Computerized Social-Emotional Assessment Measures for Early Childhood Settings

Susanne A. Denham [a, 1] sdenham@gmu.edu

Hideko H. Bassett [a] hhamada1@gmail.com

Katherine M. Zinsser [a] kzinsser@uic.edu

Isabel S. Bradburn [b] isbrad@vt.edu

Craig S. Bailey [a] craig.bailey@yale.edu

Elizabeth A. Shewark [a] lizshewark1@gmail.com

David E. Ferrier [a] david-ferrier@utc.edu

Kristi H. Liverette [a] kliverette23@gmail.com

Jessica Steed [a] jessica.steed1216@gmail.com

Samantha P. Karalus [a] spkaralus@fcps.edu

Saeid Kianpour [b] skian@vt.edu

[a] George Mason University, Department of Psychology, 4400 University Drive, MS3F5, Fairfax, VA 22030-4444

[b] Child Development Center for Learning and Research, Department of Human Development 0416, Virginia Tech, Blacksburg, VA 24061

The present study was funded by NICHD grant #R21HD068744 and IES grant #IES R305A110730. We are grateful to the many children, families, and teachers who participated in this study, and the directors of the facilities who so cooperatively worked with us. Special thanks go to Andrew Steed, Adrian Grajdeanu, and Dustin Rhodes for programming our computerized measures. We also thank Claire Christensen, Candace DeVito, Charlotte Edwards, Nicole Fettig, Grace Howarth, Sarah Lim, Sara O'Malley, Tanvai Rawat, Caitlin Tarr, Ruvi Tsokodayi, and Naomi Watanabe for their major contributions to this work, as well as the numerous other graduate and undergraduate research assistants who also administered the measures. [1] Corresponding Author: Susanne Denham sdenham@gmu.edu. Permanent Address: 1716 Seddon Rd., Richmond, VA 23227

Abstract

Social-emotional competence (SEC) is increasingly acknowledged by parents, educators, and lawmakers as central to school success. Given the tremendous SEC gains made by preschoolers, early childhood educators need access to sensitive assessment tools that enable them to monitor and tailor instruction to individual children's needs. Computerized direct assessment tools have several advantages to meet these needs, including inherent interest to children and ease of use for teachers. Thus, we evaluated the psychometric adequacy of computerized assessment tools measuring two key aspects of preschoolers' SEC: emotion knowledge and social problem solving. Participants included 450 preschoolers from three regions. We used two versions each of two measures widely used in research: The Affect Knowledge Test, Shortened (AKT-S) and Challenging Situations Task (CST). Both were administered via in-person and computerized modes, in counterbalanced orders. For each computerized administration, observers rated children's computer competence and interest in the assessment process. Analyses examined internal consistency reliability of the computerized measures. Interrelations and mean differences between computerized and in-person modes for each measure were used to demonstrate concurrent validity of the computerized measures. Because of the importance of SEC for early school success, associations of the computerized measures with aggregate teacher ratings of social-emotional behavior and learning behaviors/attitudes were used as indicators of predictive validity. Findings showed that the computerized AKT-S and CST appear reliable. Further, for concurrent validity, both are related to, and do not differ from, the in-person mode. Predictive validity relations were stronger for the AKT-S than the CST, therefore validity of the CST should be probed further. Discussion centers on advantages of using these computerized measures, and how teachers could be supported to use them.

Computerized Social-Emotional Assessment Measures for Early Childhood Settings

Ensuring that children are on positive school developmental trajectories at school entry is essential for later academic and social success. Previous research has highlighted social-emotional competence (SEC) as being especially important in establishing such success (Denham, Brown, & Domitrovich, 2010; Murray & Harrison, 2011; Romano, Babchishin, Pagani, & Kohen, 2010). In particular, the extant literature has identified educators as key promoters of Social Emotional Learning (SEL) across five important areas of competence: (a) self-management, the ability to regulate thoughts, emotions, and behaviors, (b) self-awareness, including the ability to recognize one's emotions, (c) social awareness of culture, beliefs, and feelings of others, (d) relationship skills, the abilities to effectively communicate with and work well with peers, and build meaningful relationships, and (e) responsible decision-making, including the abilities to make plans for the future, and solving social problems (Oberle et al., 2016; Payton et al., 2000). A preschooler who has attained such age-appropriate SEC skills is more likely than one who has not to become the kindergartner who can better plan and pay attention to academic tasks and devote more resources to learning. S/he can benefit more from teachers' instructions, sharing academic information and resources with peers and modeling peers' learning skills (Denham et al., 2010; Romano et al., 2010). In this study we focus on two SEC skills that are important during early childhood: emotion knowledge and social problem solving.

Growing Attention to SEC

There is increasing stakeholder acknowledgement that SEC is key in the transition from preschool to elementary school. Both parents and early childhood educators point to an urgent need for SEC programming, especially for children who live in low income families and/or are members of minority groups who historically have faced challenges to academic achievement (Bridgeland, Bruce, & Hariharan, 2013; Piotrkowski, Botsko, & Matthews, 2000). Further, all state education systems and Head Start programs now incorporate SEC skills, including emotion knowledge and social problem solving, in prekindergarten learning standards (albeit with fewer indicators and implemented less systematically in comparison to cognitive skills; see Dusenbury et al., 2015). National legislation has also been introduced in the US authorizing funds for technical assistance, training, and programming on SEC (O'Connor, DeFeyter, Carr, Luo, & Romm, 2017). As a result, SEC is receiving increasing attention and incorporation in early childhood instruction.

Therefore, because SEC skills are so important and viewed as such not only by academicians, but vehemently by early childhood educators, parents, and even legislators, it would behoove us to assess them well. However, educators lack the requisite psychometrically sound tools to identify, track and assess SEC skills (McKown, 2017). Without sound measurement tools to facilitate early assessment of SEC, educators cannot best tailor instruction to each child, and stakeholders cannot ascertain progress or whether a given SEC curriculum is more or less fruitful (Denham, 2015; Denham, Bassett, & Zinsser, 2012). Hence, our goal in this study is to promote useful SEC assessments measuring preschoolers' emotion knowledge and social problem solving.

SEC Assessment

There are, of course, definite requirements for such assessment tools (Denham, 2015; Kendziora, Weissberg, & Dusenbury, 2011). They should be developmentally appropriate, integrated with curricula, beneficial to all parties, primarily reliant on the child's everyday activities in realistic situations, culturally and linguistically responsive, and reflect children's actual performance (Copple & Bredekamp, 2009). Data emanating from such assessment should, moreover, not be used for high stakes decisions, such as retention in kindergarten. Instead, formative and summative functions of assessment should be undertaken to effectively identify children needing intervention or higher-level services, highlight specific needs of children and classrooms in terms of instruction, and show overall effects of programming (Denham, 2015; Denham, Ferrier, Howarth, Herndon, & Bassett, 2018).

With these stipulations in mind, our research team has developed several well-validated assessments for aspects of SEC, including direct in-person assessments of emotion knowledge with the Affect Knowledge Test – Shortened (AKT-S) and social problem solving with the Challenging Situations Task (CST). The AKT-S and CST measures have been widely used by researchers around the world (e.g., Leyva, Berrocal, & Nolivos, 2014; Kılıç, Ş., & Aytar, 2016; Rebelo, Verissimo, Machado, & Silva, 2013; Sette, Bassett, Baumgartner, & Denham, 2015; Sheard, Ross, & Cheung, 2013; Upshur, Heyman, & Wenz-Gross, 2017). In fact, one recent meta-analysis found that over a third of all effect sizes in studies of children's emotion socialization were measured with the AKT (Zinsser, Gordon & Jiang, 2019).

As theoretically predicted, research has consistently shown that the AKT and CST predict social competence, classroom adjustment, and academic readiness (e.g., Authors, 2012b, 2013a, 2013b, 2014b). The theoretical proposition underlying the AKT's relations with these attributes is that emotion knowledge "frees up" personal resources, allowing young children to interact positively with others in the preschool environment, and focus more specifically on learning. Regarding social problem solving as assessed by the CST, being calmly "ok" or sad instead of angry during a peer provocation may allow the child to take time to think about a prosocial solution, whereas the angry child may merely lash out aggressively, such that sad or "just ok" self-emotion choices on the CST are related to their teacher-rated social competence and classroom adjustment (Denham et al. 2013; Izard, Stark, Trentacosta, & Schultz, 2008; Schultz et al., 2010). In contrast, choosing happy emotion responses connotes lack of understanding or denial that a provocation has occurred (or perhaps social desirability in the context of assessment). Choosing angry emotion responses would more directly relate to deficits of social competence and classroom adjustment (Orobio de Castro, Merk, Koops, Veerman, & Bosch, 2005). Regarding behavior response choices on the CST, children who choose socially competent behaviors as part of response decision-making, as opposed to aggressive or passive behaviors, are likely to demonstrate greater social competence and overall classroom adjustment.

