Received: 28 November 2019 | Accepted: 27 January 2021

DOI: 10.1111/sode.12512

METHODS ARTICLE

WILEY

Psychometric properties of Virtual Environment for Social Information Processing, a social information processing simulation assessment for children

Nicole M. Russo-Ponsaran | **Clark McKown** | **Ashley Karls** | **Irene Y. H. Wu**

Department of Psychiatry & Behavioral Sciences, Rush NeuroBehavioral Center, Rush University Medical Center, Chicago, IL, USA

Correspondence
Nicole Russo-Ponsaran, Department of Psychiatry & Behavioral Sciences, Rush NeuroBehavioral Center, Rush University Medical Center, 4711 Golf Road, Suite 1100, Skokie, IL 60076, USA.
Email: nicole_russo@rush.edu

Funding information
Institute of Education Sciences; National Institute on Disability and Rehabilitation Research; Rush Translational Sciences Consortium

Abstract

Virtual Environment for Social Information Processing (VESIP™) is a web-based social information processing assessment developed for youth in grades 3–7. VESIP was developed to address: (a) the broader assessment of social information processing in a variety of socially challenging situations and (b) the need for technically strong and scalable assessments that can be administered universally in schools as a growing number of states adopt social and emotional learning standards. Consistent with the Crick and Dodge theoretical model of social information processing, VESIP assesses six different dimensions: solution preference, problem identification, emotion response, intent attribution, goal preference, and social self-efficacy. This study summarizes technical properties of VESIP based on the evaluation of two general education samples: a multi-state group of students whose data were part of a large-scale norming study (N = 2,156), and a subset of local students from that group who participated in a validation study (n = 334). Confirmatory factor analyses supported a model that has three distinct facets: (a) one that includes five social information processing factors, (b) one that includes five situational factors, and (c) one overall social information

processing factor. This model closely parallels the Crick and Dodge model and suggests that social information processing is somewhat situation specific. Internal consistency and test–retest reliabilities for social information processing factors were good. VESIP scores were consistently associated with an alternate measure of social information processing and other criterion measures. Implications for theory and practice are considered.

KEYWORDS
assessment, education, peer relationships, problem solving, social cognition, social information processing

1 | INTRODUCTION

1.1 | Social information processing

Everyday life is filled with opportunities for social success and the risk of social failure. Daily interactions at school, at play, and at home require a range of key interpersonal skills, including social information processing (SIP) skills. SIP skills, defined as social-cognitive processes that allow us to navigate challenging social situations are critical for successful daily interactions. As described in a theoretical model established by Crick and Dodge (1994, 1996), SIP skills are made of dimensions including cue encoding, cue interpretation, goal generation, solution generation, evaluation of the merits of those solutions, and choosing and enacting a solution. Cue encoding involves the process of recognizing, through encoding verbal and nonverbal cues, that a social problem exists. The model includes both internal (e.g., feelings, empathy) and external encoding (e.g., what a peer did) in cue encoding. Cue interpretation involves developing a theory about the cause of a problem and determining the intent of those involved. Goal generation involves defining a desired outcome. Solution generation involves identifying a set of solutions that can be enacted in response to a social problem in order to achieve an important social goal. Choosing an effective solution requires consideration of the pros and cons of each possible solution, the likelihood of being able to enact a response, and the anticipated results of that response. Crick and Dodge considered the actual enactment piece separate from the cognitive decision-making process. Taken together, these dimensions are inter-related and do not necessarily occur in a forward sequence. At any point in the process, past experiences and emotional responses can influence decision-making (Lemerise & Arsenio, 2000). After an encounter, the sequence of events and results may be evaluated to determine whether the chosen response produced the desired social outcome. Crick and Dodge also considered this step auxiliary to the central model.

1.1.1 | Importance of SIP

Effective SIP enables people to successfully interact and form relationships with their peers, resulting in greater social support and acceptance (Bauminger et al., 2005; Crick & Dodge, 1994; McKown, 2007; McKown et al., 2009; Weissberg et al., 1997). Social support and acceptance are, in turn, partially causally related to positive behavioral and mental health outcomes. Additionally, SIP dimensions are reported to be associated with functional academic and general life outcomes (Crick & Dodge, 1994, 1996; Duncan et al., 2007; Liberman et al., 1986; McFall, 1982).

There is a long-established history of research linking SIP skills and particular diagnoses. While not an exhaustive review, many studies show that children with disruptive behavior disorders are more likely to attribute hostile intent to ambiguous situations and overvalue aggressive solutions to those problems (e.g., de Castro et al., 2005; Lansford et al., 2010; Matthys et al., 1999; Verhoef et al., 2019). Individuals with anxiety or depression tend interpret benign situations as hostile, favor avoidant solutions, and demonstrate less self-confidence in problem-solving abilities (e.g., Dickson & MacLeod, 2004; Luebbe et al., 2010). Youth with autism spectrum disorder may struggle with identifying social problems and generating and selecting appropriate solutions (e.g., Channon et al., 2014; Russo-Ponsaran et al., 2018, 2019; Shulman et al., 2012). Although not a clinical diagnosis, victimized individuals also show distinct SIP patterns including, for example, more hostile intent attributions, greater encoding of negative cues, and less relationship-oriented goals (Garner & Lemerise, 2007; van Reemst et al., 2016).

Certainly, SIP skills are critical thinking skills which all youth need to develop. The ability to understand and successfully navigate one's social environment facilitates peer acceptance and integration into a community, which has a significant effect on a person's success and quality of life (Baron-Cohen, 1989; Denham, 2006; Dubow & Tisak, 1989; Dubow et al., 1991). At stake are not just the material gains one receives from being part of a community, but also the emotional and psychological benefits (Parker & Asher, 1987). Few large-scale studies exist that evaluate the ability to process social information in non-clinical samples and across a variety of social situations.

1.1.2 | Relevance to education

Despite increasing awareness of the importance of SIP for all youth, including pre-school and school-aged students in general education programs (Cooke, 2017; Ziv, 2013), educators lack access to scalable assessments feasible for universal administration. In response to an increasing number of states adopting social and emotional learning standards, many school districts nationwide are implementing related learning curricula. For example, SIP skills are addressed in curricula like Second Step (Committee for Children, 2008), I Can Problem Solve (Shure, 1992), and the Social Decision Making/Problem Solving Program (Elias & Butler, 2005a, 2005b). Even so, educators are not well-prepared to measure SIP ability or progress. Acknowledging the importance and relevance of SIP skills for youth, we endeavored to develop a scalable, user-friendly, feasible, and technically sound method to evaluate children's SIP skills.

1.1.3 | Open questions about SIP

SIP skills are often measured in the context of one or two hypothetical situation types—ambiguous provocation and peer entry—and are limited to negative situational contexts (see review in Verhoef et al., 2019) rather than benign, complex ones. Because hypothetical situations in SIP assessments are typically homogenous in nature, it is difficult to ascertain the extent to which SIP skills operate in the same way across the varied types of challenging situations children routinely confront. Is SIP the same when a child is being bullied and when they are trying to join an ongoing activity? What about when they are choosing what game to play with a friend or when they are responding to a parent's request to do a chore? Presumably the same cognitive processes are used across these situations, but less understood is whether these processes are applied in the same way and whether children's skill levels differ by situation type.

One prior study of which we are aware has examined this question. In a sample of 387 early-elementary aged children, Dodge and colleagues (2002) examined the factor structure of a SIP assessment that included several vignettes focused on ambiguous provocation and others focused on peer entry. They found that a model including only SIP factors (intent attribution, goal generation, response generation, and response evaluation) fit the data

moderately well, and that model fit improved substantially when they added factors reflecting situation type. These findings suggest that the SIP skills may be applied differently across different situation types.

1.2 | Measurement of SIP

While there are quality SIP assessments available for use, surprisingly few SIP assessments assess the broad range of SIP dimensions as described by Crick and Dodge; fewer assess a broad range of challenging social scenarios; and even fewer engender a combination of high usability, technical soundness, and feasibility for universal school-based administration. Although not meant as an exhaustive list, Table 1 highlights some of the more commonly used measures. Typically, the most widely used tools for measuring SIP skills in schools are teacher rating scales. While these scales are useful for measuring observable behavior (e.g., externalizing behaviors), SIP skills are not easily observed. For a rater to score children's thinking skills (e.g., how skillfully they recognize the onset of a social problem), a high level of inference is required. As a result, rating scale measures of SIP may be vulnerable to inaccuracy. Most rating scales are generalized and provide only coverage of broad social skills rather than specific SIP dimensions. Self-report questionnaires can be used to assess SIP, but they require respondents to be able to accurately rate their own skills and willing to honestly report them. As a result, self-report of SIP is also vulnerable to inaccuracy.

