Journal of Applied Developmental Psychology 55 (2018) 39–51

Contents lists available at ScienceDirect

Journal of Applied Developmental Psychology

journal homepage: www.elsevier.com/locate/jappdp

Development and initial validation of a social emotional learning assessment for universal screening

Stephen N. Elliott[a,f,*], Michael D. Davies[b], Jennifer R. Frey[c], Frank Gresham[d], Greta Cooper[e]

[a] Arizona State University, United States
[b] Griffith University, Australia
[c] George Washington University, United States
[d] Louisiana State University, United States
[e] Queensland Department of Education, Australia
[f] Australian Catholic University, Australia

ARTICLE INFO

Keywords:
Social emotional learning
Universal assessment
Screening
Identification of at-risk children

ABSTRACT

Social Emotional Learning (SEL) is a critical aspect of schooling. While a theoretical model put forward by the Collaborative for Academic, Social, and Emotional Learning (CASEL) has defined five well-accepted components of SEL, few assessments claim to measure these SEL components. This study examined the initial validation of scores for a new universal screening measure called the Social Emotional Learning Screening Assessment (SELA). The SELA's content and internal structure were based on the CASEL five model and the existing SSIS Performance Screening Guide. As part of a larger project, experienced Australian teachers of 268 children from prep through year 3 provided initial user and psychometric evidence for the SELA. The results indicated the teacher-completed SELA is well aligned with the CASEL model and offers educators a time-efficient, sensitive, and reliable measure that effectively identifies students at-risk socially and academically. Although preliminary but promising, further research with the SELA is required to replicate and extend these findings to educators in US schools and to test its application with larger, more diverse samples of students.

A contention is growing world-wide that children need more than traditional academic skills to thrive in the 21st century. In fact, many people believe children and youth need social emotional skills to complement and enable their academic skills (e.g., ACARA, 2013; Weissberg & Cascarino, 2013; World Economic Forum, 2016). This assertion about the importance of social emotional skills is based on research in several countries – Australia, Canada, Denmark, England, and the United States – that demonstrates social emotional learning (SEL) skills are essential for students to meet the challenges of learning, to help them prevent risky personal behaviors, to prepare them for the demands of a changing workplace, and ultimately, to promote their wellbeing (e.g., DiPerna, Volpe, & Elliott, 2002; Durlak, Weissberg, Dyminicki, Taylor, & Schellinger, 2011; Elliott, Frey, & Davies, 2015; Guhn et al., 2016; Miyamoto, Huerta, & Kubacka, 2015; Nielsen, Meilstrup, Nelausen, Koushede, & Holstein, 2015).

This article focuses on the conceptualization and initial validity evidence for a new universal screening assessment that measures SEL competencies consistent with those advanced by the Collaborative for Academic, Social, and Emotional Learning (CASEL, 2012). Specifically, we briefly review universal screening assessments of children's social emotional skills and identify the paucity of measures conceptually aligned with the widely-adopted CASEL SEL competency model. Second, we examine the development and initial psychometric attributes of a universal SEL screening measure called the Social Emotional Learning Assessment[1] (SELA; Elliott, 2016). The SELA was (a) designed with experienced teachers at the conclusion of a multi-year, class-wide social skills intervention program, (b) inspired by the CASEL SEL competencies, and (c) modelled on the SSIS Performance Screening Guide (Elliott & Gresham, 2007), which screens both social and academic behaviors of children.

1. Social emotional learning and social skills: defined and aligned

Social emotional learning has been defined as the "process of acquiring knowledge, skills, attitudes, and beliefs to identify and manage emotions; to care about others; to make good decisions; to behave

* Corresponding author at: Australian Catholic University, Australia.
E-mail address: steve_elliott@asu.edu (S.N. Elliott).

[1] The *Social Emotional Learning Assessment or SELA* was integrated into the SSIS SEL Edition Assessments and renamed the *SSIS SEL Screening and Progress Monitoring Scales* in 2017. This assessment is published by Pearson Assessments and authored by S.N. Elliott & F.M. Gresham.

http://dx.doi.org/10.1016/j.appdev.2017.06.002
Received 19 September 2016; Received in revised form 23 May 2017; Accepted 1 June 2017
Available online 08 June 2017
0193-3973/

ethically and responsibly; to develop positive relationships and to avoid negative behaviors" (Elias & Moceri, 2012, p. 424). Based on this comprehensive definition of SEL, leaders in the Collaborative for Academic, Social, and Emotional Learning (CASEL) advanced a theoretical model of SEL, often referred to as the CASEL five (CASEL, 2012), which includes:

- *Self-Awareness,* defined as the ability to accurately recognize one's emotions and thoughts and their influence on behavior. This includes accurately assessing one's strengths and limitations and possessing a well-grounded sense of confidence and optimism.
- *Self-Management*, defined as the ability to regulate one's emotions, thoughts, and behaviors effectively in different situations. This includes managing stress, controlling impulses, motivating oneself, and setting and working toward achieving personal and academic goals;
- *Social Awareness*, defined as the ability to take the perspective of and empathize with others from diverse backgrounds and cultures, to understand social and ethical norms for behavior, and to recognize family, school, and community resources and supports;
- *Relationship Skills*, defined as the ability to establish and maintain healthy and rewarding relationships with diverse individuals and groups. This includes communicating clearly, listening actively, cooperating, resisting inappropriate social pressure, negotiating conflict constructively, and seeking and offering help when needed; and
- *Responsible Decision-Making Skills,* defined as the ability to make constructive and respectful choices about personal behavior and social interactions based on consideration of ethical standards, safety concerns, social norms, the realistic evaluation of consequences of various actions, and the well-being of self and others.

To date, the CASEL five model (CASEL, 2012) has directly influenced the development of dozens of school based intervention programs in the United States and an entire national curriculum on Personal and Social Capability in Australia. It has not until recently, however, directly influenced the development of assessments of the social, emotional, and academic skills commonly targeted within these programs. The relative paucity of assessments is likely the result of assessment tools being developed and published prior to the recent burgeoning interest in SEL and the disconnect between professionals involved in social behavior assessments and others involved with classroom interventions (Elliott et al., 2015). Interestingly, however, many of the skills representative of the CASEL five have been traditionally part of SSIS social skills assessments and intervention programs that have been in use for decades (Elliott et al., 2015). Numerous definitions of social skills exist, and nearly all describe behaviors that facilitate the initiation and maintenance of positive social relationships, contribute to peer acceptance, allow for individuals to cope with and adapt to the demands of the social environment, and result in satisfactory school adjustment while enabling academics (Gresham, 2002; Gresham & Elliott, 1990, 2008). These behaviors are part of a number of well-regarded assessments and clearly are embedded within many, if not all, the SEL core competencies advanced by CASEL.