Despite these theoretical foundations, the popularity of these tools among researchers, and the evidence base supporting their use, their administration and scoring by trained examiners makes them prohibitive for classroom teachers' use. To provide educators with the most psychometrically rigorous tools, we have adapted these tools to maximize their potential utility and feasibility in preschool, Head Start, and childcare classrooms. Our goals are to (a) allow teachers to track students' progress and inform instruction (formative assessment), (b) measure child outcomes (summative assessment), and (c) evaluate program outcomes. To accomplish these goals, we adapt these in-person measures for electronic administration, using tablet devices.

In sum, given the growing importance of SEC and its concomitant assessment in the early childhood classroom, the focus of the current study is to evaluate the psychometric efficacy of two direct assessment tools measuring emotion knowledge and social problem solving, adapted for computer administration. The goal is, with continued improvement based on these evaluations, to use these tools to inform early childhood instruction. In accessing what the child knows and thinks, direct assessment can complement and add to information emanating from the observable and ratable behavioral aspects of SEC (McKown, 2017).

Use of Technology in Assessment

Why computerize measures of early childhood SEC? In the busy early childhood classroom, teachers must have easy-to-administer assessment tools. Tablets have become very common in preschool classrooms in the last two decades (e.g., Fletcher, Whitaker, Marino, & Anderson, 2014; NAEYC, 2012), and meet this need. Other principal advantages of computer-based systems over conventional assessment methods are that: (a) stimuli are standardized, making assessment more precise; (b) significant savings can be made in both time and labor; (c) scoring can be immediately available, without error-prone optical scoring; and (d) training for assessment administrators (i.e., teachers or educational resource personnel) can be dramatically reduced and streamlined. In attempts to computerize our preschool SEC measures, we strove to meet these criteria, to help teachers move toward use of electronic portfolios of both formative and summative SEC assessment.

Such assessment also can take advantage of the capabilities of technology for animation, speech, and sound. In fact, young children engage with educational software as soon as they can manipulate a touchscreen (NAEYC, 2012; Vandewater, Rideout, Wartella, Huang, Lee, & Shim, 2007). They show intense interest and pleasure, and surprising stamina, in interacting with computers (Ellis & Blashki, 2004; McCarrick & Li, 2007). Thus, computer use, whether for instruction or assessment, can be a highly motivating, positive experience for young children.

The Current Study

In summary, to facilitate teacher use of these SEC assessments in effective, efficient, and valid ways, it is crucial to develop appropriate means of standardizing and streamlining direct assessments emotion knowledge and social problem solving, via computer usage. Although our original research-based measures are valuable in predicting early school success, the training, coding, and administration requirements that attend them can be "deal breakers" for applied use in the early childhood classroom. If we are to make these means of assessment useful, to move toward both formative and summative assessment, we feel that they must be computerized, with much thought given to user support (for, e.g., preschool, Head Start, and childcare teachers, Head Start mental health consultants, and others). We also consider making the assessment attractive and fun to children to be an important goal, to maintain their engagement and enjoyment of the learning process.

In this study, we present findings from our direct assessments of emotion knowledge and social problem solving. A first step in creating usable assessment tools is to examine psychometric properties of the measures. To that end, data from in-person and computerized modes of each measure were collected. We used with two versions of each measure within each mode, to yield answers for four research questions. Research Question 1 (RQ1) asks whether internal consistency and between-version temporal stability for each version of the computerized mode are adequate. The adequacy of concurrent validity, via mean equivalence of in-person and computerized modes, is addressed in Research Question 2 (RQ2). For Research Question 3 (RQ3), we further examine the adequacy of the computerized versions' concurrent validity via associations between corresponding versions across in-person and computerized modes. Finally, in Research Question Four (RQ4) we determine the adequacy of the computerized versions' predictive validity via associations between the computerized mode and both teacher-rated SEC and learning behavior/attitudes (given the importance of emotion knowledge and social problem solving to early school success, and in accordance with the theoretical propositions related to these associations).

Method

Participants

Participants included 450 preschoolers from urban, suburban, and more rural preschools ($n_{suburban}$= 337; n_{urban}= 40; n_{rural} = 73). Because we collected data in three locations, we examined regional differences. Participants' average age across all regions was 51 months, but age differed across regions (F (2, 442) = 11.25, partial η^2 =.05, p < .001); mean age in the suburban group was 52 months, but in both the urban and rural samples, the average age was 48 months (ranges 35-71 mos, 35-60 mos, and 36-58 mos, respectively, for each region). The total sample included 48.1% boys and 51.9% girls, and gender distribution did not differ by region. One hundred fifty-one of the children came from families experiencing economic stress (as indicated by enrollment in Head Start or faith-based preschools serving these populations), and such children were overrepresented in the suburban group (χ^2 (2) = 41.03, Cramer's V = .30, p < .001). We compared the AKT-S and CST scores across regions to assess any group differences; Bonferroni comparisons following one-way ANOVAs showed region differences on fewer measures than expected by chance for computerized and in-person modes; thus, only cross-region results are presented.

All teachers reporting on demographics and computer usage in the classroom, and subsequently providing predictive validity ratings, were women, with a median age of 25-34 years. About eight percent of those reporting were Latina, whereas 78.8 % reported as Caucasian, 15.1% as African American, and 6.1 percent as Asian. Their median education was BA- level college graduation (with a median six courses in early childhood education), and median reported experience in early childhood settings was 10 – 12 years. Most reported no training in the past year regarding use of computers in their workplace, a majority (58.1 to 72.0 %) reported using computers for texting, email, or web usage in their daily lives. They reported that computers do not hinder their ability to complete their job, and that they ask for help when encountering a computer problem. They also noted that they used computers in their classroom work for an average of 7.49 hours per week (*SD* = 8.05), Finally, teachers reported most often having one computer in the classroom, with each child using it on average 14.4 minutes per day (*SD* = 13.1).

Measures

Social Problem Solving: **Challenging Situation Task (CST**; Authors, 2013a, 2013b**)**. The CST is a pictorial, forced choice measure, assessing children's ability to describe their own behavior choices and their attendant emotions. The CST includes six unambiguous physical and relational provocation scenarios (e.g., "Mary/John is having a good time playing in the sandbox when Bobby hits her/him", "Mary/John drew a picture of a dog. Bobby saw it and said, "It doesn't look like a dog. It looks like an ugly monster!" and started laughing"). These scenarios were presented via pictures and a short description of the transgression scenario. The picture depicted an androgynous character, given a name matching the child's gender.

Children were then asked, "If this happened to you, how would you feel?" about each situation, given four randomly ordered emotion choices using schematic drawings and verbal labels of 'happy,' 'sad,' 'angry,' and 'just ok'. Next, they were asked "What would you do?", and given four randomly ordered behavior choices reflecting socially competent (e.g., Tell him it's not a nice thing to do?), aggressive (e.g., Hit him?);

avoidant/passive (e.g., Go play somewhere else?), and crying (e.g., Cry?) for the first story above. For more details, see Supplemental File A.

For each of the four emotion and four behavior response choices, an aggregate scale score across the six scenarios was created, reflecting the number of endorsements of each emotion or behavior response type. Earlier work has demonstrated both test-retest reliability and validity for the original in-person CST, which was comprised of three scenarios (e.g., Authors, 2012b, 2012c, 2013c, 2014a; Bierman et al., 2014; Nix, Bierman, Domitrovich, & Gill, 2013). For this and subsequent study, we expanded the CST to include three more scenarios in version A, and created a parallel version B with six scenarios, so that relational provocations were included in the measure and the entire new version B was created.

Computerized adaptation of the CST. On the computer tablet, children heard audio explanations of each scenario and response for the Challenging Situations Task; all scenarios' content, their oral presentations, and use of randomization of stimuli were identical to the in-person mode. Further, the same picture cards depicting situation, attendant emotions, and response choices used in person were displayed onscreen, and children selected responses on the tablet's touchscreen. In every respect except mode of presentation the in-person and computerized versions were identical.

Emotion knowledge: Affect Knowledge Test Shortened (AKT-S; Authors, 2016**)**. The AKT-S assesses preschoolers' understanding of emotion using puppets with detachable felt faces that depict happy, sad, angry, and afraid expressions. For the *emotion labeling* portion of the measure (four items), children were asked to identify happy, sad, angry and afraid facial expressions by verbally naming them (expressive recognition; this subscale was administered in person only because there would be no way to record children's verbal reactions to the expressive stimuli on the computer table, so this score was not included in analyses and will not be discussed further), and then by pointing to them (receptive recognition). After this portion, the puppeteer briefly reviewed the correct emotion for each face to ensure that enough support had been given for subsequent portions of the measure. Then, for the *situation knowledge* portion, nine vignettes were enacted using the puppets, accompanied by vocal and visual affective cues emitted by the puppet/experimenter. For three stereotypical emotion knowledge vignettes, the puppet depicted the same negative emotion most people would feel (e.g., fear during a nightmare). In the remaining six nonstereotypical emotion knowledge vignettes, the puppet depicted emotions different from reports of the child's likely feelings (teachers provided this information). Among nonstereotypical situations, three vignettes pitted positive versus negative emotion (e.g., happy or sad to come to preschool); the rest pitted negative versus negative emotion (e.g., angry at or afraid of a peer's aggression). Children affixed the felt face of their choice to report the puppet's emotion.

