Furthermore, self-report is only modestly correlated with skill level (Shrauger & Osberg, 1981) and vulnerable to social desirability response bias (Crowne & Marlowe, 1960).

An alternative that addresses these limitations is direct assessment, defined here as a method of measuring social-emotional comprehension through performance on items that demonstrate mastery of skills (McKown, 2015). This is distinct from and complementary to teacher report and self-report. Our work is guided by the premise that direct assessments are optimal for assessing social-emotional comprehension.

To address the need for direct assessments of social-emotional comprehension, we developed a web-based system called SELweb. SELweb assesses four dimensions of social-emotional comprehension, three of which are adapted from Lipton and Nowicki (2009)'s model. "Social Awareness," the ability to understand others' emotions, draws on research on nonverbal communication (Nowicki & Duke, 1994). "Social Meaning," the ability to interpret others' mental states, draws on research on theory of mind and perspective-taking (Happé, 1994; Wellman & Liu, 2004). "Social Reasoning," the ability to reason about social problems, draws

on social information-processing research (Bauminger, Edelsztein, & Morash, 2005; Crick & Dodge, 1994). Extending the Lipton and Nowicki (2009) model of social-emotional comprehension, we include "Self-Control," which includes mental processes involved in delaying gratification and controlling emotions to achieve a goal (Duckworth, 2011).

     This paper summarizes lessons learned from a four year IES-funded Goal 5 project to develop and evaluate SELweb. In keeping with the theme of this year's SREE conference, this paper will review scientific and practice lessons learned. In service to summarizing the science, the paper will summarize key findings from two studies of the psychometric properties of SELweb. In service to summarizing practice lessons learned, the paper will describe strategies used throughout the research and development process, and mid-course corrections made along the way, to ensure the practical usefulness of the research process and research data to education partners.

Setting
     This research took place in elementary school districts in 20 elementary schools spanning nine school systems in six American states.

Population / Participants / Subjects
     Participants included an ethnically and socioeconomically diverse sample of 8,881 children from kindergarten through third grades.

Intervention / Program / Practice
     SELweb is a web-based system to assess social-emotional comprehension in kindergarten through third grade. SELweb administration takes about 45 minutes. SELweb was designed to maximize usability in schools. Throughout the research project, education partners' input was sought and integrated into assessment procedures, and the design of assessment report output. From the first field trial, assessment results were shared with education partners to use to understand their students and programs. Next, we describe SELweb's modules.

     Social awareness. Six photographs of child faces with neutral facial expressions, including three girls and two ethnic minorities, were used to create the Social Awareness module. With FaceGen software (Singular Inversions, 1998), the photographs were digitized and altered into high-intensity displays of happy, sad, angry, and frightened. For each face and emotion, we created a set of 10 faces ranging from low- to high-intensity affect displays. From this item pool, several test forms were created. Faces were assigned to test forms to ensure a balance of emotions, intensities, and child faces within a given form. Sixteen to 20 items on each test form were included on more than one form. After each face was presented, children clicked to indicate whether the face reflected happy, sad, angry, scared, or just okay. Item scoring is described in Table 1. To adjust for differences in test form difficulties and thereby equate scores, item scores were summed and standardized within form.

     Social Meaning. We created 12 illustrated and narrated vignettes in which a character is disappointed, scared, sarcastic, lying, hiding feelings, or harboring a false belief. After each story, children were asked a question whose correct answer required accurate inferences about the story character's mental state. Item scoring is described in Table 1. Item scores were summed across vignettes.

     Social Reasoning. We created five illustrated and narrated vignettes involving ambiguous provocation and five involving peer entry. After each vignette, children selected: (a) a description of the problem, (b) a social goal, and (c) solution preference. Each question was scored as described in Table 1. Scores for each question were summed across vignettes and

standardized within test form to equate scores. To reduce respondent fatigue, we created five test forms with six vignettes each. Each form included three ambiguous provocation vignettes and three peer entry vignettes. Each vignette was included on three forms.