2. Assessment of social emotional skills

Several methods exist for assessing children's social emotional skills, including direct observations, interviews, role-plays, and rating scales. Over the past two decades, however, the most frequently used method for assessing social emotional skills has been rating scales (Crowe, Beauchamp, Catroppa, & Anderson, 2011; Humphrey et al., 2011). There are a number of reasons for this rating scale preference. First, rating scales are relatively efficient tools for representing summary characterizations of individuals' observations of other people or their own behavior. As noted by Elliott and Busse (2004), rating scales are imperfect "mirrors" for reflecting images of individuals' social, emotional, and personal functioning; yet, in many cases, the information reflected by a well-constructed rating scale can be very useful. Second, rating scales are relatively easy for teachers, parents, and in many cases students to complete. Third, rating scales have been demonstrated to be more time-efficient and as valid as direct observations for assessing social skills (e.g., Doll & Elliott, 1994; Elliott, Gresham, Freeman, & McCloskey, 1988).

Three comprehensive reviews of measures of social and emotional skills for children and youth have been published (i.e., Crowe et al., 2011; Halle & Darling-Churchill, 2016; Humphrey et al., 2011) over the past five years. All three teams of researchers conducted a comprehensive search of the research literature for measures used in empirical investigations of children's and youth's social behavior. While the Humphrey et al. team identified 189 measures, the Crowe et al. team, using a more restrictive set of search criteria, identified 86 measures all of which also were identified by Humphrey and colleagues. Finally, more recently, Halle and Darling-Churchill (2016), focusing on early childhood (ages 0 to 5 years), identified 75 social emotional measures. None of these reviews, however, included assessments specifically designed for universal screening of children and youth; although short-forms or briefer versions of some the reviewed measures have been used for screening students (e.g., Behavioral and Emotional Screening System from the Behavioral Assessment of Children Scales-2).

A critical review of social-emotional and behavioral screeners was recently published by Jenkins et al. (2014). The Jenkins' team of reviewers identified five common measures, and along with our search of the research literature and publishers websites, two more measures were identified. The resulting seven published SEL screening measures currently available for use are: Behavioral and Emotional Screening System (BESS; Kamphaus & Reynolds, 2007), Behavior Intervention Monitoring Assessment System (BIMAS; McDougal, Bardos, & Meier, 2011), Devereux Student Strengths Assessment (DESSA; LeBuffe, Shapiro, & Naglieri, 2009/2014), Social Academic and Emotional Behavior Risk Screener (SAEBRS; Kilgus & von der Embse, 2015), Social Skills Improvement System Performance Screening Guide (SSIS PSG; Gresham & Elliott, 2008), Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997), and Systematic Screening for Behavior Disorders (SSBD; Walker & Severson, 1992). Each of these screening measures has positive qualities and could be used for efficiently screening large numbers of students (Jenkins et al., 2014). Only one of the assessments, the DESSA, reports to align reasonably well with the CASEL five model and has some psychometric evidence to support its score inferences (Naglieri, LeBuffe, & Shapiro, 2011; Nickerson & Fishman, 2009). As a result, the rationale for the invention of the present study's research problem is the paucity of a psychometrically sound screening measure that is content-aligned with the CASEL SEL competency model. This need, coupled with ongoing SEL research in Australia where a CASEL inspired national curriculum is influencing school and classwide interventions, motivated the invention of the universal screening measure investigated.

3. The present study: research problem and questions

The rationale for the solution of this problem was to design and validate the Social Emotional Learning Assessment[1] (SELA). Given our ongoing social emotional skills intervention programs in Australian schools, we explored the transformation of the highly efficient and reliable SSIS Performance Screening Guide (PSG; Elliott & Gresham, 2007), which was already in use in the schools, into a CASEL aligned screening measure. The SSIS PSG is the class-wide and universal screening component of the SSIS and is an example of a broad-band screening measure that allows for prosocial behavior to be contextualized along with academic skills. It was designed to be used as the first measure in any programmatic intervention effort for children within classes or entire schools to provide a quick, general sense of the

Level	Prosocial Behavior
5	Students at this performance level demonstrate most of the following: • excellent skills to communicate and/or cooperate with others • excellent skills to initiate and sustain conversations/interactions with others • excellent self-control and/or a high level of concern for others Students at this performance level generally do not need additional instruction to improve their social skill level, and their current skill level is considered to be high and/or advanced for their age.
4	Students at this performance level demonstrate most of the following: • a general competence when communicating or cooperating with others • adequate skills to initiate and sustain conversations/interactions with others • adequate self-control and/or an appropriate level of concern for others Students at this performance level may benefit from additional instruction to improve their social skills, but their current skill level is considered to be appropriate for their age.
3	Students at this performance level demonstrate most of the following: • occasional difficulty communicating or cooperating with others • occasional difficulty initiating and sustaining conversations/interactions with others • somewhat less-than-expected self-control or concern for others Students at this performance level are often in need of additional instruction to improve their social skills.
2	Students at this performance level demonstrate most of the following: • frequent difficulty communicating or cooperating with others • frequent difficulty initiating and sustaining conversations/interactions with others • limited self-control or little concern for others Students at this performance level are often in clear need of additional instruction to improve their social skills.
1	Students at this performance level demonstrate most of the following: • very limited communication or cooperation skills • extreme difficulty initiating or sustaining conversations/interactions in an age-appropriate manner • poor self-control or little or no concern for others Students at this performance level often need remedial instruction and/or coaching to improve their social skills.

Fig. 1. The SSIS performance screening guide's prosocial behavior subscale.

status of children's social and academic functioning based on observed interactions with a teacher. As a criterion-referenced, teacher completed measure, the PSG efficiently measures student performance against age- or grade-level expectations in four inter-related areas: prosocial behavior, motivation to learn, reading, and mathematics. Each of the performance areas has a behavior-anchored, 5-level set of performance descriptors that summarizes teachers' observations and interactions with students in their classrooms over several weeks. For each skill area, a classroom teacher chooses the performance level that best represents each student's current level of functioning. In addition to listing the behaviors that define each performance level, the SSIS PSG provides a brief statement about how much intervention, if any, is indicated by each performance level. In general, a performance evaluation of 1 indicates a high level of concern, and intervention is needed immediately, whereas a performance evaluation of 5 indicates no concern and no need for intervention at this time. Fig. 1 provides an illustration of the rubric for Prosocial Behavior within the PSG. The content of this rubric reflects highly valued social skills as rated by teachers and parents on the Importance Scale from the SSIS Rating Scale. Therefore, when students are rated as exhibiting few social skills (Red/Level 1 or Yellow/Levels 2 and 3), there is concern that their social emotional functioning is likely to be below expectations of key adults in their environment.