Children received one point for correct identification in the labeling portion. In the situation knowledge portions, they received two points for correct identification of emotion and one point for identifying the correct (positive or negative) valence but not the correct emotion (e.g., sad for afraid). Mean scores for negative emotion recognition (because of virtually no variation for the happy expression), stereotypical situation knowledge, nonstereotypical knowledge were used in subsequent analyses. For more detail, see Supplemental File 2).

The in-person AKT-S has also demonstrated good internal consistency and validity via relations with social competence and early classroom adjustment (Authors, 2015). Further, prior research with the original AKT, a longer version encompassing all AKT-S items as well as others, has been related to two other measures of preschoolers' emotion knowledge (Morgan, Izard, & King, 2010).

Computerized adaptation of the AKT-S. As with the CST, all scenarios' content, their oral presentations, and use of randomization of stimuli were identical to the in-person mode. The receptive portion was performed with audio questions, as it was in person; that is, a videotaped puppeteer asked children to choose faces and select choices from pictured options that popped up on the screen following enacted scenarios. For situation portions, videos were displayed in which a puppeteer performed the situation scenarios identically the in-person AKT. For all items, children were asked to select emotions by touching one of four onscreen line-drawn faces (the same faces used in person). As with in-person administration, puppet ethnicity and gender were matched to the child's.

Classroom adjustment: Preschool Learning Behaviors Scale (PLBS). The PLBS (McDermott, Leigh, & Perry, 2002) is a 29-item teacher behavior rating instrument assessing preschool children's approaches to learning, used here as a predictive validity measure to show how our measures of emotion knowledge and social problem solving relate to early school success. On a 3-point scale, teachers rated children's observable behaviors that occurred during classroom learning activities over the previous two months. The instrument yields three reversed-scored learning behavior dimensions: Competence Motivation (10 items, e.g., "reluctant to tackle a new activity"), Attention/Persistence (nine items, e.g., "tries hard, but concentration soon fades and performance deteriorates"), and Attitudes Toward Learning (eight items, e.g., "doesn't achieve anything constructive when in a sulky mood"). Multi-method, multi-source validity analyses further substantiate the PLBS dimensions for preschool children, and reliability estimates were similar for Caucasian and non-Caucasian portions of the sample

(Fantuzzo, Perry, & McDermott, 2004). In this study, standardized scores for these scales were averaged to form the Learning Behaviors/Attitude score (α = .86).

Social Competence and Behavior Evaluation (SCBE-30). The SCBE-30 (LaFreniere & Dumas, 1996) measures socio-emotional competence of 3- to 6-year-olds and was also used as a validity measure to show how our measures of emotion knowledge and social problem solving relate to social aspects of early school success. Teachers provided ratings on six-point scales for child behaviors such as “easily frustrated" (Anger/Aggression scale), "avoids new situations" (Anxiety/ Withdrawal scale), and "comforts or assists children in difficulty" (Sensitivity/ Cooperation scale). LaFreniere and colleagues report excellent internal consistency and test-retest reliabilities, as well as construct and convergent validity (see also Authors, 2012a). Finally, in a multi-national study, structural equivalence was noted across demographic groups (LaFreniere et al., 2002). In this study, standardized scores for these scales (Anger/Aggression and Anxiety/ Withdrawal reversed) were averaged to form the Social Competence score (α = .69).

Observer ratings. After each session with computerized measures, research assistants rated children’s computer skills (e.g. “Child presses the screen once for once for each action, avoids excessive tapping”) and interest (e.g., “Child maintains attention on the screen”) on five one/zero items. Both skill and interest were high across the two sessions, with overall item means .90 for skill for both computer versions and .84 for interest for both computer versions (*SD*s .16 and .18 for version A and B computer skill ratings, respectively, and..24 and .23 for version A and B computer interest ratings, respectively). These scores confirm impressions that children understood the tasks and were engaged with them. Scores for all ratings across the two sessions were standardized and summed to be used in subsequent analyses as a covariate (internal consistency reliability expressed as mean interitem association = .33, $p < .001$).

Thus, anecdotal observation, supported by quantitative skill and interest ratings, showed that children were well engaged in the computerized tasks. Nonetheless, it was important to consider whether language spoken at home might hamper performance on the computerized measures for children who might be English language learners (ELLs). Approximately 61% of parents reported on home language; their reports of speaking English at home (n = 215) versus those reporting speaking Spanish (n = 37) or some other language (n = 22) did not differ on any individual rating of skill or interest for either version of the computerized AKT-S or CST, or the aggregate (Fs (2, 249-265) < 0.97, *ns*).

Procedure

We refer to in-person and computerized administrations of each measure as modes, and for each mode there were two versions of the measure, notated as version A and version B. Both modes are identical in all pertinent respects, including verbal and pictorial item content, training between recognition and situation portions of the AKT-S, presentation of stories, and randomized placement of response choices. At each administration, children were individually assessed in a quiet area in or just outside their classrooms, with two measures, either in-person or computerized modes. Research assistants introduced the computer tablet to each child individually, making sure that the instructions were understood, and allowed each child to complete the assessment on their own. Children had few if any difficulties working independently with the tablet, although a few had to be encouraged to finish. Most chose to wear child-sized earphones while completing the assessment. Administration time averaged approximately 10 to 15 minutes per session, including rapport building.

The order of administration was counterbalanced across modes and versions for the two measures, resulting in a total of eight combinations (see Appendix A). Within each of these eight orders, each child was administered the in-person versions (A and B) twice and the computerized version (A and B) twice. More specifically, children did the same version of both measures, either computerized or in-person, depending on what counterbalanced order was being followed in Administration 1 and 3, and the other version in Administration 2 and 4. Which mode came first in each pair of weeks was randomized. A streamlined version of this ordering is depicted in Figure 1.

After each of the computerized versions, administrators rated children on their interest in task and skill with the tablet computer. Teachers completed the two validity measures (i.e., the PLBS and SCBE-30) approximately two-three months after the child assessments.

Analytic Plan

For both CST and AKT-S computerized modes, cross-version correlations were used to approximate temporal stability, and mean interitem associations were calculated to demonstrate reliability (RQ1). Using MANOVAs, mean differences in the measures were examined in terms of mode (computerized and in-person) and version (A and B), with covariates of age, gender, and observer rating aggregate (RQ 2).

Next, concurrent validity associations of in-person and computerized modes were calculated (RQ 3). Finally, predictive validity associations using the Learning Behaviors/Attitudes and Social Competence aggregates were calculated for both versions of the computerized mode (Research Question 4). Because data from assessments are nested within classrooms, a hierarchical linear modeling (HLM) approach was taken for RQ 3 and 4 analyses. Level 1 models for RQ3 included the in-person mode predicting the corresponding computerized mode, for each scale and each version. Level 1 models for RQ4, in separate equations for each scale and each version, included computerized measures' scores predicting teacher rating aggregates. Level 2 for RQs 3 and 4 included classroom membership, to appropriately weight the mean and slope of child-level variables. Intraclass correlation coefficients (ICCs) were calculated for each child outcome and mode, revealing between 0 and 31% of the variance in child outcomes explained by their classroom, justifying a multi-level analytical approach. Unsurprisingly, teacher-reported outcomes (i.e., social competence and learning behaviors/attitudes) tended to have higher ICCs, .13–.31, than direct child assessments, .00–.15 (see Tables 3 to 5).

Effect sizes (Equation 1) and percentage variance explained (Equation 2) were calculated for HLM coefficients in RQ3 and RQ4 (Rimm-Kaufman, Curby, Grimm, Nathanson, & Brock, 2009)**.** As noted by these authors, calculation of each involves the following:

Eq. 1: ((predictor's unstandardized β) X its standard deviation))/(outcomes' standard deviation).

Percentage of variance explained at each level involved the following:

Eq. 2: ((total variance from unconditional model) - (residual variance from the final model))/(total variance from unconditional model).