Self-Control. We developed a choice-delay task (Kuntsi, Stevenson, Oosterlaan, & Sonuga-Barke, 2001) and a frustration-tolerance task (Bitsakou, Antrop, Wiersema, & Sonuga-Barke, 2006). Scoring is described in Table 1.

Research Design

This research included three psychometric field trials in which SELweb and validation measures were concurrently administered. In all field trials, a subset of children completed SELweb twice during the school year, permitting an estimate of temporal stability.

Data Collection and Analysis

All students in kindergarten through third grade completed SELweb and the University's IRB granted a waiver of informed consent for the research team to use de-identified SELweb and academic data. School personnel administered SELweb in one or two group sessions.

Parents of all children in kindergarten through third grade were invited to consent to their child's participation in an "add-on" study. From add-on study students and their teachers we collected additional measures of social-emotional comprehension and teacher rating scales. SELweb data that were reasonably interpretable were shared with education partners for use in understanding their students and guiding instruction. Data analysis methods are integrated into the summary of Findings below.

Findings / Results

SELweb's composite score reliabilities averaged .84 and .85 and six-month temporal stabilities averaged .66 and .63 in field trials one and two, respectively. In all field trials, scores on the assessment modules fit a hypothesized four-factor model of Social-Emotional Comprehension. Using the Complex Samples facility in MPlus (Muthén & Muthén, 2012) to account for the nesting of students in classrooms, data the first two field trials yielded an excellent fit of module scores to a four-factor model (Trial 1: CFI = .95, RMSEA = .046, 90% CI .037 - .056; Trial 2: CFI = .95, RMSEA = .049, 90% CI = .044 - .055). The four-factor model fit the data significantly better than several alternatives (comparisons to the four-factor model, $\Delta\chi^2/df > 160$, p < .05).

Analyses supported SELweb's criterion-related validity. Using hierarchical linear models (Raudenbush, Bryk, & Congdon, 2004) to account for the nesting of students in classrooms, and controlling for age, IQ (in Field Trial 1), sex, and ethnicity, we found that overall performance on SELweb was positively associated with teacher reported social skill, peer acceptance, and academic competence, and negatively associated with teacher reported problem behavior. As seen in Appendix B, Table 2, standardized regression coefficients reflecting the association between SELweb performance and criterion measures averaged .31 (range = .21 to .46)

We tested convergent and discriminant validity with nested structural equation models. First, a model was constructed in which the four latent factors created from SELweb indicators were modeled as predictors of four parallel factors reflecting Social Awareness, Social Meaning, Social Reasoning, and Self-Control, created from alternative indicators. Paths between factors representing the same construct are "convergent" paths and paths between factors representing different constructs are "discriminant" paths. In two field trials, convergent paths were moderate to large and significant. When each convergent path was constrained to zero, it significantly reduced model fit. Discriminant paths were small and non-significant. When discriminant paths

were constrained to zero, model fit was not significantly changed. These findings support the convergent and discriminant validity of SELweb factor scores. See Figure 1 for a summary.

Educators reported that SELweb was easy to use and informative, but that they did not always know how to use assessment findings to guide action. In addition, they provided important, and sometimes surprising, feedback about SELweb features they found particularly useful, and the features they would like to see incorporated into the system to maximize its usefulness.

Conclusions

Both studies provided evidence that: (a) composite assessment scores exhibited high reliability, (b) together, assessment modules demonstrated a theoretically coherent factor structure, (c) factor scores demonstrated convergent and discriminant validity, and (d) a score reflecting overall SELweb performance demonstrated good evidence of criterion-related validity. In addition, SELweb is highly usable, easily scalable, and feasibly administered in school settings. Educators find SELweb informative, and further work is needed to ensure that SELweb assessment findings guide educators to use evidence-based instructional practices.

Appendices

Appendix A. References

Bauminger, N., Edelsztein, H. S., & Morash, J. (2005). Social information processing and emotional understanding in children with LD. *Journal of Learning Disabilities, 38*, 45-60.