Research by Miller et al. (2015) provides strong support for the technical and practical soundness of the SSIS PSG. Specifically, these researchers examined nearly 2000 students in grades 1 through 8 from 20 different schools across the country found that the SSIS PSG Prosocial Behavior and Motivation to Learn Scales, respectively, correctly identified 86% to 91% of students at-risk for social, emotional, and behavior problems at three points (fall, winter, spring) during an academic year. This measure was compared to several other screening measures (i.e., Direct Behavior Rating-Single Item Scales; Behavioral

Table 1
Demographic information by disability group[a].

Variable	Non-SWAN n (% of total)	SWAN n (% of total)	Total n (% of total)
Gender			
Male	50 (18.7%)	94 (35.1%)	144 (53.7%)
Female	58 (21.6%)	66 (24.6%)	124 (46.3%)
Grade			
Prep	2 (0.7%)	63 (23.5%)	65 (24.3%)
Year 1	33 (12.3%)	41 (15.3%)	74 (27.6%)
Year 2	42 (15.7%)	30 (11.2%)	72 (26.9%)
Year 3	31 (11.6%)	26 (9.7%)	57 (21.3%)
English as a Second Language			
No	99 (36.9%)	138 (51.5%)	237 (88.4%)
Yes	9 (3.4%)	22 (8.2%)	31 (11.6%)
Aboriginal, Torres Strait Islander			
No	97 (36.2%)	140 (52.2%)	237 (88.4%)
Yes	11 (4.1%)	20 (7.5%)	31 (11.6%)
Total	108 (40.3%)	160 (59.7%)	268 (100%)

[a] *Note*. Students were classified as having a disability (SWAN) if they received specialized assistance due to diagnosed disability or special needs requiring support.

and Emotional Screening System; Student Record Review Form; and Office Discipline Referrals) and overall performed the best psychometrically for the purpose of identifying students at-risk for social emotional difficulties.

Although technically sound and user friendly, the PSG is not well aligned with the CASEL five model or intervention programs designed to teach skills consistent with the five SEL components of this model. Therefore, a need existed for the development of a universal screening measure that more directly assesses children's self-awareness, self-management, social awareness, relationship skills, and responsible decision making.

Recently, Humphrey and Wigelsworth (2016) made the case for just such a universal screener. Specifically, they noted the critical role a school-based screening measure plays in identifying children early on who are in need of mental health services before their problems emerge or grow difficult to address. In the process of making the case for a screener, they identified both the SSIS Rating Scales and the PSG as promising assessments and went on to specify criteria for a high quality screener.

In the present study, we report on the initial and preliminary reliability and validity research regarding the development of a universal screening measure of social emotional learning and academic functioning that is an extension of the SSIS Performance Screening Guide (Elliott & Gresham, 2007) and is based on the CASEL five model. This new screening measure, called the Social Emotional Learning Screening Assessment[1] (SELA), was designed to specifically measure all five of the CASEL SEL Competencies and basic academic skills for purposes of (a) identifying children in need of interventions to improve social emotional learning skills and (b) monitoring progress of children's skill development during and after interventions. The SELA was developed at the end of a multi-year SSIS Classwide Intervention Program (Gresham & Elliott, 2008) with Australian elementary teachers who had used the SSIS PSG for several years. Data from both social behavior and academic measures of students concurrently were collected with the SELA, thus allowing for a comparison with measures of known psychometric evidence.

Seven months after the completion of the first part of the study and with new classes of students, a follow-up data collection with the same teachers in the original study was conducted to establish the preliminary test-retest reliability and the usability of an online version of SELA. Specifically, the teachers assessed new classes of their students with an online version of the SELA on two occasions one month a part and then completed a questionnaire on the usability of the online assessment tool.

The research questions of interest were: (a) How well aligned is the content of the SELA with the CASEL framework for social emotional learning? (b) Is the SELA a usable and feasible universal screening measure of elementary students' social emotional learning skills? (c) Does the SELA yield reliable scores? (d) Can the SELA be used to identify students known to be at-risk for social emotional difficulties? (e) Can the SELA be used to identify students known to be at-risk for academic difficulties? In the remainder of this article, we provide initial preliminary evidence to address these questions.

4. Method

4.1. Sample

The primary participants in this study included 12 teachers who taught prep (ages 4–5) through year 3 (ages 7–8) and 268 of their students. A secondary sample of 266 students in these same teachers' classrooms one school year later participated in a test-retest follow-up study. All participants were all from one suburban school in Queensland, Australia that was involved in a multi-year SSIS Classwide Intervention Program to improve fundamental social skills (see Davies, Cooper, Kettler, & Elliott, 2015). This school served a majority of low income families and a large number of students with additional needs (SWANs).

Table 1 provides demographic data for the entire primary sample of 268 students who had complete data on all measures. This sample of students was 54% male and 46% female, nearly equally distributed across four school years, and was comprised of 160 (61%) students with additional needs (SWAN) and 108 (40%) students without additional needs (Non-SWAN). Students identified as having an additional need received all their instruction in the general education classroom, occasionally with some additional in-class support from special education personnel. The secondary sample of students for the test-retest study was 52% male and 46% female and nearly equally distributed across the same four years of school. Information on their status as a SWAN or Non-SWAN was not made available.

The 12 teachers (83% female and 17% male) were an experienced group who had been at the school for at least three years. Two of the teachers taught Prep level, four taught Year 1, 3 taught Year 2, and three taught Year 3. The teachers also were very familiar with the SSIS assessment measures and Classwide Intervention Program (CIP). All teachers assented to volunteer in this study as part of their evaluation of the CIP. All students were included because the CIP was a standard part of the school program for students in prep to year 3; however, all student data were de-identified to ensure confidentiality.

4.2. Measures

Five measures were used during the school year to document student's social and academic functioning. The SSIS PSG and SSIS Rating Scales were described earlier in detail, and evidence about their psychometric qualities documented. The PAT-Reading and NAPLAN tests are diagnostic and formative assessments, respectively, that are published and secure tests used across Queensland to document all students' achievement. Finally, the SELA is a newly developed screening measure and the primary focus of this report. Key details about each of these measures follow.