Results

Research Question 1: Reliability

First, all associations between versions of both measures were generally significant, exemplifying short-term test-retest reliability (see Table 1). Over one- to three-week intervals, correlations for the AKT ranged from weak in magnitude for negative recognition to moderate for other subscales. For CST emotion choices, all correlations were moderate in magnitude, and for CST behavior choices these ranged from weak for avoidant/passive choices to moderate for the other three choices. Thus, over 80% (9 of 11) of cross-version reliability correlations were of moderate magnitude.

Second, regarding internal consistency, each scale has only three to six items (with only three or four possible responses). Thus, there were too few items in CST and AKT-S scales for traditional indices of internal consistency (e.g., Cronbach's α) to be meaningful (Briggs & Cheek, 1986; Spiliotopoulou, 2009); mean interitem association marked internal consistency here. Mean interitem association above .14 are considered acceptable (Clark & Watson, 1995); Briggs and Cheek recommended between .20 and .40 for best item-item homogeneity. Internal consistency by these criteria is generally acceptable across versions for each measure, except for version A avoidance/passive choices, with 72% of scales meeting Briggs and Cheek's more stringent benchmark.

Research Question 2: Concurrent Validity via Assessing Mean Differences

Mixed model MANOVAs were performed to examine differences across in-person/ computerized mode, and version (A or B) as within-subject factors, with age, gender, and observer rating aggregate as between-subject covariates (see Table 2 for adjusted means for each mode and version, for each measure). Findings for covariates are reported for each measure. Significant interactions did not occur at a frequency greater than expected by chance and will not be reported.

For the AKT-S, there were no main effects of mode or version for any aggregate. Age differences were found favoring older children on negative emotion recognition, stereotypical situation, and nonstereotypical situation knowledge (across mode and version, *Fs* (1, 362-370) = 16.98 – 39.14, partial $\eta^2 s$ = .04 -- .10 *ps* < .001). Further, children who obtained higher observer ratings on computer usage obtained higher scores on all indices of emotion knowledge (*Fs* (1, 362-370) = 47.15 – 54.27, partial $\eta^2 s$ = .11 – .13, *ps* < .001). These are generally medium to large effects. Girls performed better on nonstereotypical situation knowledge (*F* (1, 362) = 5.80, partial η^2 = .02, *p* < .01, a small effect).

For CST, there were no main effects of mode or version. Several age differences were found. Sad emotion choices were made more often by older children (*F* (1, 363) = 11.16, partial η^2 =.03, *p* < .01); conversely, happy emotion choices were made less often by older children (*F* (1, 363) = 28.03, partial η^2 =.07, *p* < .001). In terms of behavior choices, older children chose prosocial behavior responses more often (*F* (1, 363) = 26.80, partial η^2 =.07, *p* < .001). These are small to medium effects. Regarding gender, there were two differences: Boys more often chose avoidant/passive behaviors (*F* (1, 363) = 3.89, partial η^2 =.01, *p* < .05, a small effect). Girls were more likely to choose crying responses (*F* (1, 363) = 9.80, partial η^2 =.03, *p* < .05, a small effect). There were observer rating effects for happy and sad emotion choices (*Fs* (1, 363) = 13.01 and 6.82, partial η^2s =.04 and .02, respectively, *ps* < .001, small effects). Associations between this covariate and these scores demonstrated that children who were more skilled with and interested in the computer were likely to choose sad, not happy, emotions.

Research Question 3: Concurrent Validity via Between Mode Associations

Significant associations were found between in-person and computerized modes for all AKT-S and CST behavior choice scales, for both tests' versions (see Tables 3 and 4). More specifically, no pair of version associations for any scale differed significantly. We calculated effect sizes for the HLM coefficients with in-person versions' scores predicting computerized versions' scores and these ranged from medium (*d* = .20) negative recognition version B, to large for stereotypical version B (*d* = .42) (all remaining coefficients for version B, and all for version

A, also had large effect sizes). Finally, we computed percent of variance explained; this ranged from .09 to .23 for the aforementioned coefficients.

For the CST emotion choices, significant associations occurred for all emotions except “just ok”, and version differences were generally small. Effect sizes for significant associations ranged from small (*d* = .10) for angry version A, to large for happy version A (*d* = .42). Percent of variance explained ranged from .06 to .37 for the aforementioned coefficients.

As for CST behavior choices, version differences in associations were again relatively small. Effect sizes were all large, from version B of aggressive choices (*d* = .44) to passive/avoiding version A (*d* = .62). Percent of variance explained ranged from .41 to .60 for the aforementioned coefficients.

In general, then, inter-modal associations were evident for both the AKT-S and for behavior choices of the CST; most effect sizes were large. In contrast, effect sizes for CST emotion choices ranged from small to large.

Research Question 4: Predictive Validity Associations

Computerized AKT-S total scores showed significant associations with teacher aggregates for children’s SEC and learning behaviors/attitudes (see Table 5); version differences in associations sometimes occurred. Effect sizes for the social competence aggregate were medium to large, from stereotypical version B (*d* = .18) to for nonstereotypical version A (*d* = .34), with percent variance explained .14 for both aforementioned coefficients. Effect sizes for the learning behaviors/attitudes aggregate were medium, from negative recognition version A (*d* = .18) to nonstereotypical version B (*d* = .29), with percent variance explained ranging from .19 to .23 for the aforementioned coefficients.

CST scale scores showed fewer, but expected, associations with teacher reports of child SEC and learning behaviors/attitudes (see Table 5). Overall, for both emotion and behavior choices, significant validity associations occurred at a rate greater than expected by chance. Version differences in associations were generally small (see Table 5 for some exceptions, for happy and “just ok” emotion choices, and socially competent behavior choices). Specifically, happy emotion choices were negatively related, and “just ok” emotion choices were positively related to both the social competence and learning behavior/attitudes aggregate. The effect sizes for these significant associations with the social competence aggregate were medium to large, from “just ok” version A (*d* = .19) to happy version A (*d* = .41). Percentage variance explained for both these associations was .15. Effect sizes for the learning behaviors/attitudes aggregate were large, from happy choices version B (*d* = .38) to happy choices version A (*d* = .58), with percent variance explained ranging from .19 for both.

In terms of behavior choices on the CST, certain behavior choices were related, in expected directions, with social competence and learning behaviors/attitudes aggregates. For the social competence aggregate, effect sizes for significant associations were medium to large, from aggressive behavior choice version A to socially competent choices version A, respectively (*d* = .31 to *d* = .44). Concomitant percent variance explained were .16 to .17. In terms of learning behaviors/attitudes, effect sizes for significant associations were medium to large for socially competent choices, ranging from the version B (*d* = .31), version A (*d* = .52). Concomitant percent variance explained were .19 to 22.

Overall, then, the computerized AKT-S generally showed indices of validity regarding early school success. The CST’s relation with indices of validity were more complex; happy and “just ok” emotion choices, as well as socially competent behavior choices, were the main scales relating to teacher aggregates of social competence and learning behavior/attitudes; sad and angry emotion choices, as well as (generally) aggressive, crying, and passive/avoidant choices were unrelated.

Discussion

For early childhood educators to make informed decisions about how best to support children's social-emotional development, they must have access to rigorous formative assessments of children's competencies that can easily be administered in classroom settings. This study set out, therefore, to develop and validate computer-based parallel versions of two highly regarded measures of children's emotion knowledge (the AKT-S) and social problem solving (the CST). These measures are psychometrically valid and reliable when administered in person by trained researchers, but inappropriate for classroom use. In evaluating the performance of computerized versions of the AKT-S and CST, our study's goals were to specify: (a) reliability, (b) presence or absence of inter-modal mean differences and associations, and (c) validity in terms of relations with teacher ratings of social competence and learning behaviors. Overall, we found that children generally tended to perform similarly across modes (in-person and computerized) and versions of the assessments (allowing for formative re-testing). Further, computerized versions of both tools were associated with teachers' ratings of children's social-emotional competence and learning behaviors, although the pattern was stronger for the AKT than the CST. Some variations in performance are discussed below. Findings and limitations are considered considering the overarching goal of making SEC measurement more accessible and useful in the preschool classroom.

Research Question 1: Reliability

Specifically, reliability of the computerized measures was generally adequate. In introducing computerized measures of SEC for kindergartners to third graders, McKown and colleagues found cross-version/test-retest associations of similar magnitude, albeit across six months rather than several weeks (McKown, 2018; McKown, Russo-Ponsaran, Johnson, Russo, & Allen, 2016). Despite finding the measures reliable overall, two scales – AKT-S negative emotion recognition and CST avoidant/passive behavior choices – showed relatively weaker reliability, suggesting the need for further research to continue improving each scale. Assessing reliability over longer time periods could also strengthen the measures' psychometric profile.