Bitsakou, P., Antrop, I., Wiersema, J. R., & Sonuga-Barke, E. J. (2006). Probing the limits of delay intolerance: Preliminary young adult data from the delay frustration task (DeFT). *Journal of Neuroscience Methods, 151*, 38-44. doi:10.1016/j.jneumeth.2005.06.031

Common Core State Standards Initiative (2010). *English Language Arts Standards.* Washington, DC: Author.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115*, 74-101. doi:10.1037/0033-2909.115.1.74

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349-354. doi:10.1037/h0047358

DiPerna, J. C., & Elliott, S. N. (1999). Development and validation of the academic competence evaluation scales. *Journal of Psychoeducational Assessment, 17*, 207-225. doi:10.1177/073428299901700302

Duckworth, A. L. (2011). The significance of self-control. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 2639-2640. doi:10.1073/pnas.1019725108; 10.1073/pnas.1019725108

Dusenbury, L., Zadrazil, J., Mart, A., & Weissberg, R. (2011). *State learning standards to advance social and emotional learning: The state scan of social and emotional learning standards, preschool through high school*. Chicago, IL: CASEL.

Happé, F. G. E. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders, 24*, 129-154.

Kuntsi, J., Stevenson, J., Oosterlaan, J., & Sonuga-Barke, E. (2001). Test–retest reliability of a new delay aversion task and executive function measures. *British Journal of Developmental Psychology, 19*, 339-348. doi:10.1348/026151001166137

Lipton, M., & Nowicki, S. (2009). The social emotional learning framework (SELF): A guide for understanding brain-based social emotional learning impairments. *Journal of Developmental Processes, 4*, 99-115.

McKown, C., (2015). Challenges and opportunities in the direct assessment of children's social-emotional comprehension (pp. 320-335). In J.A. Durlak, C. Domitrovich, R.P. Weissberg, and T.G. Gullotta (Eds.), *Handbook of Social and Emotional Learning*. New York, NY: Guilford Press.

McKown, C., Allen, A.A., Russo-Ponsaran, N.M., Johnson, J.K. (2013). Direct Assessment of Children's Social-Emotional Comprehension. *Psychological Assessment*, 25, 1154-1166.

McKown, C., Russo-Ponsaran, N.M., Allen, A.A., Johnson, J., & Russo, J. (2015). Web-based direct assessment of children's social-emotional comprehension. *Journal of Psychoeducational Assessment.* Article published online.

McKown, C., Russo-Ponsaran, N.M., Allen, A., & Warren, H.K. (2015). Social-emotional factors and academic outcomes in elementary-aged children. *Infant and Child Development.* Article published online.

Nagaoka, J., Farrington, C.A., Ehrlich, S.B., & Heath, R.D. (2015). Foundations for Young Adult Success: A Developmental Framework. University of Chicago Consortium on Chicago School Research.

National Research Council. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century.* Committee on Defining Deeper Learning and 21st Century Skills, J.W. Pellegrino & M.L. Hilton (eds). Board on Testing and Assessment and Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Nowicki, S., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy scale. *Journal of Nonverbal Behavior, 18*, 9-35.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2004). *HLM 6 for windows* [computer software]. Skokie, IL: Scientific Software International, Inc.

Shrauger, J. S., & Osberg, T. M. (1981). The relative accuracy of self-predictions and judgments by others in psychological assessment. *Psychological Bulletin, 90*, 322-351. doi:10.1037/0033-2909.90.2.322

Singular Inversions. (2005). *Facegen main software development kit* (version 3.1). Vancouver, BC: Author.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75*, 523-541.