4.2.1. SSIS PSG

The Performance Screening Guides measure preschool through early middle school students' skills against grade-level expectations in four performance areas: prosocial skills, motivation to learn, reading, and math. Each of the performance area guides is a criterion-referenced, behaviorally-anchored, multi-level performance descriptor that summarizes several weeks of teachers' observations and interactions with students in their classrooms. An examination of the Prosocial Behavior

subscale's rubric (see Fig. 1) reveals that each of its five levels is comprised of a set of three specific social skills modified by a statement concerning quality or frequency with which the collection of skills is exhibited. In general, the more frequently a student exhibits a behavior, or the more competently the skills are expressed, the higher the level of performance as described by the descriptor. As previously outlined, teachers choose the performance level that best represents each of their students' current levels of functioning in a given performance area. The performance rating ranging from a low of 1 (red band) to a high of 5 (green band) has implications for the importance and type of instructional intervention needed. Students earning a performance rating of 1 in a skill area are considered in need of direct and remedial instructional actions. Students earning a 2 are in the yellow band and require "caution" and additional instruction, teacher attention, and monitoring to ensure that they more consistently use their skills. For this study, we considered students receiving a rating level of 1 or 2 on the Prosocial Scale at-risk socially or on one of the academic scales at-risk academically. Initial findings on the reliability and utility of the PSG were positive. Test-retest reliability estimates ranged from $r = 0.68$ to $r = 0.74$ across skill areas, and inter-observer reliability estimates ranged from $r = 0.55$ to $r = 0.68$ across skill areas. Teachers who used the PSG field test study unanimously agreed that the instructions were clear and easy; information was sufficient; definitions of skill domains were useful; descriptors were clear referred to useful behaviors, and helped sort students into levels; colors and numbers were helpful; and guides were easy to use.

4.2.2. SSIS Rating Scale

The Rating Scales are the second generation of the Social Skills Rating System (SSRS), a frequently researched multi-rater (teacher, parent, student) measure of students' social skills. In this study, only the teacher rating scale was used. Initial arguments for the reliability and validity of the Ratings Scales are based on their foundation in the SSRS. The Rating Scales provide scores based on United States national norms for Prosocial Skills, Problem Behaviors, and Academic Competence. The SSIS - Rating Scales provide a well-rounded picture of children's social behaviors and a brief assessment of academic competence using nationally (USA) standardized behavior rating scales. The Social Skills Scale measures seven positive social behaviors: Communication, Cooperation, Assertion, Responsibility, Empathy, Engagement, and Self-Control. All seven subscales are included across all three rating forms. The Problem Behaviors Scale measures behaviors that can compete or interfere with the performance of prosocial behaviors. It assesses behavior in five subscales: Externalizing, Bullying, Hyperactivity/Inattention, Internalizing, and Autism Spectrum. The Academic Competence Scale provides a quick estimate of overall academic functioning. Teachers rate reading and mathematics performance, general cognitive functioning, and motivation. Substantial psychometric evidence for the validity of SSiS-Rating Scales' scores is provided in its Technical Manual (Gresham & Elliott, 2008).

4.2.3. PAT-reading

The Progressive Achievement Tests in Reading (4th Edition, ACER) is a well-researched and normed test for measuring and tracking student achievement in reading comprehension, word knowledge, and spelling. PAT Reading Comprehension assesses the retrieving of directly stated information, interpreting explicit information, interpreting implied information, and reflecting on texts in multiple-choice format. These skills reflect comprehension skills described in the Australian National Curriculum for English. PAT Vocabulary assesses word knowledge through synonyms using questions that are all multiple choice. There are two PAT spelling tests. Dictated spelling requires the teacher to read a sentence aloud containing a word that students then spell, and the teacher then scores. Written spelling requires students to correct a misspelt word in a written sentence. Detailed diagnostic information is provided for the spelling of each word. The combined PAT-R provides teachers with objective information for setting realistic learning goals and planning effective programs for learning. While data on all three components are collected, for the purposes of this study, the scores for PAT Reading Comprehension were used.

4.2.4. NAPLAN reading tests

The National Assessment Program – Literacy and Numeracy (NAPLAN) is a program of tests of basic skills of Australian students in years 3, 5, 7 and 9 that has been annually administered since 2008. These standardized tests assess students' reading, writing, language conventions (spelling, grammar and punctuation), and numeracy and are administered by the Australian Curriculum, Assessment and Reporting Authority (ACARA). NAPLAN is not a pass or fail type test, but rather, its scores indicate how individual students are performing in numeracy and literacy skills against the national standards for Australian children of the same year level. For this study, we only had access to the Reading test.

4.2.5. SELA

As described earlier, the SELA is a CASEL inspired extension of the SSIS PSG. Like the PSG, it uses a criterion-referenced approach with 5-level performance rubrics. However, in place of the PSG broadband Prosocial Behavior domain (that privileges communication, cooperation, and self-control skills), it measures the five narrower band social emotional domains of Self-Awareness, Self-Management, Social Awareness, Relationship Skills, and Responsible Decision-Making. Like the PSG, it also situates these social emotional skills domains within the context of academic learning and thus continues to assess Motivation to Learn, Reading, and Mathematics performance. Thus, the SELA measures behaviors in eight domains and offers scores (1 = lowest performance level to 5 = highest performance level) for each domain, as well as a Social Emotional (SE) composite score ranging from 5 to 25 and an Academic Functioning (AF) composite score ranging from 3 to 15. For purposes of screening and intervention planning, scores of 1 and 2 on a given rubric are indicative of a high need of attention; students with an SE composite scores of 10 or less are considered at-risk socially, and students with an AF composite of 6 or less are considered at-risk academically. A comparison of the content covered by the SSIS PSG and SELA rubrics against the CASEL competency domains is provided in Fig. 2. Examples of a SELA SE subscale rubric and AF subscale rubric are provided as Figs. 3 and 4, respectively.

The SELA is available in both paper and online versions. The paper version was used initially with teachers in the present study. Eight months later, after an online version was developed, the same teachers used it to conduct a small test-retest reliability study and to document the usability of the assessment.

4.3. Procedures

The primary data collection for this project took place at two points during a school year. The first data collected was within 6 weeks of the beginning of school and included SSIS PSG data for all students in Prep through Year 3 classrooms. For those students who were identified on the PSG at Time 1 as functioning at a Level 1 or 2 in any of the four PSG domains, their teacher followed up by completing the SSIS Rating Scale to gain further information about the students' social skills, problem behaviors, and general academic functioning. Eight weeks later, only students in Year 3 participated in the NAPLAN testing in Numeracy and Literacy.

The second data collection point occurred late in the same school year, nearly seven months after the first data collection. At this Time 2 data point, all students were again assessed by their teachers with the SSIS PSG, and students who were deemed at-risk (Level ratings of 1 or 2) were again re-assessed on the SSIS RS. In addition, at this second data point, students were administered the PAT-Reading Test as required by Queensland Department of Education.