Research Questions 2 and 3: Concurrent Validity via Assessing Mean Differences and Between Mode Associations

There were no mean differences between modes or versions for either the AKT-S or CST. Covariate analyses revealed one aspect of validity. Specifically, older children showed greater proficiency on the AKT-S, replicating earlier research (e.g., Authors, 2015). Age differences on the CST also echo earlier work (Authors, 2014a), with younger children more often choosing happy emotions and less often choosing sad emotions and prosocial behaviors. Further, inter-modal associations independent of classroom membership were almost unanimously statistically significant, with mostly medium to large effect sizes, except for angry and "just ok" CST emotion choices.

Critical consideration of the magnitude of inter-modal associations would be most useful given others' work. However, we found no research reports that have begun with successful in-person measures and compared them to new computerized versions (e.g., Parker, Mathis, & Kupersmidt, 2013, created a computerized emotion knowledge measure but did not begin with an in-person version with which to compare it). Those who have evaluated in-person and computerized assessments have focused on cognitive aspects of early childhood development, and found computerized assessment scales to be internally consistent and temporally stable, as well as comparable across the two modes in terms of mean differences (e.g., for phonological awareness, Carson, Gillon, & Boustead, 2011; see also Csapó, Molnár, & Nagy, 2014, who examined first graders' cognitive abilities and found internal consistency but a few mode differences). None has examined inter-modal associations.

What, then, contributes to inter-modal consistency? Assessment elements related to inter-modal similarity were isolated by Csapó et al. (2014); these included lack of change between the in-person versus the computerized version in item content, required type of response, and scoring. It may be that the general lack of mean difference for our measures, as well as their inter-modal associations, reside at least partially in the invariance of these criteria. That is, the in-person and computerized modes' items looked identical, their child response requirement was quite similar, and scoring was identical.

It could be argued that associations should thus be even stronger due to such similarity. However, given that observer ratings of computer use/skill were related to AKT-S scores and CST emotion choices, variability in children's independent tablet usage may somewhat attenuate inter-modal associations. We made subsidiary efforts to pinpoint how observers' ratings of children's computer ability might indeed moderate the intermodal associations, but these were uninformative. Perhaps future researchers could use more precise measures of children's computer ability to investigate this potential source of attenuation and continue to examine cross-modal associations.

Research Question 4: Predictive Validity Associations

Computerized assessments have been found predictively valid in other early childhood research as well. For example, Carson and colleagues garnered evidence for validity of a phonological awareness assessment via its association with reading difficulties (Carson, 2017; Carson et al., 2011; Carson, Boustead, & Gillon, 2014). McKown and colleagues (McKown, 2018; McKown et al., 2016) showed strong relations of both components of their SELWeb computerized measure's scales, both individually and aggregated, to three- to eight-year-olds' social skills and behavior problems, and indirectly to their math and reading skills.

The present study also highlights the predictive validity of our computerized assessments with well-documented indices of social competence and positive learning behaviors. In line with others' results, then, validity of the AKT-S demonstrated strong support in predicting teacher reports of learning behaviors/attitudes and social competence. The CST scales significantly related to teacher-rated social competence and learning behaviors/attitudes made sense – saying one would be happy when difficult situations occurred or that one would enact aggression were negatively related to teacher ratings. In contrast, teacher ratings were positively related with maintaining a "just OK" demeanor and performing socially competent behaviors. Nonetheless, given somewhat stronger validity evidence for the earlier in-person mode of the CST (which did not include difficult relational situations or two versions; Authors, 2013a, 2013b, 2014a), and some differences in the earlier measures' patterns of validity associations as compared to the current computerized mode, CST validity should be probed further and work could be done to improve it.

Overall, however the current results suggest promising preliminary findings, and show that the computerized mode has comparable validity with the in-person mode. These results coincide with a growing movement to create valid, digitized direct assessments to increase their usage in early childhood education settings.

Considerations to Bring Computerized SEC Assessment to Preschool Classrooms

If our goal is to move toward using these measures in the early childhood classroom, what other qualities may be important to consider? First, children must be able and interested in performing the assessments. Our observer ratings and children's own words (e.g., one boy said after his first assessment session, "That was FUN! I want to do it again!") suggest that this need has been met. Our findings also are supported by recent evidence of preschoolers' fluency with and enjoyment of digital media (Clarke & Abbot, 2016; Palaiologou, 2016).

We can also examine important qualities of computerized instruction for preschoolers (Ihmeideh, 2015). These include age appropriateness of item presentation via software and methods for using hardware. In our computerized measures, such considerations included use of a touchscreen rather than a mouse (e.g., use of a mouse can be difficult for preschoolers; Barnes, 2015). Other important qualities related to usability that were met in our computerized versions included allowing the child a measure of control in setting the pace; providing clear instructions, with simple, precise directions and visual prompts; and ability for the child to use the program independently. These qualities add to the potential usability of the measures in the classroom.

Second, teachers must be able to moderate/perform the assessments and find them useful in linking to instruction and programming (Denham, 2015; McKown, 2017). Reducing requisite training and time demands for teachers is also essential, especially because some of them may have concerns limiting their integration of technology in the classroom (e.g., Blackwell, Lauricella, & Wartella, 2014). Regarding ease of administration, Carson and colleagues (2011) noted success in their study of computerized assessment, as well as quicker assessment time (see also Tanyel & Knopf, 2011). Further, Stevenson, Touw, and Resing (2011) noted that not only were computerized measures quicker to administer than noncomputerized, but also cheaper. Johnson (2012), in an early review of our plans to create these measures, stated that over and above these strengths, data from computerized measurement must be accessible so teachers can use data in a timely fashion. Attempting to meet these needs, we will continue work on two fronts: (a) improvement of reliability and validity via examination of item content and modification, and (b) making improved measures available in an online platform, complete with automatic scoring for both individuals and classrooms.

Further, regarding how computerized assessment could be useful to instruction and programming, Tanyel and Knopf (2011) suggested that use of computerized assessment gave teachers more time in the classroom, increased their organization (via use of computerized results files), and even motivated them to deeper observation of and interaction with children. They also asserted that early childhood teachers find computerized measures useful in multi-tiered assessment that informs instruction aligned with SEC standards (see also Kamler, Moiduddin, & Malone, 2014; Reilly, 2007; Rosen, & Jaruszewicz, 2009). In short, progress monitoring through technology-based data collection can allow for timely change in teacher behaviors impacting student behaviors (Johnson, 2012). With respect to usefulness for instruction, we are adding activity suggestions for teachers to our measures' platform (see Supplemental File 3). Central to the entire platform, we are trying to conform to Johnson's assertion of that computerized measures need to be easily understood by early childhood educators.

All these assertions, however, require that teachers feel confidence and competence in using any digital medium (Kamler et al., 2014), and we would assert this caveat is particularly pertinent for tying assessment to instruction. Teachers will benefit from professional development focused on administration and scoring of these computerized measures to facilitate their use (Johnson, 2012; Kamler et al., 2014), especially given the paucity of training in educational computer use that they report here. Finally, teachers may need support in sharpening their focus on SEC. To that end, our research team is also creating training materials (e.g., modules on Introduction to Child SEL, Supporting Children's SEL in the Classroom, and Teachers' Own Emotional Competence), to be made available alongside the computerized assessments and any training subsequently created for their use.

Further, regarding child users, more attention may be needed into the potential role of ELL status on facility with these computerized measures, before more widespread dissemination. Although we did not measure language competency (specifically regarding English, or more generally), arguably it might be expected to impact child performance broadly and not specifically the focus of this work – psychometric adequacy of the computerized mode of the SEC measures. Nonetheless, it would advantageous to pinpoint its contributions more specifically.

In sum, findings across urban, suburban, and more rural settings strengthen our conclusions regarding our measures' potential generalized usefulness. Next steps include standardization/validation with a larger, even more demographically diverse sample, with attention given to language competency. Our focus would be to continue improvement of reliability and validity, especially for AKT-S recognition and angry, "just ok", and avoidant/passive choices on the CST. At the same time, work on improvement of the computer platform, along with automatic scoring, activity suggestions, and allied professional development modules needs to continue in parallel with measure improvement. Over and above this continued work, results here point to a very promising set of measures of SEC in young children, which are linked with early school success.

References

Authors (2012a, 2012b, 2012c; 2013a, 2013b, 2014a, 2014b, 2015).

Barnes, S. K. (2015). Computer-based testing and young children. In O. Saracho (Ed)., *Contemporary perspectives on research in assessment and evaluation in early childhood education.* (pp. 373-395). Charlotte, NC: Information Age Publishers.

Bierman, K. L., Nix, R. L., Heinrichs, B. S., Domitrovich, C. E., Gest, S. D., Welsh, J. A., & Gill, S. (2014). Effects of Head Start REDI on children's outcomes 1 year later in different kindergarten contexts. *Child Development*, *85*(1), 140-159.

Blackwell, C. K., Lauricella, A. R., & Wartella, E. (2014). Factors influencing digital technology use in early childhood education. *Computers & Education*, *77*, 82-90.