With direct assessment of SIP, children demonstrate their skills by answering questions about how they think about and would respond to hypothetical challenging social situations. While there is still an appreciable gap in available direct assessments (Denham et al., 2010; Halle & Darling-Churchill, 2016; Hamilton et al., 2018; Weissberg et al., 2013), an increasing number of tools is becoming available. However, many are intended for younger populations (e.g., Challenging Situations Task; Denham et al., 1994; Southampton Test of Empathy for

TABLE 1 Review of common social information processing assessments

Measure	Age/Grade level	Format	Based on Crick and Dodge?	Citation
Test of Problem Solving (TOPS 2/3)	6–17 years	Teacher rating scale	No	Bowers et al. (2005, 2007)
Taxonomy of Problematic Social Situations for Children (Short/Long forms)	7–10 years	Teacher rating scale	Yes	Dodge et al. (1985), Matthys et al. (2001)
Social Problem-Solving Inventory Revised	13 years and up	Self-report	No	D'Zurilla et al. (2002)
Home Interview with Child	Grades K–3	Structured interview	No	Conduct Problems Prevention Research Group (1991)
Children's Evaluation of Everyday Situations (ChEESE-Q)	Grades 3–6	Structured questionnaire	Yes	Bell et al. (2009)
Challenging Situations Task	3–6 years	Vignette-based interview	No	Denham et al. (1994)
Schultz Test of Emotion Processing (STEP)	Grades PK–5	Computerized video vignettes	Yes	Schultz et al. (2010)
Social Information Processing Application (SIP-AP)	8–12 years	Web-based, self-administered	Yes	Kupersmidt et al. (2011)

Preschoolers; Howe et al., 2008) or require a one-on-one administration format (e.g., Test of Problem-Solving Elementary; Bowers et al., 2005). Such assessments require special training and skilled administration or scoring. Other existing web-based direct assessments (e.g., SELweb EE; McKown et al., 2016) offer improvements over one-on-one assessments because they are scalable, but often utilize less ecologically valid still images and have less content coverage of SIP dimensions and scenarios. Video-based direct assessments attempt to capture children's social thinking more closely, but most of these were developed to assess SIP specifically with respect to externalizing behavior and aggression (e.g., Video-SEIP; Coccaro et al., 2017; Social Information Processing Application [SIP-AP]; Kupersmidt et al., 2011), making generalization and broader application unclear.

Few assessments exist that are ideal for broad use in school systems through automation and allowance for group administration, specifically target a range of SIP skills based on the Crick and Dodge theoretical model, and exhibit technical soundness. While existing rating scales, self-report questionnaires, and digitally administered direct assessments have some of the desirable qualities of usability and technical soundness, none have all and, while psychometric data are available on many of these measures, sample size is often limited. A combination of these characteristics would reflect a widely relevant assessment that is suitable for universal administration.

1.3 | Virtual Environment for Social Information Processing™

1.3.1 | Our working model

Virtual Environment for Social Information Processing, or VESIP, is a web-based SIP assessment that uses an avatar and dynamica (animated) scenarios to create a customizable virtual environment through which children navigate challenging social situations. The goal was to create a SIP assessment with broad coverage of key SIP dimensions as identified by the Crick and Dodge SIP model. As such, the assessment was designed to measure SIP dimensions of: (a) problem identification, reflecting the identification of the relational problem at hand; (b) emotion encoding, reflecting internal cue encoding; (c) intent attribution to determine the degree of hostile intent; (d) goal preference for a desired situational outcome; (e) solution preference, reflecting both Crick and Dodge dimensions of response construction and decision; and (f) social self-efficacy. VESIP is brief (30–35 min) and presents 10 animated challenging situations across five categories (a-e below) that are particularly salient for late elementary and middle school students (e.g., Crick & Dodge, 1994; Erdley & Asher, 1996; Renshaw & Asher, 1983). Categories include: (a) ambiguous provocation, (b) peer entry into a group, (c) friendship initiation, (d) compromise, and (e) response to bullying (online Appendix). Each category is presented within two different school settings. Consistent with this model, we present confirmatory factor analyses with latent variable that correspond to these SIP dimensions and situation types in Results.

VESIP was intended to be ecologically valid and engaging such that evoked responses were more natural and indicative of what happens in real life. For example, VESIP allows for personalization to increase engagement, which, subsequently, is believed to increase validity (Stapleton, 2004; Verhoef et al., 2019). The child's avatar incorporates designated demographic features (e.g., gender, ethnicity) which are rarely addressed in existing assessments and allows for customization of preferred features (e.g., hairstyle, wardrobe) to increase the child's ability to relate to their avatar. To further increase engagement, the child customizes several scenes by selecting a preferred item that gets incorporated into the scenario.

While certainly working memory (e.g., van Nieuwenhuijzen & Vriens, 2012) and language abilities (e.g., Bauminger-Zviely et al., 2019; McKown et al., 2013) support SIP, to reduce cognitive demands, users do not need to rely solely on imagination, working memory, or language skills to access the interactions. Animated and narrated scenarios are played out twice and response options can be repeated through a hovering mechanism. Assessment questions are presented in the form of an exchange with a peer character matched in age, gender, and ethnicity to the user, emphasizing a first-person perspective for the question and answer portions of the assessment (Kupersmidt et al., 2011). As a result, children remain immersed in the experience throughout the assessment

rather than responding as an observer to questions about hypothetical situations as they would in a typical SIP interview. In this way, the narrative flow is better preserved, and the concept of an adult interrogator is removed to encourage a more honest and candid response.

VESIP was designed to be accessible, scalable, and suitable for universal administration in schools. It may be completed on any non-tablet computer device with an internet connection and a web browser, making it highly usable and suitable for mass administration. VESIP has already shown promise for use by children with and without autism spectrum disorder (Russo-Ponsaran et al., 2018). Specifically, VESIP exhibited high usability and feasibility ratings, internal consistency reliability from .72 to .82, and expected performance differences between diagnostic groups. Previously collected data from smaller scale usability studies also showed that VESIP was more engaging than semi-structured interviews about hypothetical social challenges.

1.4 | Study goals

The goals of the present study were to: (a) report psychometric properties of VESIP in general education students from grades 3–7 (internal consistency reliability, factor structure, and evidence of criterion-related validity); (b) provide a preliminary examination of SIP skills with respect to a broader range of scenario types than typically probed; and (c) provide evidence of its use as a viable tool for universal assessment of SIP skills in general education settings with minimal administrative burden. In response to the rise in social and emotional learning in schools, this study also aimed to establish the benefit of VESIP for use with general education students and the relationship between performance on VESIP and academic competence (Cooke, 2017).

2 | METHODS

2.1 | Procedures

All recruitment methods, consent procedures, and protocols were approved by the Institutional Review Board (IRB) at Rush University Medical Center (RUMC).

2.1.1 | Recruitment

For the validation sample, participants were recruited during 2016–2017 from three suburban school districts in Illinois based on long-standing partnerships with our research team. For the norming study, those districts and new partner schools were recruited nationally through an open call for participants for a large-scale norming study. Recruitment involved word of mouth, paid electronic advertisements through select educational listservs, cold-calling advertising via postal and electronic mail to school districts nationwide, advertisement through free listservs (e.g., CASEL SEL Assessment workgroup), and vendor tables at psychology and education conferences. Ultimately, 10 districts from six states responded to our call (IL, PA, MI, MS, Washington D.C., and NY). These school districts offered a diverse population from urban, suburban, and rural areas (Table 2).

2.1.2 | Informed consent

For the validation study, parent/guardian consent forms were sent home by participating schools to all students in grades 3–7. For those families who consented, trained members of the research team (bachelor's

TABLE 2 *Demographic information*

	Norming study (*N* = 2,156)		Validation (*n* = 334)	
Measure	***n***	**(%)**	***n***	**(%)**
Ethnicity				
White	460	(21.3)	98	(29.3)
Hispanic	1,169	(54.2)	206	(61.7)
Asian	58	(2.7)	19	(5.7)
Black	195	(9.0)	3	(0.9)
Other	16	(0.7)	2	(0.6)
Unknown	258	(12.0)	6	(1.8)
Grade				
3	206	(9.6)	91	(27.2)
4	456	(21.2)	80	(24.0)
5	618	(28.7)	74	(22.2)
6	422	(19.6)	50	(15.0)
7	454	(21.1)	39	(11.7)
Total	2,156 (1,159 male)		334 (164 male)	
M_{age}	10.99		10.34	
SD_{age}	1.34		1.37	
Districts	10		3	
Schools	>14[a]		7	
States	6		1	

[a]One district partner did not disclose how many schools were represented in their sample.

degree or above) met one-on-one with each student in a spare office or classroom at their school to obtain child assent. In the one case where a student chose to withdraw from the study after parental consent was obtained, no additional information was collected. After acquiring a student's assent, additional assessment measures were administered as described below. Data from consenting students were linked to VESIP data by schools.

Participating school partners either had social and emotional learning curricula in place or were considering use of a curricula, and so they were interested in VESIP data as it related to their program evaluation purposes. The RUMC IRB issued a waiver of informed consent, allowing districts to administer VESIP free of charge to any, or all, students in grades 3–7. In exchange, districts provided de-identified VESIP data for our norming study research purposes. Partner schools received assessment results that would allow them to better understand their students' SIP skills and to inform teaching and learning.

2.2 | Measures

For the validation study, assessment validity of VESIP was examined through an alternate measure of SIP, teacher rating scales of social behavior and academic competence, performance on academic achievement measures, and sociometric assessment.

2.2.1 | VESIP

Students can complete VESIP independently. Once logged in, the student clicks on their user profile from a drop-down menu and selects either to begin a new assessment or continue an incomplete assessment. All students completed a tutorial module, an avatar customization module, and the full 10-scenario assessment (Figure 1; online Appendix). During group administrations, students wore headphones and worked individually on their own computers, separated by open spaces or dividers when available. For each animated scenario, the user's avatar (named 'Alex' in all cases) encounters a challenging social situation which prompts a conversation with a virtual friend ('Dana') about the user's reactions, perspectives, goals, and feelings, mirroring key dimensions of the Crick and Dodge theoretical framework for SIP. After a scenario plays out, Dana asks Alex what their initial reaction is to the situation: 'What do you want to do?' (solution preference, multiple choice). After selecting a solution, the user is asked 'How sure are you that you could do that?' (social self-efficacy, slider scale). Then, the scenario plays out a second time, after which the user is asked 'What just happened?' (problem identification, multiple choice), 'How'd you feel?' (emotion response, multiple choice), 'How mean were they?' (intent attribution, slider scale), and 'How do you want things to turn out?' (goal preference, multiple choice). Responses are recorded automatically. Item scoring rules and descriptions are summarized in Table 3; these have also been described in Russo-Ponsaran et al. (2018).