SSIS Performance Screening Guide (PSG) Subscales	Social Emotional Learning Assessment (SELA) Subscales	CASEL Social Emotional Learning Competency Domains
Prosocial Behavior	Self-Awareness	Self-Awareness
	Self-Management	Self-Management
	Social Awareness	Social Awareness
	Relationship Skills	Relationship Skills
	Responsible Decision Making	Responsible Decision Making
Motivation to Learn	Motivation to Learn	
Reading	Reading	
Mathematics	Mathematics	

Fig. 2. Comparison of the SSIS PSG to SELA subscales and CASEL competency domains.

Level 5	A student at this performance level demonstrates most of the following: • Exceptional skills for communicating and/or cooperating with others. • Excellent skills to initiate and sustain conversations/interactions with others. • Excellent self-control and/or high level of concern for others. A student at this performance level generally does not need additional instruction to improve relationship skills. This skill level is high for a student of his/her age.
4	A student at this performance level demonstrates most of the following: • Proficient skills for communicating and/or cooperating with others. • Very good skills to initiate and sustain conversations/interactions with others. • Very good self-control and/or an appropriate level of concern for others. A student at this performance level may benefit from additional instruction to improve relationship skills. This skill level is appropriate for a student of his/her age.
3	A student at this performance level demonstrates most of the following: • Adequate skills for communicating and/or cooperating with others. • Adequate skills to initiate and sustain conversations/interactions with others. • Good self-control and/or an appropriate level of concern for others. A student at this performance level often is in need of additional instruction to improve relationship skills. This skill level is lower than desired for a student of his/her age.
2	A student at this performance level demonstrates most of the following: • Few skills for communicating and/or cooperating with others. • Poor skills to initiate and sustain conversations/interactions with others. • Limited self-control and/or an appropriate level of concern for others. A student at this performance level is in clear need of additional instruction to improve relationship skills. This skill level is significantly lower than desired for a student of his/her age.
1	A student at this performance level demonstrates most of the following: • Very few skills for communicating and/or cooperating with others. • Very poor skills to initiate and sustain conversations/interactions with others. • Very Limited self-control and/or an appropriate level of concern for others. A student at this performance level is in clear need of additional and intensive instruction to improve relationship skills. This skill level is seriously lower than desired for a student of his/her age.

Fig. 3. The SELA relationship subscale.

Level	Motivation to Learn
5	Students at this performance level demonstrate most of the following: • a high level of engagement in almost all instructional activities • a high capacity to stay on task until the activity is completed • a high amount of effort, even when confronted with difficult or complex activities • a high level of attentiveness in almost all situations Students at this performance level generally do not need additional instruction to improve their motivation to learn, and their current level is considered to be high and/or advanced for their age.
4	Students at this performance level demonstrate most of the following: • an adequate level of engagement in a variety of instructional activities • infrequent difficulties staying on task to complete an activity • a general level of effort when work is difficult • an adequate level of attentiveness in a variety of situations Students at this performance level may benefit from additional instruction to improve their motivation to learn, but their current level is considered to be adequate.
3	Students at this performance level demonstrate most of the following: • occasional difficulty engaging in instructional activities • some difficulty staying on task to complete an activity • inconsistent levels of effort when confronted with difficult work • frequent difficulty being attentive in many situations Students at this performance level are often in need of additional instruction to improve their motivation to learn.
2	Students at this performance level demonstrate most of the following: • frequent difficulty engaging in instructional activities • frequent difficulty staying on task to complete an activity • a less-than-expected level of effort when confronted with difficult work • frequent difficulty being attentive in many situations Students at this performance level are often in clear need of additional instruction to improve their motivation to learn.
1	Students at this performance level demonstrate most of the following: • little or no engagement in instructional activities • significant difficulty staying on task to complete an activity • little or no effort when confronted with difficult work • little or no attending skills in most situations Students at this performance level often need incentives to learn and remedial instruction to improve their motivation to learn.

Fig. 4. The SSIS PSG and SELA motivation to learn subscale.

Two weeks after the Time 2 data were collected, teachers completed the SELA on all their students and a questionnaire concerning the usability and feasibility of the SELA for conducting universal screenings of students. Since this data collection closely followed the completion of the social behavior measures, it was considered a concurrent assessment.

The final data collection for the study was conducted seven months later when the same 12 teachers were invited to conduct a test-retest study with a new group of 266 students and to try out an online version of the SELA. This study was done online line during a 1-month period.

4.4. Data analysis

To address the five research questions that motivated this study, a series of descriptive and psychometric analyses were conducted. First, we conducted descriptive statistics with all measures for the total sample, then for various subgroups (boys and girls; SWAN and Non-SWAN), and finally by groups of students per grade. Following these analyses, we conducted a number of correlational analyses. Specifically, we computed coefficient alphas and test-retest correlations for SEL and Academic Functioning rubrics, concurrent correlations with SSIS PSG and Rating Scale subscale scores, and predictive relationships between social behavior measures and measures of students' academic outcomes (i.e., PAT-Reading and NAPLAN scores). Finally, we conducted conditional probability analyses, or ROC analyses, to examine the ability of the PSG and SELA rubrics to predict the likelihood of identifying a student considered to be at-risk socially or academically or identified as a SWAN or Non-SWAN. These latter analyses also provided evidence regarding the potential use of the SELA to predict membership in a risk category and indices regarding its sensitivity (true positives) and specificity (true negatives).

5. Results

5.1. Evidence to support the alignment of SELA with CASEL framework

Two activities were conducted to establish the degree of alignment between the content of the five SELA social emotional rubrics and the

five social emotional domains advanced in the CASEL framework. The first activity involved three social behavior researchers working independently to assign the 46 social skills items from the SSIS Improvement System to a CASEL social emotional skills category. This activity helped to establish sets of objective behaviors likely to be associated with each social emotional domain. The first round of this activity resulted in item assignment agreements of 61%, 65%, and 71% among the three possible pairs of reviewers. For the second round, disagreements were discussed and resulted in a consensus agreement for item assignments of 93% (43 of the 46 items).

The second activity that provided evidence for the alignment of the SELA content with the CASEL social emotional domain was a Q-sort method with six teachers naïve to the purpose of the SELA and to the CASEL model. These teachers were provided definitions of each of the five social emotional domains (as specified in the introduction section of this manuscript) on a separate sheet of paper and 25 randomly ordered slips of paper with each of the 5-level performance descriptors for each of the SELA rubrics without level numbers or colors. The teachers were asked to work independently and to first assign each performance descriptor slip of paper to a social emotional domain and then to order the slips from lowest level to highest level within each domain by taping them to the appropriate sheet of paper. Five of the six teachers independently completed this task 100% correct; the sixth teacher was more ambivalent and slower, and gave up when she saw others had finished the activity. At the conclusion of the activity, the teachers provided feedback about the wording of the descriptors, but had no difficulty with the meaning of the content of the descriptors, nor the relative levels of performance they communicated.