Bridgeland, J., Bruce, M., & Hariharan, A. (2013). *The missing piece: A national teacher survey on how social and emotional learning can empower children and transform schools. A Report for CASEL*. Civic Enterprises.

Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality*, *54*, 106-148. doi10.1111/j.1467-6494.1986.tb00391.x

Carson, K. L. (2017). Reliability and predictive validity of preschool web-based phonological awareness assessment for identifying school-aged reading difficulty. *Communication Disorders Quarterly*, *39*, 259-269. doi10.1177/1525740116686166

Carson, K., Boustead, T., & Gillon, G. (2014). Predicting reading outcomes in the classroom using a computer-based phonological awareness screening and monitoring assessment (Com-PASMA). *International Journal of Speech-Language Pathology*, *16*(6), 552-561.

Carson, K.L., Gillon, G., & Boustead, T. (2011). Computer-administrated versus paper-based assessment of school-entry phonological awareness ability. *Asia Pacific Journal of Speech, Language and Hearing*, *14*(2), 85-101. doi10.1179/136132811805334876

Chen, J. Q., & Chang, C. (2006). Using computers in early childhood classrooms: Teachers' attitudes, skills and practices. *Journal of Early Childhood Research*, *4*(2), 169-188. doi10.1177/1476718X06063535

Clark, L. A. & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*, 309–319. doi:10.1037/1040-3590.7.3.309

Clarke, L., & Abbott, L. (2016). Young pupils', their teacher's and classroom assistants' experiences of iPads in a Northern Ireland school: "Four and five years old, who would have thought they could do that?". *British Journal of Educational Technology*, *47*, 1051-1064. doi10.1111/bjet.12266

Copple, C., & Bredekamp, S. (2009). *Developmentally appropriate practice in early childhood programs serving children from birth through age 8*. Washington, DC: National Association for the Education of Young Children.

Couse, L. J., & Chen, D. W. (2010). A tablet computer for young children?: Exploring its viability for early childhood education. *Journal of Research on Technology in Education, 43*, 75-98

Csapó, B., Molnár, G., & Nagy, J. (2014). Computer-based assessment of school readiness and early reasoning. *Journal of Educational Psychology*, *106*, 639-650. doi10.1037/a003575

Denham, S. A. (2015). Assessment of SEL in educational contexts. In Durlak, J. A., Domitrovich, C. E., Weissberg, R. P., & Gullotta, T. P. (Eds). *Handbook of Social and Emotional Learning: Research and Practice* (pp. 285-300). New York: Guilford Press.

Denham, S. A., Bassett, H. H., & Zinsser, K. (2012). Computerizing social-emotional assessment for school readiness: First steps toward an assessment battery for early childhood settings, *Journal of Applied Research on Children: Informing Policy for Children at Risk*: *3(2)*, Article 3. Available at: http://digitalcommons.library.tmc.edu/childrenatrisk/vol3/iss2/3

Denham, S. A., & Brown, C. A., & Domitrovich, C. E. (2010). “Plays nice with others”: Social-emotional learning and academic success. *Early Education and Development, 21,* 652-680. doi10.1080/10409289.2010.497450

Denham, S. A., Ferrier, D.E., Howarth, G.Z., Herndon, K. J., & Bassett, H. H. (2019). Key considerations in assessing young children’s emotional competence. In N. Humphrey, A. Lendrum, M. Wigelsworth, & M. T. Greenberg (Eds)., *Social and emotional learning.* New York: Routledge Group.

Denham, S. A., Kalb, S. C., Way, E., Warren-Khot, H. K., Rhoades, B. L, & Bassett, H. H. (2013). Social and emotional information processing in preschoolers: Indicator of early school success? *Early Child Development and Care, 183,* 667-688.

Dusenbury, L. A., Newman, J. Z., Weissberg, R. P., Goren, P., Domitrovich, C. E., & Mart, A. K. (2015). The case for preschool through high school state learning standards for SEL. *Handbook of social and emotional learning: Research and practice,* 532-548.

Ellis, K., & Blashki, K. (2004). Toddler techies: A study of young children's interaction with computers. *Information Technology in Childhood Education Annual*, *16*, 77-96.

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*, 212-230. doi:10.1521/scpq.19.3.212.40276

Fletcher, E. N., Whitaker, R. C., Marino, A. J., & Anderson, S. E. (2014). Screen time at home and school among low-income children attending head start. *Child Indicators Research*, *7*(2), 421-436.

Ihmeideh, F. M. (2015). Assessment of children's digital courseware in light of developmentally appropriate courseware criteria. *British Journal of Educational Technology*, *46*, 649-663. doi10.1111/bjet.12163

Izard, C. E., Stark, K., Trentacosta, C., & Schultz, D. (2008). Beyond emotion regulation: Emotion utilization and adaptive functioning. *Child Development Perspectives. 2*(3), 156-163. doi: 10.1111/j.1750-8606.2008.00058.x

Johnson, U. (2012). A step in the right direction: Development of the Computerized Assessment for Preschool Social Emotional Learning (CAPSEL). *Journal of Applied Research on Children: Informing Policy for Children at Risk*, *3*(2), 12. Available at: http://digitalcommons.library.tmc.edu/childrenatrisk/vol3/iss2/12

Kendziora, K., Weissberg, R. P., Ji, P., & Dusenbury, L. A. (2011). *Strategies for social and emotional learning: Preschool and elementary xx student learning standards and assessment*. Newton, MA: National Center for Mental Health Promotion and Youth Violence Prevention, Education Development Center, Inc.

Kılıç, Ş., & Aytar, A. G. (2016). Emotion understanding of Turkish preschoolers and maternal emotional socialization. *Journal of Human Sciences*, *13*(1), 2102-2112.

LaFreniere, P. J., & Dumas, J. E. (1996). Social competence and behavior evaluation in children ages 3 to 6 years: The short form (SCBE-30). *Psychological Assessment*, *8*, 369-377. doi10.1037/1040-3590.8.4.369

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,* 201-219. doi: 10.1207/s15566935eed1302_6

Leyva, D., Berrocal, M., & Nolivos, V. (2014). Spanish-speaking parent–child emotional narratives and children's social skills. *Journal of Cognition and Development*, *15*(1), 22-42.

McCarrick, K., & Li, X. (2007). Buried treasure: The impact of computer use on young children’s social, cognitive, language development and motivation, *AACE Journal, 15*, 73-95.

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. doi: 10.1002/pits.10036

McKown, C. (2017). Social and emotional learning: A policy vision for the future. *The Future of Children. Princeton, New Jersey: Princeton University. Retrieved from www. futureofchildren. org*.

McKown, C. (2018). Reliability, factor structure, and measurement invariance of a web-based assessment of children’s social-emotional comprehension. *Journal of Psychoeducational Assessment,* 0734282917749682.

McKown, C., Russo-Ponsaran, N. M., Johnson, J. K., Russo, J., & Allen, A. (2016). Web-based assessment of children’s social-emotional comprehension. *Journal of Psychoeducational Assessment*, *34*(4), 322-338.

Morgan, J. K., Izard, C. E., & King, K. A. (2010). Construct validity of the emotion matching task: Preliminary evidence for convergent and criterion validity of a new emotion knowledge measure for young children. *Social Development*, *19*(1), 52-70.

Murray, E., & Harrison, L. J. (2011). The influence of being ready to learn on children's early school literacy and numeracy achievement. *Educational Psychology*, *31*, 529-545. doi10.1080/01443410.2011.573771

Nix, R. L., Bierman, K. L., Domitrovich, C. E., & Gill, S. (2013). Promoting children's social-emotional skills in preschool can enhance academic and behavioral functioning in kindergarten: Findings from Head Start REDI. *Early Education & Development*, *24*(7), 1000-1019.

Oberle, E., Domitrovich, C. E., Meyers, D. C., & Weissberg, R. P. (2016). Establishing systemic social and emotional learning approaches in schools: A framework for schoolwide implementation. *Cambridge Journal of Education*, *46*(3), 277-297.

O'Connor, R., DeFeyter, J., Carr, A., Luo, J. L., & Romm, H. (2017). *A review of literature on social and emotional learning for students ages 3-8: Implementation strategies and state and district support policies (part 2 of 4).* U.S. Department of Education Institute for Education Sciences National Center for Education Evaluation and Regional Assistance: Regional Educational Laboratory (Mid-Atlantic) at ICF International.

Orobio de Castro, Merk, Koops, Veerman, & Bosch, 2005. Emotions in social information processing and their relations with reactive and proactive aggression in referred aggressive boys. *Journal of Clinical Child and Adolescent Psychology*, *34*(1), 105-116.