FIGURE 1 This is a screenshot from VESIP demonstrating the scenario for peer entry into a group. Here, the child's avatar, Alex, watches two children play ball on the playground. Alex wants to join them, but the children do not invite the child to play

2.2.2 | Alternate assessment of SIP

The SIP-AP was chosen as the best available assessment for validation purposes due to its applicability to the same age range, its basis in the Crick and Dodge theoretical model, and its web-based direct assessment format. To conserve resources (e.g., less interruption in the school day) and to utilize scenarios most like VESIP, we opted to administer an abbreviated form of the SIP-AP. Using the Spearman–Brown prophecy formula, we estimated the fewest number of SIP-AP scenarios that would likely produce internal consistency reliabilities ≥ .70. As a result of those calculations, we estimated that a customized four-scenario version requiring 10 min to complete would yield scores with the desired reliabilities. While the SIP-AP covers a broad range of processing dimensions, we included six of the 16 dimensions available that overlapped with VESIP SIP dimensions. These questions include hostile attributional bias ('do you think the boy intended to be mean?'), intentionality attribution readiness ('would you need more information to make a decision about why the boy…?'), angry emotion ('how angry would you feel…?'), revenge goal ('would you want to get back at the boy or get the boy in trouble?'), no prosocial goal ('would you want to get along with the boy?'), and aggressive response ('would you push, hit, call names, or insult the boy?'). Because the SIP-AP measures aggressive tendencies, a higher score indicates lower SIP competency. The internal consistency of the scoring categories as measured by Cronbach's α in our sample ranged from .47 (hostile attributional bias) to .82 (aggressive response) which is consistent with expectations based on the reduction in number of scenarios administered.

2.2.3 | Intellectual ability

The literature suggests that social and emotional skills like SIP are associated with outcomes above and beyond IQ (McKown et al., 2013, 2016). However, it is possible that performance on VESIP is a function of IQ, and any association between VESIP and other variables is because students with higher VESIP scores tend to have higher IQs. To rule out this possibility, we covaried IQ in analyses focused on criterion-related validity. Specifically, we administered a brief, two-subtest form of the Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II; Groth-Marnat, 2003; Kaufman & Lichtenberger, 1999, 2006; Sattler, 2008; Wechsler, 2011) to all consented participants in the validation study. Administration time was approximately 20 min. Matrix Reasoning and Vocabulary make up the two-subtest version. In our sample, reliability was $\alpha = .81$ for Matrix Reasoning and $\alpha = .82$ for Vocabulary.

2.2.4 | Teacher report: School Social Behavior Scale, Second Edition

The School Social Behavior Scale, Second Edition (SSBS-2; Merrell, 2002) is a behavior rating scale that targets social functioning in schools. It is quick to administer and is normed and standardized for students ages of 5 to 18 years (Merrell, 2002). The assessment includes a Social Competence scale ($\alpha = .97$) and an Antisocial Behavior scale ($\alpha = .96$), each with 32 items. Social Competence includes subscores for Peer Relations ($\alpha = .96$), Self-Management/Compliance ($\alpha = .93$), and Academic Behavior ($\alpha = .93$); and Antisocial Behavior includes subscores for Defiant/Disruptive ($\alpha = .91$), Antisocial/Aggressive ($\alpha = .86$), and Hostile/Irritable Behavior ($\alpha = .93$).

2.2.5 | Teacher report: Social Skills Improvement System

During the norming study, a cohort of students in grades 6–7 from one district also had teacher-report on the Social Skills Improvement System (SSIS; Gresham & Elliott, 2008) available. That data were linked to de-identified VESIP

scores and shared with us. The SSIS has a Social Skills scale (α = .98) that includes subscales of Communication (α = .92), Cooperation (α = .94), Assertion (α = .88), Responsibility (α = .94), Empathy (α = .95), Engagement (α = .92), and Self-Control (α = .96); a Problem Behaviors scale (α = .92) that includes subscales of Externalizing (α = .89), Bullying (α = .82), Hyperactivity/Inattention (α = .88), Internalizing (α = .85), and Autism Spectrum (Part A: α = .90, Part B: α = .72); and an Academic Competence scale (α = .97). We hypothesized that overall higher scores on VESIP would be related to higher ratings on Social Skills and Academic Competence, and lower ratings on Problem Behaviors. Although observed social behavior is a few steps removed from one's ability to reason through social situations, we still expected moderate relationships to exist.

2.2.6 | Sociometric assessment: Peer nominations

Some partner schools previously collected peer nomination data through a complementary web-based assessment called SELweb EE (McKown et al., 2016). Partner schools who wished to adopt VESIP and who had already administered the peer nominations assessment as part of their routine practice were asked to provide us with peer nomination data for any students in grades 3–5 who also consented to the validation study (n = 200). Because students in grades 6–7 were not typically all in one classroom, they did not have peer nomination data. In this assessment, students wore headphones and saw the names of each classmate on the screen while the integrated text was read aloud. After the presentation of each classmate, the student was asked to indicate whether they like the classmate. The student had the option to click on a 'yes' button or a 'next' button.

For students who completed the peer nomination assessment (60% of validation participants), three scores were computed. First, a most-liked (ML) score was computed by tallying the number of nominations as a liked peer each student received. To adjust for differences in classroom size, we computed a within-class z-score by subtracting the classroom's mean number of nominations received from each student's actual number of nominations received and dividing that difference by the class-wide standard deviation of nominations received. The same procedure was used to compute a within-class z-score for nominations as a least-liked (LL) peer. Finally, we computed a social preference score by subtracting the LL z-score from the ML z-score and re-standardizing that difference within class. From this procedure, the resulting three scores (ML, LL, and SP) all had a mean of zero and a standard deviation of one.

2.2.7 | Academic achievement

Because a goal of this study was to understand SIP skills in general education students and to examine the relationship between SIP and academic performance, the Reading Curriculum-Based Measurement (R-CBM) and Mathematic Concepts and Applications (M-CAP) subtests of AIMSweb (NCS Pearson, 2012) were administered. For the R-CBM, students were asked to read passages from narrative fiction stories aloud for 1 min; scores were based on words read correctly. For the M-CAP, students were asked to complete as many math problems as they could in 8 (grades 3–6) or 10 (grade 7) minutes. Points for each correct item depended on the level of difficulty. AIMSweb has been shown to have equal difficulty across grade levels and has published alternate form test–retest reliability ranges from .93 to .95 for R-CBM and from .80 to .86 for M-CAP (NCS Pearson, 2012). National percentiles were used in analyses. Districts who administered part, or all, of AIMSweb as part of their standard assessment practice provided the data for students who participated in the validation study. For those students participated in the validation study and did not have AIMSweb data available, a trained research staff member administered the appropriate subtests during the validation testing session.

TABLE 3 VESIP dimension descriptions and scoring

Dimension	Description	Response options and categories	Item score
Solution Preference	*'What do you want to do?'* The extent to which the child selected socially competent solutions to challenging social situations	a. Aggressive (e.g., yell at them) b. Third party (e.g., talk to an adult) c. Passive-avoidant (e.g., don't say anything) d. Prosocial-assertive (e.g., speak up)	0 = Aggressive 1 = Third party 1 = Passive-avoidant 2 = Prosocial-assertive
Problem Identification	*'What just happened?'* The extent to which the child accurately identified the cause of a challenging social situation	a. Benign misunderstanding (e.g., they didn't realize, or I did something) b. Recognition of a social problem with a hostile attribution (e.g., they don't like me, or they're trying to hurt my feelings) c. No understanding (e.g., nothing)	2 = Benign misunderstanding 1 = Recognition of a social problem with a hostile attribution 0 = No understanding
Emotion Response	*'How did you feel when…?'* The extent to which situationally-appropriate emotions are evoked in the child	a. Angry b. Sad c. Ok a. Happy b. Scared c. Worried	1 = Angry 2 = Sad 3 = Ok 0 = Happy 0 = Scared 1 = Worried
Intent Attribution	*'How mean were they?'* The extent to which the child thinks another child is behaving with the intention of being mean or hostile	Sliding scale ranging from 'very mean' to 'not at all mean'	0 (very mean) – 5 (not at all mean)
Goal Preference	*'How do you want things to turn out?'* The extent to which the child favors socially positive goals when confronted with a social challenge	a. No goal (e.g., not sure) b. Retribution (e.g., get back) c. Conflict reduction via third party intervention (e.g., help making things better) d. Prosocial (e.g., work it out) e. Avoidance (e.g., wish it didn't happen)	0 = No goal 0 = Retribution 2 = Conflict reduction via third party intervention 3 = Prosocial 1 = Avoidance
Social Self-Efficacy	*'You chose [solution]. How sure are you that you can do that?'* The extent to which the child believes that s/he could enact preferred solutions	Sliding scale ranging from 'very mean' to 'not at all mean'	0 (not at all sure) – 5 (very sure)
Social Information Processing composite	Global estimate of a child's general social information processing abilities		Composite of solution preference, problem identification, goal preference, emotion response, and intent attribution scores

3 | RESULTS

3.1 | Analysis

3.1.1 | Reliability

The internal consistency reliability (Cronbach's α) of VESIP dimension scores ranged from .62 (problem identification) to .75 (social self-efficacy), in the norming sample and .56 (problem identification) to .75 (social self-efficacy) in the validation sample (Table 4). Because observed dimension scores were on different scales, we calculated the internal consistency reliability (Cronbach's α) of a composite score reflecting SIP using standardized item scores. Reliabilities were .86/.84 for the two samples, respectively.