Appendix B. Tables and Figures

Table 1

Description of SELweb modules, questions, and item scoring

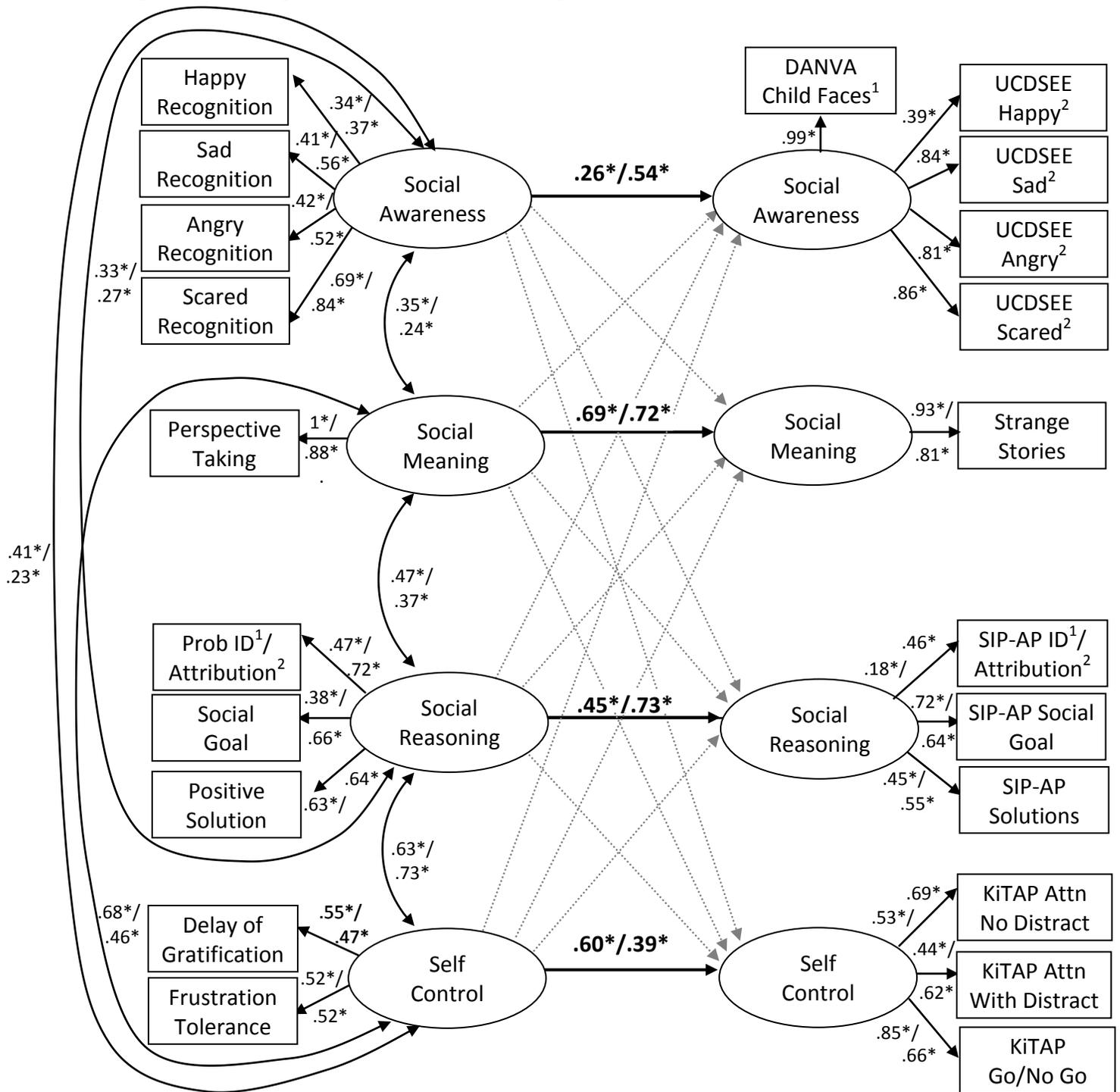| Module	Stimulus	Question and Response Options	Item Score
Social Awareness	Respondents view individual child faces and indicate emotion expressed.	What is the child feeling? Happy, Sad, Angry, Scared, Just OK	2 Correctly recognizes emotion; 1 Mistakes emotion for neutral; 0 Selects incorrect emotion
Social Meaning	Respondents hear illustrated, narrated vignette and must infer the mental state of a character. Example: A boy has a false belief about the location of a soccer ball and looks in the wrong place.	Questions about character intention (e.g., "Why did the boy look in the basket?") Illustrated, narrated forced choice, four possible responses.	2 Correct mental state inference (e "He *thinks* it is in the basket.") 1 Correct answer, no mental state inference (e.g. "He looks in the basket.") 0 Incorrect answer (e.g. "His brother told him to look there.")
Social Reasoning	Respondents hear illustrated, narrated vignettes involving either ambiguous provocation (e.g. getting bumped into by a classmate) or peer entry (e.g. trying to join an ongoing game of basketball).	**Problem Identification** (Study 1) What is the problem? Illustrated, narrated forced choice (e.g. "There is no problem"; "Someone bumped you"; "You feel bad"; "Someone bumped you & you feel bad")	2 Descriptive (e.g. "Someone bumped into you.) 1 Resilient (e.g. "There is no problem") 0 Reactive (e.g. "Someone bumped into you & you feel bad.")
		Attribution (Study 2) Did the person do it to be mean? Yes or no; if yes, a little or a lot?	2 "no" 1"yes" and "a little" 0"yes" and "a lot"
		Goal Preference How do you want it to turn out? Narrated forced choice with positive (e.g. "Become friends") or retribution (e.g. "Get back at them.") options	Study 1 1Positive goal; 0 Negative goal; Study 2 2 Positive goal; 1 Retribution goal; 0 Revenge goal
		Solution Preference What would you do? Illustrated, narrated forced choice, four response types (e.g. "Hit or yell at him,"; "Ask the teacher for help"; "Talk to him"; and "Walk away.")	2 Competent assertive (e.g. "Talk to him"); 1 Self-advocacy (e.g. "Ask the teacher for help") and ignoring (e.g. "Walk away"); 0 Aggressive (e.g. "Hit him.")
Self-Control: Choice Delay Task	Children send illustrated rocket ships to space. One is fast. One is slower. One is very slow.	Children are told to get as many points as possible in ten trials.	3 Slowest rocket; 2 Medium rocket; 3 Fast rocket
Self-Control: Frustration Tolerance	Children view pairs of shapes and indicate whether they match. Several items are programmed to get "stuck."	Children click on a "✓"if the shapes are the same and an "X" if they are different. Children do as many items as possible in 90 sec.	1 Correct response; 0 Incorrect response