Palaiologou, I. (2016). Children under five and digital technologies: implications for early years pedagogy. *European Early Childhood Education Research Journal*, *24*,5-24. doi10.1080/1350293X.2014.929876

Parker, A. E., Mathis, E. T., & Kupersmidt, J. B. (2013). How is this child feeling? Preschool-aged children's ability to recognize emotion in faces and body poses. *Early Education & Development*, *24*(2), 188-211.

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 behavior in children and youth. *Journal of School Health*, *70*(5), 179-185.

Piotrkowski, C. S., Botsko, M., & Matthews, E. (2000). Parents' and teachers' beliefs about children's school readiness in a high-need community. *Early Childhood Research Quarterly*, *15*, 537-558. doi10.1016/S0885-2006(01)00072-2

Rebelo, A. S. G., Verissimo, M., Machado, P. M., & Silva, F. (2013). A segurança dos modelos internos e o conhecimento emocional nas crianças de idade pré-escolar= Internal models and emotional knowledge in preschool children. *Psicologia: reflexão e crítica*, 591-598.

Reilly, M. A. (2007). Choice of action: Using data to make instructional decisions in kindergarten. *The Reading Teacher*, *60*(8), 770-776. doi10.1598/RT.60.8.7

Rimm-Kaufman, S. E., Curby, T. W., Grimm, K. J., Nathanson, L., & Brock, L. L. (2009). The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. *Developmental Psychology*, *45*, 958-972.

Romano, E., Babchishin, L., Pagani, L. S., & Kohen, D. (2010). School readiness and later achievement: Replication and extension using a nationwide Canadian survey. *Developmental Psychology, 46*, 995-1007. doi: 10.1037/a0018880

Rosen, D. B., & Jaruszewicz, C. (2009). Developmentally appropriate technology use and early childhood teacher education. *Journal of Early Childhood Teacher Education*, *30*, 162-171. doi10.1080/10901020902886511

Schultz, D., Ambike, A., Logie, K. S., Bohner, K. E., Stapleton, L. M., VanderWalde, H., ...Betkowski, J. A. (2010). Assessment of social information processing in early childhood: Development and initial validation of the Schultz Test of Emotion Processing—Preliminary Version. *Journal of Abnormal Child Psychology, 38*, 601–613. doi 10.1007/s10802-010-9390-5

Sette, S., Bassett, H. H., Baumgartner, E., & Denham, S. A. (2015). Structure and validity of Affect Knowledge Test (AKT) in a sample of Italian preschoolers. *The Journal of Genetic psychology*, *176*(5), 330-347.

Sheard, M. K., Ross, S., & Cheung, A. (2013). Social-emotional learning championing freedom, education and development: A vehicle for at-risk students to succeed. *Cypriot Journal of Educational Sciences*, 1-18.

Spiliotopoulou, G. (2009). Cronbach's alpha and paediatric assessment in occupational therapy. *Australian Occupational Therapy Journal 56*, 150–155 doi: 10.1111/j.1440-1630.2009.00785.x

Stevenson, C. E., Touw, K. W., & Resing, W. C. (2011). Computer or paper analogy puzzles: Does assessment mode influence young children's strategy progression? *Educational and Child Psychology*, *28*(2), 67-84.

Tanyel, N., & Knopf, H. T. (2011). Does using digital media in assessment affect teacher practices in infant and toddler classrooms? *International Journal of Early Years Education*, *19*(3-4), 297-311. doi10.1080/09669760.2011.644923

Upshur, C. C., Heyman, M., & Wenz-Gross, M. (2017). Efficacy trial of the Second Step Early Learning (SSEL) curriculum: preliminary outcomes. *Journal of Applied Developmental Psychology*, *50*, 15-25.

Vandewater, E. A., Rideout, V. J., Wartella, E. A., Huang, X., Lee, J. H., Shim, M. (2007). Digital childhood: Electronic media and technology use among infants, toddlers, and preschoolers. *Pediatrics, 119*, 1006-1015.

Zinsser, K.M., Gordon, R., & Jiang, X. (under review) Lessons from home: A systematic review of emotion-focused parenting practices to inform preschool teaching practices. Manuscript submitted for review.

Table 1

Reliability for Computerized AKT and CST Scales

Measure/Scale	Test-Retest (Version/Version) Associations	Internal Consistency (Mean Interitem Association)	
		Version A	Version B
AKT			
Negative Recognition 3 items	.24	.15	.18
Stereotypical Situations[a] 3 items	.55	.40	.36
Nonstereotypical Situations[b] 6 items	.48	.33	.35
CST (each scale has 6 items)			
Emotion Choices			
Happy Choices	.55	.27	.32
Sad Choices	.47	.21	.25
Angry Choices	.45	.15	.22
"Just OK" Choices	.57	.24	.29
Behavior Choices			
Socially Competent Choices	.49	.19	.25
Aggressive Choices	.52	.24	.28
Crying Choices	.48	.21	.21
Avoidant/Passive Choices	.36	.12	.14

*Note. df*s=399 to 422. All version/version associations significant, p <.001. [a] Stereotypical=all people feel the same way. [b] Nonstereotypical =people may differ in their emotional reactions to these situations.

Table 2

Descriptive Data for AKT and CST

Measure/Scale	Version A		Version B	
	In-person	Computerized	In-person	Computerized
AKT				
Negative Recognition	1.88 (0.27)	1.74 (0.35)	1.89 (0.25)	1.76 (0.35)
Stereotypical Situations	1.81 (0.33)	1.72 (0.43)	1.80 (0.37)	1.69 (0.45)
Nonstereotypical Situations	1.78 (0.36)	1.69 (0.38)	1.76 (0.36)	1.65 (0.44)
CST				
Emotion Choices				
Happy Choices	0.17 (0.23)	0.17 (0.23)	0.19 (0.25)	0.19 (0.25)
Sad Choices	0.35 (0.31)	0.30 (0.27)	0.36 (0.31)	0.27 (0.27)
Angry Choices	0.28 (0.29)	0.22 (0.23)	0.28 (0.27)	0.24 (0.25)
“Just OK” Choices	0.20 (0.23)	0.30 (0.29)	0.18 (0.23)	0.30 (0.29)
Behavior Choices				
Socially Competent Choices	0.40 (0.29)	0.35 (0.27)	0.37 (0.27)	0.39 (0.29)
Aggressive Choices	0.19 (0.27)	0.21 (0.25)	0.23 (0.27)	0.23 (0.27)
Crying Choices	0.09 (0.17)	0.12 (0.19)	0.12 (0.19)	0.12 (0.19)
Passive Choices	0.32 (0.27)	0.32 (0.25)	0.28 (0.23)	0.26 (0.23)

Note. All means are per item of the measure, adjusted for age, gender, and observer ratings of children’s computer use competence and interest in the assessment process. AKT-S item scores vary from 0 to 2, and CST scores vary from 0 to 1. Standard deviations are in parentheses. *N*s = 408-436.

Table 3

HLM Coefficients Depicting Relations Among Computerized and In-Person AKT-S Scales

Computerized AKT-S	Version/ICC	Negative Recognition	Stereotypical Total	Non-Stereotypical Total
Negative Recognition	A/.02	0.268 *** (0.08)		
	B/.06	0.252 *** (0.08)		
Stereotypical	A/.08*		0.500 *** (0.07)	
	B/.10**		0.514 *** (0.07)	
Non-Stereotypical	A/.09*			0.482 *** (0.05)
	B/.09**			0.466 *** (0.07)

Note. Total *N*s = 450 children in 106 classes. *df*s = 99 to 101 for intercept, 301 to 318 for predictor. [a] Adjusted means with age, gender, and observer rating of computer interest and skill accounted for. $^{***}p < 0.001$.

Table 4 *HLM Coefficients Relations Computerized and In-person CST Scales*

Computerized CST	In-person CST										
	Version/ ICC	Emotion Choices: Happy	Sad	Angry	Just OK	Version /ICC		Behavior Choices: Socially Competent	Aggressive	Crying	Avoidant/ Passive
Emotion Choices							**Behavior Choices**				
Happy	A/.00	0.591*** (0.06)				A/.15***	Socially Competent	0.469*** (0.04)			
	B/.14***	0.556*** (0.05)				B/.09*		0.585*** (0.05)			
Sad	A/.04^{+}		0.157* (0.07)			A/.04^{+}	Aggressive		0.472*** (0.05)		
	B/.06**		0.161** (0.06)			B/.08*			0.598*** (0.04)		
Angry	A/.04			0.079* (0.06)		A/.04**	Crying			0.563*** (0.06)	
	B/.06^{+}			0.158** (0.05)		B/.02				0.511*** (0.06)	
Just OK	A/.10**				0.108 (0.10)	A/.00	Avoidant/ Passive				0.406*** (0.05)
	B/.11**				-0.012 (0.08)	B/.00					0.444*** (0.04)

Note. Total *N*s = 450 children in 106 classes. *df*s = 101 to 103 for intercept, 298 to 308 for predictor. $^{***}p < 0.001$.