5.2. Usability and feasibility evidence for SELA

The 12 Australian teachers who completed the paper versions of SELA and PSG for their students also completed a Post-Project Evaluation Questionnaire. The results of this questionnaire indicated that these teachers agreed or strongly agreed that the PSG was “easy to use” (93.8%) and it “provided useful information about my entire class” (81.3%). When asked what they liked most about the SSIS assessments, they singled out the PSG and noted the following: “It made me look at each of my students closely,” “It gives me a clear snapshot of my entire classes' strengths and weaknesses,” “It is easy to use and time efficient,” “It nicely covers both social skills and academic areas and connects them,” and finally “The PSG's red-yellow-green color signals are helpful communicators about risk.” Because the Post-Project Evaluation Questionnaire did not directly ask questions about the SELA, teachers were asked during a focus group session to express their reactions to the measure. Ten of the 12 teachers with several years of experience using the PSG and SSIS Rating Scales expressed strong support for SELA and indicated they would use it as a screening measure and a progress monitor if it were available to them. Two teachers indicated the SELA was more detailed than needed for their purposes and stated a preference for the PSG.

The same 12 teachers, eight months later, completed the initial online version of the SELA as part of the test-retest reliability study. After using the online version, all 12 teachers indicated it was easy to access and navigate, and equally easy to use the electronic version of the SELA in comparison to the paper version. There was no missing data with the online version and scoring was more efficient.

5.3. Descriptive statistics and evidence regarding the relationships among teacher-completed social behavior measures

Basic descriptive statistics for all measures are displayed in Tables 2 through 6. The data sets were 98.2% complete with missing items determined to be missing completely at random, so no special treatment was needed for the missing data. An inspection of Table 2 indicates that teachers used all levels of the 5-level SELA rubrics to describe their students' social emotional skills and academic functioning. It also suggests that the three AF rubrics common to both the SELA and the PSG (completed between 2 and 3 weeks apart in time) functioned similarly, although the SELA rubrics were rated slightly lower. The SELA SE rubrics (range from 2.96 to 3.43) also were consistently rated lower than the more broadband Prosocial rubric (3.6) of the PSG. It should be noted that based on a school decision to respect teachers' time, the SSIS Rating Scale scores are only for the subset of 55 students deemed to be at-risk either socially or academically. The mean standardized scores (mean = 100; SD = 15) for each of the SSIS Rating Scale subscales clearly indicated that these students were, on average, functioning 1 SD below average for Social Skills and 1 SD above average on Problem Behaviors and nearly 2 SDs below average academically.

Table 2
Means and standard deviations on all measures for total sample.

Subscale	Mean	SD
PSG: prosocial	3.60	1.13
PSG: motivation	3.60	1.25
PSG: reading	3.31	1.25
PSG: math	3.13	1.19
RS: social skills	84.02	11.26
RS: problem behaviors	121.6	15.89
RS: academic competence	75.89	9.35
SELA: self-awareness	2.96	1.13
SELA: self-management	3.10	1.24
SELA: social awareness	3.32	1.12
SELA: relationship skills	3.14	1.12
SELA: decision making skills	3.43	1.20
SELA: motivation to learn	3.19	1.27
SELA: reading	3.01	1.24
SELA: mathematics	2.97	1.15
PAT: reading	86.87	19.28
NAPLAN: reading	375.80	73.35

Note. PSG and SELA $n = 268$; SSIS-RS $n = 55$ which were all students at-risk socially; SSIS academic competence $n = 44$ which were all students at-risk academically; PAT reading $n = 245$; NAPLAN reading $n = 50$ which were only year 3 students.

For a more refined understanding of the students' ratings on the SELA for the entire sample, examine Tables 3 for score means (standard deviations) of female and male students and Table 4 for developmental differences in scores across prep (average age 5) through year 3 (average age 8). The data for the different gender groups were as expected based on previous social skills studies; that is, girls were consistently rated higher than boys on nearly all social and academic measures (Gresham & Elliott, 2008). A MANOVA with the dependent variables of SE Composite and AF Composite scores was conducted and confirmed that girls were rated statistically higher than boys on these two summary indices and all subsequent univariate analysis of seven of the eight specific rubrics, with the exception being the Mathematics rubric.

The developmental data (in Table 4) represent a 4-year age/grade span. For all eight SELA rubrics and the associated SE and AF Composite scores, there is a clear progression from prep year to year 3 with average scores increasing by approximately 1/2 a standard deviation with the exception for the Motivation to Learn rubric.

Table 5 presents a comprehensive correlation matrix for all the teacher-completed social behavior measures in the study. As documented, all of the SELA rubrics correlate moderately to highly with the PSG rubrics. Of particular interest are the correlations between the PSG Prosocial rubric and the five SELA social emotional rubrics where the correlations are all statistically significant and range from a low of 0.63 to a high of 0.76. These same SELA rubrics, however, correlated substantially lower with the SSIS Rating Scale Social Skill subscale, with only the Social Awareness, Relationship Skills, and Decision Making rubrics being correlated significantly with the Social Skills subscale.

The three SELA academic functioning rubrics correlated highly (range 0.83 to 0.85) with each of their identical PSG academic

Table 3
Means and standard deviations for SELA subscales: total sample and females/males.

SELA scale and subscales	Total sample $N = 268$	Female subsample $N = 124$		Male subsample $N = 144$
Self–awareness mean(SD)	2.96 (1.13)	3.17 (1.02)	> .004	2.78 (1.19)
Self–management mean (SD)	3.10 (1.24)	3.41 (1.13)	> .000	2.83 (1.28)
Social awareness mean (SD)	3.32 (1.12)	3.62 (0.99)	> .000	3.06 (1.17)
Relationship skills mean (SD)	3.14 (1.12)	3.35 (1.05)	> .004	2.96 (1.15)
Decision–making skills mean (SD)	3.43 (1.20)	3.65 (1.12)	> .006	3.24 (1.25)
SE composite mean (SD)	15.94 (5.03)	17.19 (4.63)	> .000	14.87 (5.13)
Motivation to learn mean (SD)	3.19 (1.27)	3.48 (1.22)	> .000	2.95 (1.26)
Reading skills mean (SD)	3.01 (1.24)	3.17 (1.18)	> .053	2.88(1.28)
Mathematics skills mean (SD)	2.97 (1.15)	2.99 (1.06)	=	2.94 (1.22)
AF composite mean (SD)	9.17 (3.34)	9.64 (3.15)	> .034	8.77 (3.46)

Note. The direction and probability of the difference in mean rating scores for Female to Male is indicate in each

functioning rubric. Both the SELA Reading and PSG Reading rubrics correlated highly with the direct measures of reading provided to students. Specifically, the SELA reading correlated 0.64 with NAPLAN Language Arts (which includes writing) and 0.68 with PAT reading. The SSIS PSG correlated 0.63 with NAPLAN Language Arts and 0.63 with PAT reading. These SELA academic rubrics also correlated moderately (range 0.57 to 0.69) with the PSG prosocial rubric.