3.1.2 | Correlations between variables and age

All VESIP dimension raw scores were significantly correlated with age, except for intent attribution (r = .04, p = .06). Pearson r correlation coefficients ranged from .07 (emotion response, p < .001) to −0.16 (problem identification, p < .001).

3.1.3 | Factor structure

There were no missing data for the confirmatory factor analyses. Each VESIP question was designed to measure a separate dimension of SIP. In addition, we included five distinct situation types. This provided us an opportunity to evaluate the fit of the data to hierarchically nested models reflecting three conceptualizations: one that includes five SIP dimensions, a second that adds an overall SIP factor, and a third that adds latent variables reflecting situation type.

To reduce the number of observed variables, we averaged each item score across the two parallel vignette types. For example, we reduced 10 goal preference items to five by averaging the two item scores associated with ambiguous provocation, averaging the two items associated with peer entry, and so on.

We ran preliminary analyses with these observed scores in preparation for confirmatory factor analyses. Skewness and kurtosis for all observed variables was < |3|, a value recommended by Kline (2015) for testing univariate normality. Table 5 presents zero-order correlations between VESIP dimension raw scores. A correlation including all observed scores in the models is available upon request (Supporting Information Table S8).

TABLE 4 Internal consistency reliability (Cronbach's α) across two data sets

	Norming study (N = 2,156, 1,159 male)		Validation study (n = 334, 164 male)	
VESIP dimensions (score ranges)	**α**	***M* (*SD*)**	**α**	***M* (*SD*)**
Solution Preference (0–2)	0.73	1.55 (.32)	0.66	1.61 (.27)
Problem Identification (0–2)	0.62	1.50 (.27)	0.56	1.59 (.23)
Goal Preference (0–3)	0.75	1.93 (.67)	0.70	2.07 (.59)
Emotion Response (0–3)	0.69	2.11 (.44)	0.67	2.15 (.41)
Intent Attribution (0–5)	0.71	3.06 (.83)	0.72	3.19 (.79)
Social Self-Efficacy (0–5)	0.75	3.88 (.79)	0.75	3.95 (.77)
Social Information Processing composite	0.86		0.84	

TABLE 5 Relationship between VESIP dimension raw scores in the norming data set

	1	2	3	4	5	6
1. Solution preference						
2. Problem identification	0.41***					
3. Goal preference	0.59***	0.41***				
4. Emotion response	0.31***	0.12***	0.12***			
5. Intent attribution	0.29***	0.22***	0.17***	0.48***		
6. Social self-efficacy	0.08***	−0.01	0.10***	0.16***	0.07***	

Note: Table shows Pearson's bivariate correlation coefficients.
***$p < .001$

We used Amos version 24.0.0 (Arbuckle, 2014) Maximum Likelihood estimation. Consistent with recommended practices (Jackson et al., 2009), we report the Comparative Fit Index (CFI) and Root Mean Squared Error of Approximation (RMSEA). Although there is not perfect agreement on the cutoff for determining 'good enough' model fit, consistent with (Hu & Bentler, 1999), we considered a CFI > .95 and RMSEA < .06 to reflect a good fit of the model to the data.

We tested a series of hierarchical confirmatory models. First, we fit the data to a five-factor solution including factors reflecting emotion response, problem identification, degree of hostile intent, goal preference, and solution preference. Each latent variable included five observed variables as indicator scores. We conceptualized social self-efficacy as distinct from the SIP dimensions, and so did not include it in this model. Fit statistics suggested a marginal fit of the data to this model (CFI = .84, RMSEA = .06 (90% Confidence Interval (CI) [.06, .06])).

Next, we added an overall SIP latent variable, with all the observed scores loading on this overall score in a bifactor model. The fit of data to this model was better (CFI = .90, RMSEA = .05 (90% CI [.05, .05]). The $\chi^2(25)$ difference in model fit between the first and second model was 704.9, significant well beyond $p < .05$. Factor loadings were variable in magnitude and significance.

Next, we added five factors that loaded on the observed variables associated with each situation type. The fit of this model was substantially better than the prior model (CFI = .97, RMSEA = .03 (90% CI [.03, .03]). The $\chi^2(25)$ difference in model fit between the second and third model was 903.2 ($p < .05$). This final model is illustrated (Figure 2; Table 6).

3.1.4 | Data reduction

Based on the finding from the factor analysis, we created a summary score for each dimension of SIP by averaging the scores within that dimension and standardizing the average scores. In turn, we used all VESIP dimension scores, except social self-efficacy, to create an overall SIP composite. To do so, we averaged z-scores reflecting each dimension of SIP and re-standardized this average score to set the scale with a mean of zero and standard deviation of one. Both the SIP composite score, the domain scores that make up the composite, and social self-efficacy were all used in the analyses of criterion-related validity described below.

3.1.5 | Comparison of students in validation and norming samples

A subset of students whose parents consented to their participation completed VESIP and validation measures. Compared with the larger VESIP sample, students who completed the validation study were not significantly different in terms of the percentage of boys (45.4% vs. 50.9%). On 17 of 25 observed VESIP scores, compared with

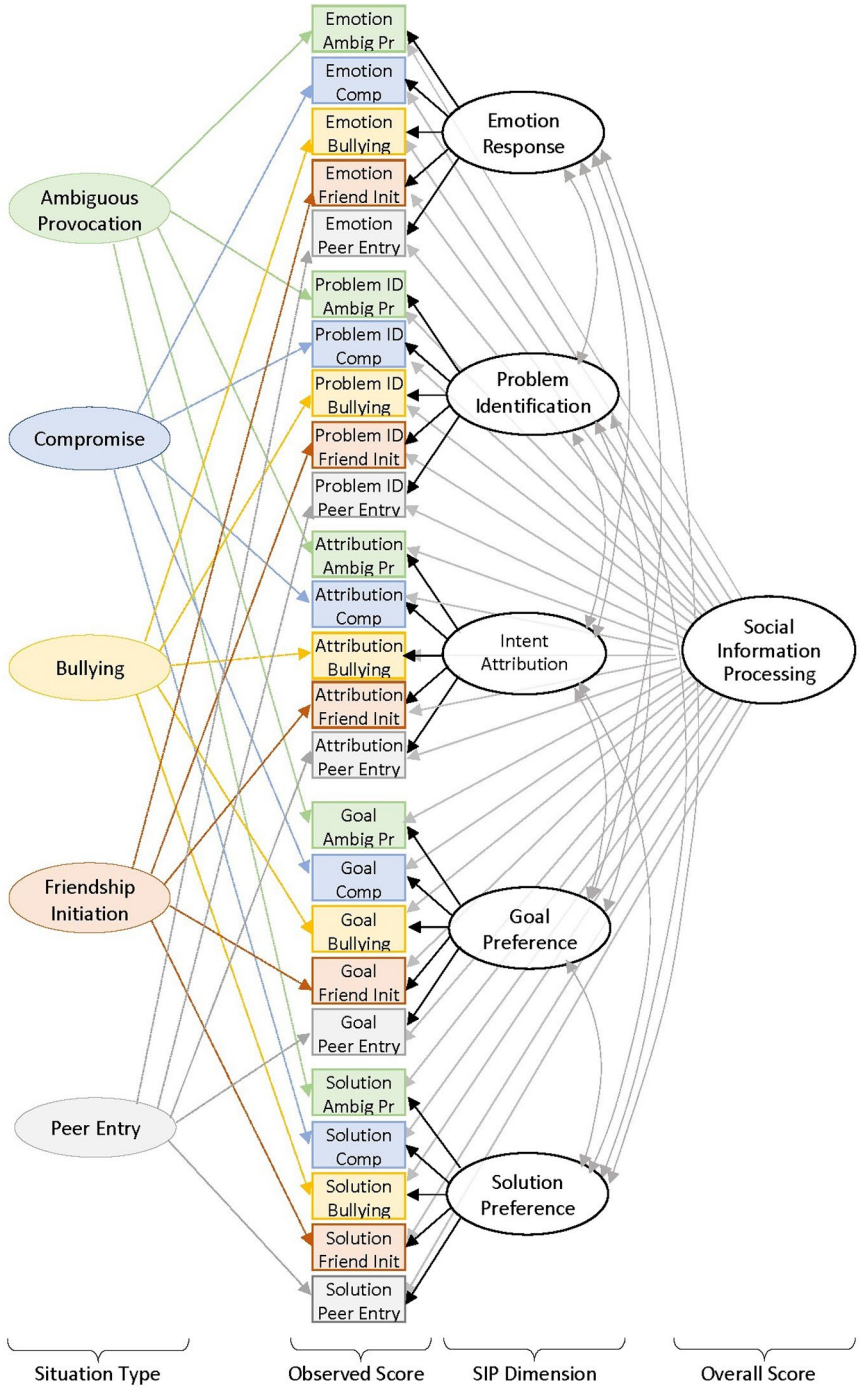

FIGURE 2 Confirmatory factor model of social information processing based on VESIP scores. Ambig Pr, Ambiguous Provocation scenario; Comp, Compromise scenario; Friend Init, Friendship Initiation scenario

the larger VESIP study, students who completed the validation study scored significantly higher than students who did not and in all cases the correlation ration (eta squared) was less than .02. In addition, compared with the larger VESIP study, students who completed the validation study differed in ethnic composition, (10.5% vs. 0.9%