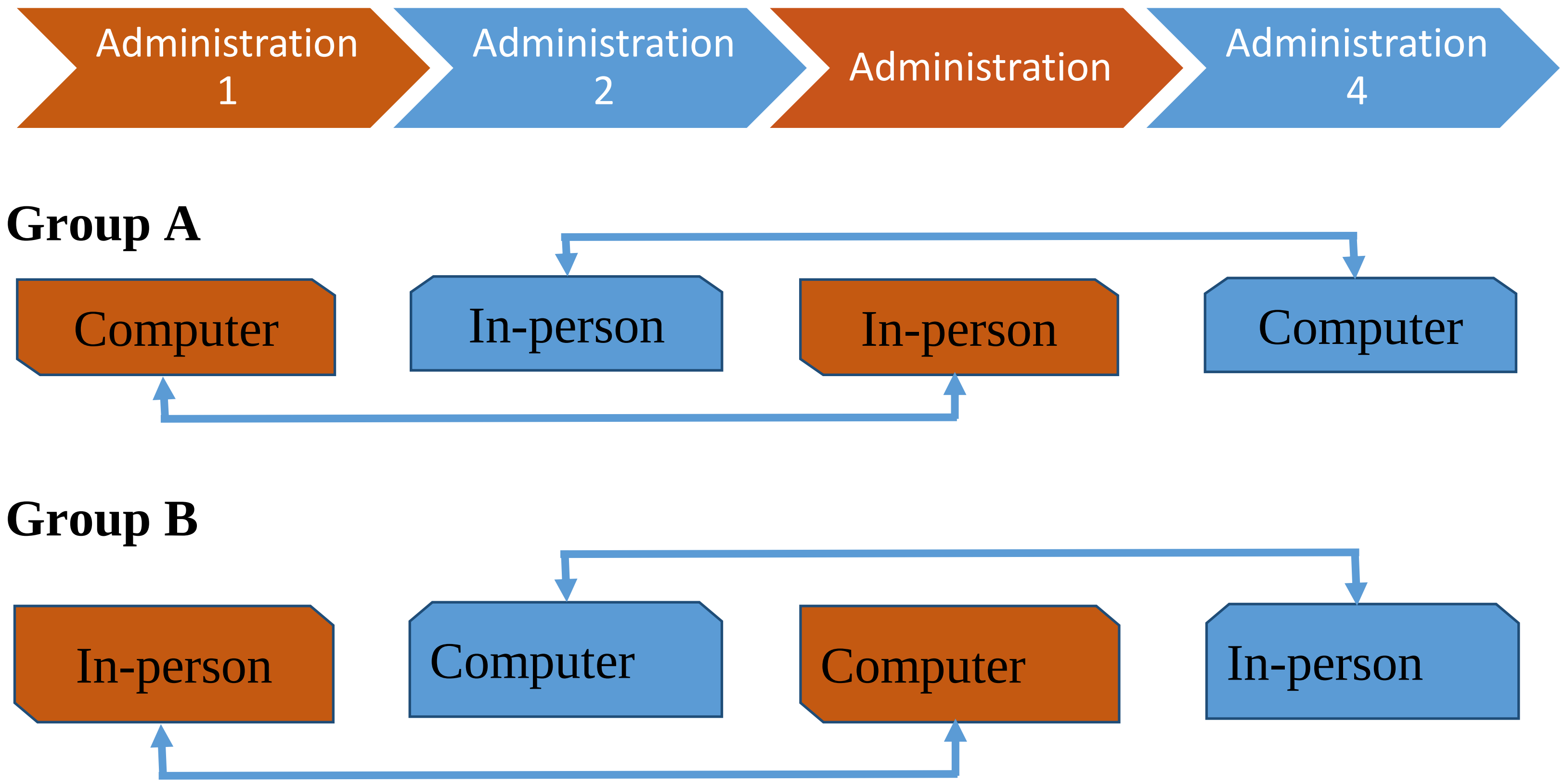
Table 5

HLM Coefficients for Validity Analyses

	Social Competence		Learning Behaviors/Attitudes	
Computerized Measure/Version for Each	Version A	Version B	Version A	Version B
AKT				
Negative Recognition	0.337** (0.11)	0.320** (0.12)	0.214+ (0.13)	0.372* (0.16)
Sterotypical Situations	0.274** (0.10)	0.223** (0.09)	0.326** (0.11)	0.246* (0.12)
Nonstereotypical Situations	0.187 (0.13)	0.355*** (0.10)	0.253* (0.12)	0.342** (0.11)
CST				
Emotion Choices				
Happy Emotion Choice	-0.407* (0.18)	-0.224 (0.17)	-0.507* (0.22)	-0.377* (0.18)
Sad Emotion Choice	0.078 (0.13)	-0.165 (0.17)	-0.041 (0.17)	0.023 (0.17)
Angry Emotion Choice	-0.014 (0.17)	0.061 (0.17)	0.030 (0.19)	0.193 (0.18)
"Just OK" Emotion Choice	0.225* (0.11)	0.254** (0.10)	0.389** (0.14)	0.107 (0.13)
Behavior Choices				
Socially Competent Behavior Choice	0.408** (0.15)	0.165 (0.12)	0.484** (0.16)	0.320* (0.17)
Aggressive Behavior Choice	-0.294+ (0.17)	-0.211(0.17)	-0.302 (0.24)	-0.270 (0.18)
Crying Behavior Choice	-0.111 (0.22)	0.126 (0.19)	-0.440+ (0.24)	-0.107 (0.23)
Avoidant/Passive Behavior Choice	-0.108 (0.18)	-0.094 (0.18)	0.440 (0.21)	-0.093 (0.18)

Notes. + $p \leq .07$. * p < .05, ** p < .01, *** p < .001. Predictors (computerized measures) grand mean centered for analysis. Standard errors in parentheses. *Dfs* = 86 for intercept B_0, 254 for B_1. *ICCs* for social competence and learning behaviors/attitudes equal .13 and .31, respectively, *p*s < .001.

Figure 1
Schematic Representation of Administration Procedures



Simplified diagram of the procedure. Children were randomly assigned either group A or B. The two boxes connected by the arrow were the same version of the measure in a different mode of assessment.

Appendix 1

Order of Assessments

Order	Administration 1	Administration 2	Administration 3	Administration 4
1	1. AKT-S(A) – Computer 2. CST(A) – Computer	1. CST(B) – in-person 2. AKT-S(B) – in-person	1. AKT-S(A) – in-person 2. CST(A) – in-person	1. CST(B) – Computer 2. AKT-S(B) – Computer
2	1. AKT-S(B) – in-person 2. CST(B) – in-person	1. CST(A) – Computer 2. AKT-S(A) – Computer	1. AKT-S(B) – Computer 2. CST(B) – Computer	1. CST(A) – in-person 2. AKT-S(A) – in-person
3	1. AKT-S(B) – Computer 2. CST(A) – Computer	1. CST(B) – in-person 2. AKT-S(A) – in-person	1. AKT-S(B) – in-person 2. CST(A) – in-person	1. CST(B) – Computer 2. AKT-S(A) – Computer
4	1. AKT-S(A) – in-person 2. CST(B) – in-person	1. CST(A) – Computer 2. AKT-S(B) – Computer	1. AKT-S(A) – Computer 2. CST(B) – Computer	1. CST(A) – in-person 2. AKT-S(B) – in-person
5	1. CST(B) – Computer 2. AKT-S(B) – Computer	1. AKT-S(A) – in-person 2. CST(A) – in-person	1. CST(B) – in-person 2. AKT-S(B) – in-person	1. AKT-S(A) – Computer 2. CST(A) – Computer
6	1. CST(A) – in-person 2. AKT-S(A) – in-person	1. AKT-S(B) – Computer 2. CST(B) – Computer	1. CST(A) – Computer 2. AKT-S(A) – Computer	1. AKT-S(B) – in-person 2. CST(B) – in-person
7	1. CST(B) – Computer 2. AKT-S(A) – Computer	1. AKT-S(B) – in-person 2. CST(A) – in-person	1. CST(B) – in-person 2. AKT-S(A) – in-person	1. AKT-S(B) – Computer 2. CST(A) – Computer
8	1. CST(A) – in-person 2. AKT-S(B) – in-person	1. AKT-S(A) – Computer 2. CST(B) – Computer	1. CST(A) – Computer 2. AKT-S(B) – Computer	1. AKT-S(A) – in-person 2. CST(B) – in-person

Note: Computer and in-person versions assessments for all measures were performed within a median of 14 days; 89-91% were completed within 1 month.

Electronic supplementary material
Click here to download Electronic supplementary material: CST appendix with pictures.docx