Table 2

Criterion-related validity of social-emotional comprehension

	Criterion							
	Child Behavior			Academic Competence			Social Acceptance	
	SSIS[1]		DESSA-mini[2]	SSIS[1]	AIMSweb[1]		Peer Nominations	
Variable	Social Skills	Problem Behavior	Total Score	Academic Competence	Reading	Math	Social Preference[1]	Social Preference[2]
Age	-.04	.32*	-.23*	-.36*	-.17	-.32*	-.09	-.25
IQ	.06	-.09	-	.32*	.28*	.47*	.07	-
Sex	.30*	-.37*	.30*	-.18	.13	-.27*	-.05	.12
White	.32	-.68	-.12	.31	.30	.74*	.09	.20
Black	-.20	-.29	-	-.05	-.09	-.04	-1.66	-
Hispanic	.53	-.82*	-.14	.74	.10	.39	.26	.26
Asian	-.24	-.48	.38	1.12*	.14	1.32*	-.32	.61
SE Comprehension	**.28***	**-.36***	**.42***	**.46***	**.27***	**.21***	**.22**	**.25***

Notes: Coefficients are standardized; $* p < .05$; [1]Study 1; [2]Study 2; SSIS = Social Skills Improvement System rating scale; DESSA = Devereux Student Strengths Assessment; SE Comprehension = social-emotional comprehension.

Figure 1 Convergent and discriminant validity

IFI = .90/.93
RMSEA =.054/.047 (90% CI = .040 - .068/.036 - .058)

Note: * *p* < .05; coefficients are standardized. Coefficients before "/" are from Study 1; those after "/" are from Study 2. For simplicity of presentation, not all modeled covariates, errors and covariances are represented; [1]Study 1 only; [2]Study 2 only; DANVA = Diagnostic Analysis of Nonverbal Accuracy; UCDSEE = U.C. Davis Set of Emotion Expressions; SIP-AP = Social Information Processing Application; KiTAP = Test of Attentional Performance for Children.