TABLE 6 Standardized factor loadings

	SIP dimension					Situation type					
Item	Emo	ID	IA	Goal	Sol	Ambiguous provocation	Compromise	Bullying	Friendship initiation	Peer entry	SIP
Emotion Response											
Ambiguous Provocation	.55					.34					.15
Compromise	.58						.11				−.11
Bullying	.49							.33			.00
Friendship Initiation	.55								.17		−.02
Peer Entry	.61									.41	.06
Problem Identification											
Ambiguous Provocation		.38				.28					.34
Compromise		.49					.08				.11
Bullying		.40						.30			.10
Friendship Initiation		.46							.28		.21
Peer Entry		.55								.44	.19
Intent Attribution											
Ambiguous Provocation			.61			.39					.24
Compromise			.59				.16				−.06
Bullying			.38					.46			−.12
Friendship Initiation			.64						.45		.06
Peer Entry			.64							.55	.01
Goal Preference											
Ambiguous Provocation				.33		−.19					.57
Compromise				.37			.46				.40
Bullying				.44				.22			.42
Friendship Initiation				.57					.19		.27
Peer Entry				.64						.26	.16

(Continues)

TABLE 6 (Continued)

Item	SIP dimension					Situation type					
	Emo	ID	IA	Goal	Sol	Ambiguous provocation	Compromise	Bullying	Friendship initiation	Peer entry	SIP
Solution Preference											
Ambiguous Provocation					.35	.09					.45
Compromise					.51		.25				.28
Bullying					.42			.27			.33
Friendship Initiation					.58				.20		.35
Peer Entry					.66					.13	.22

Note: Underlined values not statistically significant.

Abbreviations: Emo, emotion response; Goal, goal preference; IA, intent attribution; ID, problem identification; SIP, social information processing; Sol, solution preference.

African American; 2.1% vs. 5.7% Asian American; 19.9% White vs. 29.3% White; 52.9% and Hispanic vs. 61.7% Hispanic in the VESIP sample compared with the validation sample, respectively).

3.1.6 | Criterion-related validity

To evaluate the relationship between performance on VESIP and criterion measures, we ran a series of regression models with VESIP as the predictor and controlling for age and estimated IQ. Criterion measures included scores from the SIP-AP, teacher ratings scales, peer nominations, and AIMSweb. We next present results for each of the VESIP composites and their constituent scores.

SIP-AP

Performance on the SIP-AP is conceptually the closest analog to performance on VESIP because both are web-based assessments of SIP skills. VESIP is scored such that more positive responses yield higher scores, and SIP-AP is scored such that more aggressive responses yield higher scores. Relationships between VESIP and the SIP-AP were all in the expected direction, higher scores on VESIP were associated with lower scores on SIP-AP. The VESIP SIP score was significantly associated with five of six SIP-AP scores, with the exception of intentionality attribution readiness (Table 7). The VESIP overall SIP score was most strongly associated with the hostile attributional bias score, angry emotion, and revenge goal scores ($\beta = -.31$, $p < .001$, all comparisons).

Among the individual VESIP dimension scores, there were several noteworthy results. First, the VESIP solution preference score was more strongly associated with the SIP-AP aggressive solution score than with other SIP-AP scores. Second, the VESIP emotion response score was more strongly associated with the SIP-AP angry emotion score than with other SIP-AP scores. Third, the VESIP goal preference score was associated with the SIP-AP prosocial goal score, and the magnitude of the association was equal to the association between VESIP goal preference and aggressive responding, and greater than the association between VESIP goal preference and any other SIP-AP score. Fourth, the VESIP intent attribution score was more strongly associated with SIP-AP hostile attribution and angry responding than with other SIP-AP scores. Finally, VESIP problem identification was associated with several SIP-SP scores. VESIP social self-efficacy was less consistently and strongly associated with SIP-AP scores, which did not include a social self-efficacy dimension.

SSBS-2

VESIP solution preference and goal preference scores were significantly associated with social competency on the SSBS-2 and were in the expected direction ($\beta = .12$ and .14, respectively, $p < .05$). Emotion response scores on VESIP were negatively associated with antisocial behavior ($\beta = -.11$, $p < .05$). No other VESIP scores were associated with SSBS-2 scores.

Peer nominations

The overall SIP composite score on VESIP was significantly associated in the expected direction with nominations as a least-liked peer ($\beta = -.16$, $p < .05$); scores on solution preference and problem identification mirrored this association ($\beta = -.14$, $p < .05$ for both). Social self-efficacy scores on VESIP were significantly associated in the expected directions with each of the nomination scores (ML, $\beta = .15$, $p < .05$; LL, $\beta = -.20$, $p < .01$; and social preference, $\beta = -.19$, $p < .01$).

AIMSweb

Performance on VESIP was not significantly associated with performance on AIMSweb.

TABLE 7 Relationship between VESIP and SIP-AP scores, controlling for age and estimated IQ

	SIP-AP scores					
Predictors	**Host**	**IR**	**Ang**	**Rvg**	**NoPr**	**Agg**
Age	.07	−.07	−.02	.07	.18**	.12*
IQ	−.11*	−.05	−.01	−.14*	.06	−.12*
Solution Preference	−.12*	−.09	−.11*	−.18***	−.13*	−.22***
Age	.04	−.05	−.04	.04	.15**	.11*
IQ	−.12*	−.05	−.02	−.15**	.05	−.14**
Problem Identification	−.27***	.10	−.17**	−.30***	−.20***	−.19***
Age	.07	−.07	−.01	.07	.17**	.12*
IQ	−.11*	−.05	−.02	−.14**	.06	−.13*
Goal Preference	−.14**	−.10	−.04	−.15**	−.20***	−.20***
Age	−.11*	−.08	−.04	.11*	.20***	.16**
IQ	−.08	−.08	.05	−.12*	.07	−.11*
Emotion Response	−.18***	.11	−.36***	−.15**	−.11	−.15**
Age	.12*	−.08	.04	.12*	.20***	.16**
IQ	−.05	−.09	.05	−.09	.07	−.10
Intent Attribution	−.33***	.15**	−.35***	−.26***	−.08	−.16**
Age	.08	−.07	−.01	.09	.19***	.14**
IQ	−.07	−.06	.03	−.10	.09	−.09
Social Information Processing composite	−.31***	−.05	−.31***	−.31***	−.22***	−.27***
Age	.08	−.06	−.01	.09	.19***	.14*
IQ	−.11*	−.05	−.01	−.14*	.06	−.13*
Social Self-Efficacy	−.09	−.03	−.11	−.15**	−.12*	−.10

Note: Table shows standardized regression coefficients. SIP-AP scores: Agg, aggressive solution; Ang, angry emotion; Host, hostile attribution bias; IR, intent attribution readiness; NoPr, no prosocial goal; Rvg, revenge goal.
*$p < .05$; **$p < .01$; ***$p < .001$.

SSIS

Analyses supported associations between performance on VESIP and multiple subscales of the SSIS, a commonly used teacher behavior rating scale. Specifically, the SIP composite, social self-efficacy, solution preference (β = .17, .16, and .19, respectively, $p < .01$), and intent attribution (β = .13, $p < .05$) were all associated in the expected direction with Academic Competence scores. Solution preference was associated with seven of the nine subscale scores and each summary score on the SSIS (Social Skills, Academic Competence and Problem Behaviors). Emotion response on VESIP predicted the bullying subscale and problem behaviors composite scores (β = −.14 and −.11, respectively, $p < .05$). Social self-efficacy scores also predicted teacher-report of responsibility (β = .13, $p < .05$), problem behaviors (β = −.14, $p < .05$) and three of the subscales.

3.1.7 | Test–retest data

A sample of 63 students in grades 3–7 from one school that participated in our large-scale field trial also completed VESIP at a second time point. Data collection periods were separated by approximately 2 weeks. Those data supported no significant difference in response patterns (paired samples *t*-tests, $.06 \leq p \leq .88$) and SIP composite scores were highly correlated ($r = .66$, $p < .001$).

4 | DISCUSSION

VESIP builds on the strengths of SIP theory and extends the accessibility of SIP assessment. Design elements were intended to more closely approximate the experience of being involved in a wide range of socially challenging situations and to create a more engaging assessment by incorporating personal features and preferences into the assessment (Stapleton, 2004; Verhoef et al., 2019). The primary goals of this study were to: (a) report psychometric data on a newly developed web-based assessment, VESIP; (b) provide a preliminary examination of SIP skills with respect to a broader range of scenario types than typically probed; and (c) provide evidence of VESIP's use as a viable tool for universal assessment of SIP skills in general education settings across grades 3–7. It was validated in a sample of 334 students and normed in a large and diverse sample of 2,156 students.