5.4. Evidence for the reliability of SELA rating scores

The initial reliability estimates for the SELA are based on two indices, Cronbach's alphas (internal consistency) and test-retest correlations (see Table 6). Specifically, for the entire 8 rubrics with the initial sample of 12 teachers and 268 students, the alpha was 0.93, and for the SE Composite of the five social emotional rubrics and AF Composite of three academic rubrics, the alphas were 0.91 and 0.90, respectively. The test-retest reliability estimates from the follow-up study sample of the same12 teachers with a different sample of 266 students indicated the SE composite (0.89) and AF composite (0.91) ratings were highly consistent over a one-month period. The SE subscale rubric ratings ranged from 0.70 to 0.87 and the AF subscale rubric ratings ranged from 0.84 to 0.87. All the specific test-retest ratings were statistically significant (see Table 6).

Using the coefficient alphas as our reliability estimate, we calculated standard error of measurement (SEM) for each of the subscales. The resulting SEMs were all very similar and ranged from a low of 0.34 for Social Awareness and Relationship Skills subscales to a high of 0.41 for the Motivation to Learn subscale. This relatively small SEM suggests scores for the various rubric ratings can be considered to have very good precision.

5.5. Evidence for the utility of SELA for identifying students at-risk

To examine how well the SELA discriminated between students at-risk for social emotional difficulties, academic difficulties, and with additional needs, a series of Receiver Operator Characteristics (ROC) analyses were conducted. Before running the analyses, we standardized the SELA Social Composite Score (transformed raw scores to z scores; mean for the SELA Social Composite was 15.91 with a SD of 5.03 and the mean for the SELA Academic Composite was 9.11 with a SD of 3.37). We first conducted Area Under the Curve (AUC) analysis. Specifically, the area under the curve is the percentage of randomly drawn pairs for which the test correctly classifies students. The accuracy of this test depended on how well the rating rubric separated the group of children into those with and without social emotion difficulties. Accuracy is indicated by the area under the ROC curve with a result of 1 for a perfect test and an area of 0.5 a worthless test. Table 7 provides a summary of the AUC results for all the SELA subscales for predicting the Socially At-Risk group (as well as Academically At-Risk groups in Reading and Mathematics, and Students with Additional Needs (SWAN)). All the discrimination percentages for individual SELA performance rubrics and the SE and AF composites were very high (ranges from 0.83 to 0.92) for the Socially At-Risk group and moderately high for the Academically At-Risk and SWAN groups.

Following the AUC analysis, sensitivity (true positive) and specificity (true negative) percentages were calculated (see Table 8). When we compared at-risk on the PSG Prosocial subscale (which the school was using to identify socially at-risk students) and the SELA SE Composite scores, we accurately identified 60.5% of the students as true positives and 92.4% as true negatives. Coincidentally, the associated positive predictive power and negative predictive power for these social at-risk predictions were also 60.5% and 92.4%, respectively. For the comparison of the PSG Reading and the SELA, AF Composite scores we accurately identified 86.7% of the students as true positives and 91% as true negatives. The associated positive predictive power and negative predictive power for these reading at-risk predictions were 86.7% and 91.0%, respectively. Finally, when we used the PSG Mathematics and SELA AF Composite scores, we accurately identified 51.4% of students as true positives and 96.4% as true negatives. The associated positive predictive power and negative predictive power for these mathematics at-risk predictions were 84.5% and 83.9%, respectively.

Table 4
Means and standard deviations for SELA subscales: prep to Year 3 students.

SELA scale and subscales	Prep year students $N = 65$	Year 1 students $N = 74$	Year 2 students $N = 72$	Year 3 students $N = 57$
Self-awareness				
Total mean (SD)	2.63 (1.11)	2.76 (1.00)	3.10 (1.19)	3.42 (1.05)
Female	2.76 (0.99)	2.87 (0.81)	3.53 (1.00)	3.46 (1.07)
Male	2.53 (1.21)	2.67 (1.13)	2.67 (1.22)	3.38 (1.05)
Self-management				
Total mean (SD)	2.89 (1.16)	3.07 (1.11)	3.08 (1.45)	3.39 (1.19)
Female	2.97 (1.09)	3.23 (0.93)	3.72 (1.21)	3.64 (1.13)
Male	2.83 (1.23)	2.93 (1.22)	2.44 (1.40)	3.14 (1.22)
Social awareness				
Total mean (SD)	3.17 (1.17)	3.18 (1.04)	3.33 (1.32)	3.67 (0.81)
Female	3.31 (1.07)	3.32 (0.98)	4.00 (0.99)	3.79 (0.74)
Male	3.06 (1.24)	3.07 (1.08)	2.67 (1.29)	3.55 (0.87)
Relationship skills				
Total mean (SD)	2.89 (1.11)	3.03 (1.05)	3.18 (1.09)	3.51 (1.18)
Female	3.07 (1.03)	3.26 (1.09)	3.56 (0.94)	3.46 (1.14)
Male	2.75 (1.16)	2.86 (0.99)	2.81 (1.12)	3.56 (1.24)
Decision-making				
Total mean (SD)	3.12 (1.19)	3.37 (1.25)	3.61 (1.28)	3.63 (0.99)
Female	3.28 (1.00)	3.55 (1.18)	3.94 (1.22)	3.75 (0.97)
Male	3.00 (1.33)	3.23 (1.29)	3.28 (1.28)	3.52 (1.02)
SE composite				
Total mean (SD)	14.71 (5.08)	15.39 (4.38)	16.31 (5.52)	17.61 (4.73)
Female	15.38 (4.79)	16.26 (4.00)	18.75 (4.49)	18.11 (4.63)
Male	14.17 (5.30)	14.77 (4.58)	13.86 (5.43)	17.14 (4.85)
Motivation to learn				
Total mean (SD)	3.19 (1.25)	3.23 (1.21)	3.00 (1.34)	3.40 (1.27)
Female	3.38 (1.15)	3.39 (1.11)	3.56 (1.32)	3.57 (1.32)
Male	3.03 (1.32)	3.12 (1.28)	2.44 (1.13)	3.24 (1.21)
Reading skills				
Total mean (SD)	2.62 (1.25)	2.99 (1.27)	3.11 (1.13)	3.37 (1.25)
Female	2.79 (1.26)	3.23 (1.09)	3.28 (1.06)	3.36 (1.31)
Male	2.47 (1.23)	2.81 (1.37)	2.94 (1.19)	3.38 (1.21)
Mathematics skills				
Total mean (SD)	2.77 (1.10)	2.89 (1.15)	2.89 (1.04)	3.39 (1.24)
Female	2.90 (1.08)	2.87 (0.96)	2.97 (0.97)	3.25 (1.24)
Male	2.67 (1.12)	2.91 (1.29)	2.81 (1.12)	3.52 (1.24)
AF composite				
Total mean (SD)	8.57 (3.33)	9.11 (3.41)	9.00 (3.07)	10.16 (3.47)
Female	9.07 (3.23)	9.48 (2.91)	9.81 (2.97)	10.18 (3.59)
Male	8.17 (3.40)	8.84 (3.74)	8.19 (2.98)	10.14 (3.42)