4.1 | Summary and significance of findings

4.1.1 | Score reliabilities

VESIP score reliabilities were variable for individual dimensions, with problem identification as low as .62, for example, while the SIP composite was much higher at .86. This suggests two conclusions for applied use. First, the individual score reliabilities are variable enough that interpretation of these scores for understanding individual student SIP skills should be undertaken with caution. Variable reliabilities pose the risk that a student's observed score might substantially over- or under-estimate their true skill level. For these scores, educators wishing to use VESIP to understand student SIP skills would do well to look at scores aggregated at the group level, such as the classroom level. Doing so provides information about the strengths and needs of the group, without the risk of making an inaccurate inference about the skill level of an individual student. Individual student scores should include confidence intervals to communicate to those interpreting the scores a reasonable range within which the student's true score is likely to fall. In so doing, measurement imprecision will be communicated in an immediately useful way. In contrast to scores reflecting the individual SIP dimensions, the composite score reliability was .86 in the norming sample. Because of its higher reliability, this overall score provides a more consistent estimate of the skills VESIP measures. As a result, practitioners may be able to use these scores to understand an individual student's skill level. However, it is important to note that because VESIP is not designed to be a diagnostic tool and these reliabilities are below .90, VESIP is not suitable for rendering a diagnosis, screening for disorder, or for using in high-stakes decisions about student placement in special education or other programs. To better understand a student's true SIP skill level, it is recommended that VESIP be used in conjunction with more traditional measures.

4.1.2 | VESIP factor structure

Confirmatory factor analyses support and extend models of SIP. Specifically, VESIP scores fit a factor structure that includes three distinct facets. The first facet is consistent with SIP theory and includes five factors,

one for each distinct SIP dimensions. Those dimensions include students' initial response to a situation (solution preference), how they initially make sense of the problem (problem identification), their emotional response to a situation (emotion response), their interpretation of degree of hostile intent (intent attribution), and their desired outcomes (goal preference). A confirmatory model with these factors fit the data marginally well, suggesting both that the model supports SIP theory overall and that there may be room for improvement. Adding an overall SIP factor modestly improved model fit. Adding factors reflecting situation types improved model fit considerably.

This suggests two conclusions. First, the SIP dimensions may be applied somewhat differently from one situation to another. In other words, being able to navigate one socially challenging situation does not necessarily mean one can navigate a different situation with equal aplomb. On the other hand, even when ignoring situational differences in SIP, the data fit the model reasonably well, suggesting that dimensions do generalize to some degree across different situation types. Therefore, it is reasonable to estimate SIP tendencies overall.

Is it possible that some dimensions of SIP are more situation dependent than others? Inspection of the factor loadings from the full confirmatory model suggests that may be the case. Each SIP dimension loaded on five different situation factors. Larger factor loadings reflect a greater impact of situation type on an SIP dimension. By inspecting the average standardized factor loadings of each SIP dimension on situation factors, we can therefore develop some initial conclusions about which SIP dimensions are more situation dependent than others. Average factor loadings on SIP dimensions, in descending order, are .40 for intent attribution, .28 for problem identification, .27 for emotion response, .19 for goal preference, and .15 for solution preference. This suggests that intent attribution is highly situation dependent, that how students define social problems and respond emotionally to them are also situation dependent, and that students' goal and solution preferences are less situation dependent. Taken together, this means that the magnitude of the provocation in a social situation may influence early-stage information processing (intent attribution, emotion response, and problem identification) whereas students' goal and solution preferences are relatively consistent, regardless of the situation type. Further research is needed to fully understand which dimensions of SIP are situation dependent, and which are consistent across situations.

Finally, we conceptualized social self-efficacy as distinct from other dimensions of SIP. In contrast to the other scores which reflect SIP skills, social self-efficacy reflects the student's appraisal of their ability to enact a preferred solution. In prior research, self-efficacy was considered part of response decision (Crick & Dodge, 1994). Independent of other dimensions of SIP, social self-efficacy was associated with other criterion measures, suggesting it is an important and distinct predictor of functioning. Future work should validate self-efficacy more specifically.

4.1.3 | Convergent and discriminant validity

Analyses generally supported the convergent and discriminant validity of VESIP scores. Specifically, VESIP scores reflecting intent attribution, emotion response, goal preference, and solution preference were more strongly associated with SIP-AP scores reflecting the same or similar constructs than those scores reflecting different dimensions of SIP. VESIP problem identification and social self-efficacy scores were associated with several SIP-AP scores, but because there were no SIP-AP scores that directly assessed these dimensions, it was not possible to infer the convergent or discriminant validity of those scores.

4.1.4 | VESIP criterion-related validity

Performance on VESIP was generally associated with criterion measures in expected ways. SIP skills are widely reported to be associated with socially competent behavior (e.g., Dubow & Tisak, 1989; Dubow et al., 1991; Wentzel, 1991), problem behaviors (e.g., de Castro et al., 2002; Verhoef et al., 2019), and peer acceptance (e.g.,

Bauminger et al., 2005; Crick & Dodge, 1994; McKown, 2007; McKown et al., 2009; Weissberg et al., 1997). While our data provided some support for these relationships through our teacher-report (SSBS-2 and SSIS) data, these were not the strongest findings. One explanation may be that the relationship to aggressive behavior was not as robust because VESIP explores several non-confrontational yet socially challenging situations, like making a new friend and learning to compromise. Neither of these situations would necessarily invoke aggression. Dimensions of SIP were also associated with subscales on the teacher reports, but the associations were inconsistent. However, when evaluating SIP skills with respect to peer relationships, we do see a correspondence between SIP skills as measured by VESIP and least-liked status.

Because VESIP was designed, in part, to provide a universal tool for educators, we also sought to better explain the relationship between SIP skills and academic abilities. The SIP composite score from VESIP was related to teacher-report of academic competence on the SSIS. Solution preference, intent attribution, and social self-efficacy also showed significant relationships to that scale. While significant, coefficients were likely modest because VESIP has stronger associations with more proximal skills. VESIP performance was also unrelated to discrete, brief reading and math achievement tests on AIMSweb. Because those scores were taken from only one timepoint as a means for exploring VESIP's relationship to academic performance, interpretation is limited. Future work should evaluate other metrics of academic performance and success.

4.1.5 | VESIP feasibility and suitability for universal administration

A VESIP manual and supplemental training materials (e.g., video tutorials and score interpretation guides) are available to school districts. VESIP's platform makes it suitable for use in school settings and for universal administration. As evidence of its feasibility and suitability, VESIP was successfully administered to hundreds of students per day. Enhancing its usability and versatility for administrators, students can either be uploaded individually or via roster uploads. Use of VESIP is designed to be straight-forward, requiring only brief remote training sessions for school administrators, teachers, or other staff (e.g., librarians, psychologists) who will manage data or proctor the assessment. The student's self-administration process allows for rapid assessment delivery, with little burden on test administrators. Assessments are automatically saved and scored. Score reports can be downloaded for an entire classroom, multiple classrooms, or individual students. Thus, results from VESIP may be used as a formative assessment to estimate student SIP skills and use what is learned to guide instruction and help them decide what skills to emphasize for each academic year. Scores that are aggregated at the classroom level or above will provide specific enough information to guide instruction without risking an inaccurate appraisal of individual students whose observed score under- or over-estimates their actual skill level.

4.2 | Limitations and future directions

4.2.1 | Demographics

Recruitment for the norming study was open to any district nationwide who responded to our call for partners. Presumably, schools who were already knowledgeable about social and emotional learning curricula or located in states with social and emotional learning standards were more inclined to participate. Our resulting demographic for both the norming and validation samples included a majority of Hispanic participants (54% and 61%, respectively), with White alone being the second largest demographic (21% and 29%, respectively). This occurrence presents both a strength, in that not many SIP studies focus on Hispanic students, and a limitation, in that there is no certainty with how well our findings will generalize in the broader national community. Future studies should explore other user characteristics, such as first language and socioeconomic status, to better understand the

influence of participant demographics on SIP (Ellis et al., 2017). For representativeness, national norms should seek to recruit participants who more closely mirror the composition of the United States.

While this study focused on the assessment of general education students, extensions of this study to special education populations and further assessment in clinical populations will be important in understanding its broader application.

4.2.2 | Score reliabilities

While adequate, internal consistencies of individual SIP dimensions were modest. This finding has two likely sources. First, by capturing forced-choice responses to which scoring rules are applied, VESIP may reflect less nuance in students' responding than if they were to respond to Likert-type response options, as is the case with the SIP-AP. Second, the wide variety of situation types and the modest situation dependence of SIP may have attenuated internal consistency. Because situation specificity was particularly pronounced with regard to emotion response, problem identification, and intent attribution, it is not surprising that internal consistency for these dimensions was lower than the more situation-independent scores, such as goal preference and solution preference.

To address modest score reliabilities, future work might evaluate the impact of revising assessment features. In particular, the forced choice response options may attenuate reliability. The SIP-AP uses Likert-type scales for children to enter their responses to questions that are similar to VESIP's, and this resulted in score reliabilities in excess of .80 (Dodge et al., 1985; Kupersmidt et al., 2011). A revision as simple as a change to the response options might increase score reliability substantially. To some extent, the benefit of additional situational coverage may offset the lower reliabilities by providing an assessment of SIP over a substantially wider range of situations than is possible with other existing measures.

4.2.3 | Quantitative versus qualitative data

Standard scores and normative data are important to interpret performance across grade levels and students. As such, higher scores on VESIP indicate more prosocial responses. Understanding performance on individual SIP dimensions can provide information useful for designing interventions that target students' specific needs (Cooke, 2017). Recognizing also that life experience may influence responses, it is important not to take a deficit model approach to the data (Ellis et al., 2017). Youth coming from a 'high-stress' background (e.g., low socioeconomic status, violent environment) may rely on adaptive and compensatory behaviors. Being able to respond to a hostile situation in an aggressive or avoidant way may be protective. In contrast, for youth coming from a low-stress background may engage in less aggressive and more assertive ways, thus yielding a higher score on VESIP. For these reasons, in addition to quantitative data, qualitative information about responses (e.g., passive avoidant, aggressive) is also generated. While examining more qualitative aspects is outside the scope of this study, access to this type of information, in conjunction with the students' life circumstances, should be viewed in parallel to understand quantitative scores. A skilled educator or clinician may apply this information when tailoring either further evaluation or intervention.

4.2.4 | Influence of scenario on dimensions of SIP

The confirmatory factor analysis suggested the potential relevance of scenario specificity on certain aspects of SIP. To fully evaluate these relationships, future work should include a repertoire of additional scenarios depicting not only the categories of scenarios tested with VESIP, but also a wider range of social situations. It will be important to understand how many scenes of each type is optimal or necessary to elucidate the relationships and then the extent to which SIP skills are situation specific or general processing biases.

4.2.5 | Generalizability

There are few assessments that are available to measure SIP in English language learners. A trans-adapted, Spanish-language version of VESIP was also developed during the course of this study, making it one of less than a handful of Spanish-language assessments for youth (e.g., Bar-On & Parker, 2000; Gresham & Elliott, 2008; Russo et al., 2018). To date, only a small number of youth have completed that version, but a related large-scale validation effort for use by English-language learners is planned. It will be important to assess not only language experience but, as mentioned previously, other life circumstances that may influence one's SIP patterns. Finally, VESIP may also serve as a useful progress monitoring or outcome measure with respect to both educational programming and clinical use. Evaluating the utility of VESIP for these purposes is another important goal of future work.

5 | CONCLUSION

VESIP is technically sound and feasible for universal administration to large numbers of elementary and middle school general education students. Our data suggest that VESIP's dynamic assessment platform: (a) provides an engaging format for measuring theoretically supported SIP skills in both girls and boys, (b) allows for a more personalized assessment than existing measures, and (c) offers promise as a new tool for assessing a broad range of social situations and SIP dimensions.

ACKNOWLEDGMENTS

We thank the educational partners and students who contributed to this research. We also thank Soar Technology, Inc., Drs. Everett Smith, Louis Fogg, and Jaclyn Russo, and Mr. Jason Johnson (deceased) for their contributions to this work. The research reported here was supported by the Institute of Education Sciences (IES) through grant R305A15189 to Rush University Medical Center to NRP, National Institute on Disability and Rehabilitation Research (NIDRR) through grant H133S110048 to Soar Technology, Inc., and Rush Translational Sciences Consortium through a grant to NRP. The opinions expressed are those of the authors and do not represent views of IES, NIDRR, or the U. S. Department of Education.

CONFLICT OF INTEREST

N.R.P. and C.M. have financial interests in/relative to xSEL Labs, Inc., which could potentially benefit from the outcomes of this research.

DATA AVAILABILITY STATEMENT

Data available on request due to privacy/ethical restrictions.

ORCID

Nicole M. Russo-Ponsaran https://orcid.org/0000-0002-9430-152X

REFERENCES

Arbuckle, J. L. (2014). *Amos (Version 24.0.0). [Computer Program]*. IBM SPSS.

Bar-On, R., & Parker, J. D. A. (2000). *The bar-on emotional quotient inventory: Youth version (EQ-i:YV) technical manual*. Multi-Health Systems Inc.

Baron-Cohen, S. (1989). The autistic child's theory of mind: A case of specific developmental delay. *Journal of Child Psychology and Psychiatry*, *30*(2), 285–297. https://doi.org/10.1111/j.1469-7610.1989.tb00241.x

Bauminger, N., Schorr Edelsztein, H., & Morash, J. (2005). Social information processing and emotional understanding in children with LD. *Journal of Learning Disabilities*, *38*(1), 45–61. https://doi.org/10.1177/00222194050380010401

Bauminger-Zviely, N., Alon, M., Brill, A., Schorr-Edelsztein, H., David, T., Tubul, G., & Al-Yagon, M. (2019). Social information processing among children with ASD, SLD, and typical development: The mediational role of language capacities. *The Journal of Special Education*, *53*(3), 153–165. https://doi.org/10.1177/0022466918821400

Bell, D. J., Luebbe, A. M., Swenson, L. P., & Allwood, M. A. (2009). The children's evaluation of everyday social encounters questionnaire: Comprehensive assessment of children's social information processing and its relation to internalizing problems. *Journal of Clinical Child & Adolescent Psychology*, *38*(5), 705–720. https://doi.org/10.1080/15374410903103585

Bowers, L., Huisingh, R., & LoGiudice, C. (2005). *Test of problem solving-3 elementary (TOPS 3): Examiner's manual*. LinguiSystems Inc..

Bowers, L., Huisingh, R., & LoGiudice, C. (2007). *Test of problem solving-2 adolescent (TOPS-2): Examiner's manual*. LinguiSystems Inc.

Channon, S., Crawford, S., Orlowska, D., Parikh, N., & Thoma, P. (2014). Mentalising and social problem solving in adults with Asperger's syndrome. *Cognitive Neuropsychiatry*, *19*(2), 149–163. https://doi.org/10.1080/13546805.2013.809659

Coccaro, E. F., Fanning, J. R., Fisher, E., Couture, L., & Lee, R. J. (2017). Social emotional information processing in adults: Development and psychometrics of a computerized video assessment in healthy controls and aggressive individuals. *Psychiatry Research*, *248*, 40–47. https://doi.org/10.1016/j.psychres.2016.11.004

Committee for Children. (2008). *Second step: Student success through prevention program*. Author.

Conduct Problems Prevention Research Group. (1991). *Home interview with child (HIC)*. http://www.fasttrackproject.org/techrept/h/hic/

Cooke, T. (2017). Social information processing: A useful framework for educational psychology. *Educational Psychology Research and Practice*, *3*(1), 50–69. https://www.uel.ac.uk/schools/psychology/research/educational-psychology-research-and-practice/volume-3-no-1-2017

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin*, *115*(1), 74–101. https://doi.org/10.1037//0033-2909.115.1.74

Crick, N. R., & Dodge, K. A. (1996). Social information-processing mechanisms in reactive and proactive aggression. *Child Development*, *67*(3), 993–1002. https://doi.org/10.2307/1131875

de Castro, B. O., Merk, W. W., Koops, W., Veerman, J. W., & Bosch, J. D. (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. https://doi.org/10.1207/s15374424jccp3401_10

de Castro, B. O., Veerman, J. W., Koops, W., Bosch, J. D., & Monshouwer, H. J. (2002). Hostile attribution of intent and aggressive behavior: A meta-analysis. *Child Development*, *73*(3), 916–934. https://doi.org/10.1111/1467-8624.00447

Denham, S. A. (2006). Social-emotional competence as support for school readiness: What is it and how do we assess it? *Early Education & Development*, *17*(1), 57–89. https://doi.org/10.1207/s15566935eed1701_4

Denham, S. A., Bouril, B., & Belouad, F. (1994). Preschoolers' affect and cognition about challenging peer situations. *Child Study Journal*, *24*(1), 1–21.

Denham, S. A., Ji, P., & Hamre, B.. (2010). *Compendium of social-emotional learning and associated assessment measures*. Collaborative for Academic, Social, and Emotional Learning.

Dickson, J., & MacLeod, A. (2004). Approach and avoidance goals and plans: Their relationship to anxiety and depression. *Cognitive Therapy and Research*, *28*(3), 415–432. https://doi.org/10.1023/B:COTR.0000031809.20488.ee

Dodge, K. A., Laird, R., Lochman, J. E., & Zelli, A. (2002). Multidimensional latent-construct analysis of children's social information processing patterns: Correlations with aggressive behavior problems. *Psychological Assessment*, *14*(1), 60–73. https://doi.org/10.1037//1040-3590.14.1.60

Dodge, K. A., McClaskey, C. L., & Feldman, E. (1985). Situational approach to the assessment of social competence in children. *Journal of Consulting and Clinical Psychology*, *53*(3), 344–353. https://doi.org/10.1037/0022-006X.53.3.344

Dubow, E. F., & Tisak, J. (1989). The relation between stressful life events and adjustment in elementary school children: The role of social support and social problem-solving skills. *Child Development*, *60*(6), 1412–1423. https://doi.org/10.2307/1130931

Dubow, E. F., Tisak, J., Causey, D., Hryshko, A., & Reid, G. (1991). A two-year longitudinal study of stressful life events, social support, and social problem-solving skills: Contributions to children's behavioral and academic adjustment. *Child Development*, *62*(3), 583–599. https://doi.org/10.1111/j.1467-8624.1991.tb01554.x

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446. https://doi.org/10.1037/0012-1649.43.6.1428

D'Zurilla, T. J., Nezu, A. M., & Maydeu-Olivares, A. (2002). *Social problem-solving inventory-revised (SPSI-R): Manual*. Multi-Health Systems.

Elias, M. J., & Butler, L. B. (2005a). *Social decision making, social problem solving: A curriculum for academic, social, and emotional learning: Grades 4–5*. Research Press.

Elias, M. J., & Butler, L. B. (2005b). *Social decision making/social problem solving for middle school students: Skills and activities for academic, social, and emotional success*. Research Press.

Ellis, B. J., Bianchi, J., Griskevicius, V., & Frankenhuis, W. E. (2017). Beyond risk and protective factors: An adaptation-based approach to resilience. *Perspectives on Psychological Science*, *12*(4), 561–587. https://doi.org/10.1177/1745691617693054

Erdley, C. A., & Asher, S. R. (1996). Children's social goals and self-efficacy perceptions as influences on their responses to ambiguous provocation. *Child Development*, *67*(4), 1329–1344. https://doi.org/10.2307/1131703

Garner, P. W., & Lemerise, E. A. (2007). The roles of behavioral adjustment and conceptions of peers and emotions in preschool children's peer victimization. *Development and Psychopathology*, *19*(1), 57–71. https://doi.org/10.1017/S0954579407070046

Gresham, F. M., & Elliott, S. N. (2008). *Social skills improvement system: Rating scales manual*. NCS Pearson.

Groth-Marnat, G. (2003). *Handbook of psychological assessment* (4th. ed.). John Wiley & Sons Inc.

Halle, T. G., & Darling-Churchill, K. E. (2016). Review of measures of social and emotional development. *Journal of Applied Developmental Psychology*, *45*, 8–18. https://doi.org/10.1016/j.appdev.2016.02.003

Hamilton, L. S., Stecher, B. M., Schweig, J., & Baker, G. (2018). *RAND education assessment finder*. https://www.rand.org/pubs/tools/TL308.html

Howe, A., Pit-ten Cate, I. M., Brown, A., & Hadwin, J. A. (2008). Empathy in preschool children: The development of the Southampton Test of Empathy for Preschoolers (STEP). *Psychological Assessment*, *20*(3), 305–309. https://doi.org/10.1037/a0012763

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jackson, D. L., Gillaspy, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, *14*, 6–23. https://doi.org/10.1037/a0014694

Kaufman, A. S., & Lichtenberger, E. O. (1999). *Essentials of WAIS-III assessment*. John Wiley & Sons Inc.

Kaufman, A. S., & Lichtenberger, E. O. (2006). *Assessing adolescent and adult intelligence* (3rd ed.). John Wiley & Sons Inc.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford Publications.

Kupersmidt, J. B., Stelter, R., & Dodge, K. A. (2011). Development and validation of the social information processing application: A web-based measure of social information processing patterns in elementary school-age boys. *Psychological Assessment*, *23*(4), 834–847. https://doi.org/10.1037/a0023621

Lansford, J. E., Malone, P. S., Dodge, K. A., Pettit, G. S., & Bates, J. E. (2010). Developmental cascades of peer rejection, social information processing biases, and aggression during middle childhood. *Development and Psychopathology*, *22*(3), 593–602. https://doi.org/10.1017/S0954579410000301

Lemerise, E. A., & Arsenio, W. F. (2000). An integrated model of emotion processes and cognition in social information processing. *Child Development*, *71*(1), 107–118. https://doi.org/10.1111/1467-8624.00124

Liberman, R. P., Mueser, K. T., Wallace, C. J., Jacobs, H. E., Eckman, T., & Massel, H. K. (1986). Training skills in the psychiatrically disabled: Learning coping and competence. *Schizophrenia Bulletin*, *12*(4), 631–647. https://doi.org/10.1093/schbul/12.4.631

Luebbe, A., Bell, D., Allwood, M., Swenson, L., & Early, M. (2010). Social information processing in children: Specific relations to anxiety, depression, and affect. *Journal of Clinical Child and Adolescent Psychology*, *39*(3), 386–399. https://doi.org/10.1080/15374411003691685

Matthys, W., Cuperus, J. M., & Van Engeland, H. (1999). Deficient social problem-solving in boys with ODD/CD, with ADHD, and with both disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, *38*(3), 311–321. https://doi.org/10.1097/00004583-199903000-00019

Matthys, W., Maassen, G. H., Cuperus, J. M., & van Engeland, H. (2001). The assessment of the situational specificity of children's problems behaviour in peer-peer context. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *42*(3), 413–420. https://doi.org/10.1111/1469-7610.00734

McFall, R. M. (1982). A review and reformulation of the concept of social skills. *Behavioral Assessment*, *4*(1), 1–33. https://doi.org/10.1007/BF01321377

McKown, C. (2007). Concurrent validity and clinical usefulness of several individually administered tests of children's social-emotional cognition. *Journal of Clinical Child & Adolescent Psychology*, *36*(1), 29–41. https://doi.org/10.1080/15374410709336566

McKown, C., Allen, A. M., Russo-Ponsaran, N. M., & Johnson, J. K. (2013). Direct assessment of children's social-emotional comprehension. *Psychological Assessment*, *25*(4), 1154–1166. https://doi.org/10.1037/a0033435

McKown, C., Gumbiner, L. M., Russo, N. M., & Lipton, M. (2009). Social-emotional learning skill, self-regulation, and social competence in typically developing and clinic-referred children. *Journal of Clinical Child & Adolescent Psychology*, *38*(6), 858–871. https://doi.org/10.1080/15374410903258934

McKown, C., Russo-Ponsaran, N., 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. https://doi.org/10.1177/0734282915604564

Merrell, K. W. (2002). *School social behavior scales (SSBS-2)* (2nd ed.). Brookes Publishing.
NCS Pearson. (2012). *AIMSweb technical manual*. NCS Pearson Inc.
Parker, J. G., & Asher, S. R. (1987). Peer relations and later personal adjustment. *Psychological Bulletin*, *102*(3), 357–389. https://doi.org/10.1037/0033-2909.102.3.357
Renshaw, P. D., & Asher, S. R. (1983). Children's goals and strategies for social interaction. *Merrill-Palmer Quarterly: Journal of Developmental Psychology*, *29*, 353–374. http://www.jstor.org/stable/23086266
Russo, J. M., McKown, C., Russo-Ponsaran, N., & Allen, A. (2018). Reliability and validity of a Spanish language assessment of children's social-emotional learning skills. *Psychological Assessment*, *30*(3), 416–421. https://doi.org/10.1037/pas0000508
Russo-Ponsaran, N., Lerner, M. D., McKown, C., Weber, R. J., Karls, A., Kang, E., & Sommer, S. L. (2019). Web-based assessment of social-emotional skills in school-aged youth with Autism Spectrum Disorder. *Autism Research*, *12*(8), 1260–1271. https://doi.org/10.1002/aur.2123
Russo-Ponsaran, N. M., McKown, C., Johnson, J., Russo, J., Crossman, J., & Reife, I. (2018). Virtual environment for social information processing: Assessment of children with and without autism spectrum disorders. *Autism Research*, *11*(2), 305–317. https://doi.org/10.1002/aur.1889
Sattler, J. M. (2008). *Assessment of children: Cognitive foundations*. Jerome M. Sattler, Publisher, Inc.
Schultz, D., Ambike, A., Logie, S. K., Bohner, K. E., Stapleton, L. M., Vanderwalde, H., Min, C. B., & 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. https://doi.org/10.1007/s10802-010-9390-5
Shulman, C., Guberman, A., Shiling, N., & Bauminger, N. (2012). Moral and social reasoning in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *42*(7), 1364–1376. https://doi.org/10.1007/s10803-011-1369-8
Shure, M. B. (1992). *I can problem solve: An interpersonal cognitive problem-solving program: Kindergarten & primary grades*. Research Press.
Stapleton, A. (2004). *Serious Games: Serious Opportunities*. Paper presented at the Australian Game Developers' Conference. Academic Summit.
van Nieuwenhuijzen, M., & Vriens, A. (2012). (Social) Cognitive skills and social information processing in children with mild to borderline intellectual disabilities. *Research in Developmental Disabilities*, *33*(2), 426–434. https://doi.org/10.1016/j.ridd.2011.09.025
van Reemst, L., Fischer, T. F., & Zwirs, B. W. (2016). Social information processing mechanisms and victimization: A literature review. *Trauma, Violence & Abuse*, *17*(1), 3–25. https://doi.org/10.1177/1524838014557286
Verhoef, R. E. J., Alsem, S. C., Verhulp, E. E., & De Castro, B. O. (2019). Hostile intent attribution and aggressive behavior in children revisited: A meta-analysis. *Child Development*, *90*(5), e525–e547. https://doi.org/10.1111/cdev.13255
Wechsler, D. (2011). *Wechsler abbreviated scale of intelligence* (2nd ed.). Pearson.
Weissberg, R. P., Barton, H. A., & Shriver, T. P. (1997). The social-competence promotion program for young adolescents. In T. P. Gullotta (Ed.), *Issues in children's and families' lives: Primary prevention works* (Vol. *6*, pp. 268–290). Sage Publications Inc.
Weissberg, R. P., Goren, P., Domitrovich, C., & Dusenbury, L. (2013). *2013 CASEL guide: Effective social and emotional learning programs—Preschool and elementary school edition*. https://casel.org/preschool-and-elementary-edition-casel-guide/
Wentzel, K. R. (1991). Relations between social competence and academic achievement in early adolescence. *Child Development*, *62*(5), 1066–1078. https://doi.org/10.1111/j.1467-8624.1991.tb01589.x
Ziv, Y. (2013). Social information processing patterns, social skills, and school readiness in preschool children. *Journal of Experimental Child Psychology*, *114*(2), 306–320. https://doi.org/10.1016/j.jecp.2012.08.009

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

Table S8

Supplementary Material

How to cite this article: Russo-Ponsaran NM, McKown C, Karls A, Wu IYH. Psychometric properties of Virtual Environment for Social Information Processing, a social information processing simulation assessment for children. *Soc Dev*. 2021;00:1–26. https://doi.org/10.1111/sode.12512