6. Discussion

This study was a preliminary investigation to examine practical and technical aspects of the SELA, a new measure for the purpose of universal of screening elementary children with social emotional learning difficulties. The study was part of the evaluation of an authentic social behavior screening and intervention project in a high needs school in Australia. The CASEL model of social emotional learning competencies influenced this school via the national curriculum expectations characterized in the ACARA Personal and Social Capabilities for Australian children. Thus, the SELA was created to fill a measurement need for a fully-aligned measure with the behavior expectations of the influential CASEL five model of social emotional learning.

Using the multi-level rubric approach established with the SSIS Performance Screening Guide (Elliott & Gresham, 2007), we utilized detailed descriptions of each of the five CASEL SEL skill sets to formulate performance rubrics for assessing children's self-awareness, self-management, social awareness, relationship, and responsible problem solving skills. We also contextualized these SEL skills along with key academic skills - Motivation to Learn, Reading, and Mathematics - from the PSG assessment to develop a comprehensive screening assessment. This new measure was pilot tested with experienced teachers in Australia. Based on an array of qualitative and quantitative results, we found promising preliminary evidence to support continued research of the SELA as a practical and efficient measure for reliably identifying children in need of social emotional and academically related intervention services.

6.1. Major findings

Validation of the scores and use of any measure is an ongoing process and requires substantial evidence regarding the nature of the construct being measured, the reliability or precision of the measurement, and the consequences of using the measure (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). Five outcomes were observed from the initial SELA evidence resulting from the screening of several hundred Australian children regarding their social emotional wellbeing.

The first important outcome of interest concerned the alignment of the content measured on the SELA with the content of the CASEL five model. Given the specific definitions of each of the CASEL SEL skill domains were used to guide the writing of the SELA rubrics, it was not surprising to find experienced social skills researchers and teachers could assign skills and align skill description levels to the CASEL skill definitions with a very high degree of agreement. These sources of evidence suggest the SEL skills measured by the SELA are collectively a valid sample and representation of the skills and abilities advanced by the CASEL five model.

The second important evidence-based outcome came in the form of the patterns of scores for male and female students across the four years of schooling examined. Specifically, research by Gresham and Elliott (1990, 2008) and Gresham, Elliott, and Kettler (2010) has documented that girls consistently are rated higher than boys with regard to their frequency of desired social behaviors on traditional rating scale measures, such as the SSIS. Our results suggest this same sensitivity to gender differences is possible with the brief multi-level SELA performance rubric. It also indicated, unexpectedly, that this measure was very sensitive to developmental level differences in children and is even more sensitive than other rating scales. For example, most rating scales have grade or age cluster scores, rather than scores for each grade and year of age or school. In summary, our initial evidence with the SELA suggests it can be used to yield scores that are sensitive to students' gender and developmental level, an attribute that most screening measures, and many behavior rating scales, do not currently provide.

The third essential outcome of the initial research with SELA is that it has the potential to yield reliable, precise scores. Although reliability or precision evidence was limited to very high Cronbach alphas of internal consistency and high test-retest correlations for SE and AF composite and subscale scores over a 1-month period, the results indicated teachers' ratings generate scores that can be characterized as highly reliable with small standard errors of measurement. More reliability research with larger and more diverse samples, however, is needed to affirm this initial finding.

A major finding and fourth outcome of the study was that the SELA accurately discriminated students who were at-risk for social and academic difficulties from those who were not at-risk. At-risk was defined by previous ratings on the SSIS PSG, a similar type of measure to the SELA and a measure being used by schools in Australia and the United States to efficiently screen large numbers of students for universal intervention programs. The ROC analyses Area Under the Curve results and the sensitivity (true positives) and specificity (true negatives) indices suggested the SELA does a substantially better job than chance of identifying students at-risk. This result was very similar in magnitude to the findings for the SSIS PSG reported by Miller et al. (2015) with a large sample of U.S. students in grades 1 to 8. Collectively, the AUC, specificity, and sensitivity indices, along with positive and negative predictive power indices, met or exceeded widely accepted criteria for a well-functioning screening measure (Kettler, Glover, Albers, & Feeney-

Table 5
Correlations among teacher-completed social behavior measures.

SELA subscales	PSG prosocial	PSG motivation	PSG reading	PSG math	SSIS RS social skills[a]	SSIS RS problem behavior[a]	SSIS RS academic[a]
Self-awareness							
r	0.626	0.601	0.548	0.511	0.184	− 0.172	0.249
p	0.000	0.000	0.000	0.000	0.179	0.210	0.104
Self-management							
r	0.634	0.603	0.514	0.442	0.140	− 0.174	− 0.055
p	0.000	0.000	0.000	0.000	0.310	0.204	0.724
Social awareness							
r	0.626	0.549	0.497	0.479	0.310	− 0.215	0.019
p	0.000	0.000	0.000	0.000	0.021	0.115	0.904
Relationship skills							
r	0.626	0.584	0.547	0.541	0.268	− 0.049	0.108
p	0.000	0.000	0.000	0.000	0.048	0.722	0.484
Decision making							
r	0.757	0.689	0.638	0.579	0.501	− 0.510	0.194
p	0.000	0.000	0.000	0.000	0.000	0.000	0.206
Motivation							
r	0.685	0.830	0.706	0.730	0.217	− 0.364	0.462
p	0.000	0.000	0.000	0.000	0.111	0.006	0.002
Reading							
r	0.612	0.672	0.846	0.772	0.293	− 0.229	0.743
p	0.000	0.000	0.000	0.000	0.030	0.093	0.000
Mathematics							
r	0.573	0.686	0.799	0.847	0.319	− 0.289	0.877
p	0.000	0.000	0.000	0.000	0.018	0.033	0.000

[a] $n = 55$.

Table 6
SELA subscale means, standard deviations, and test-retest correlations.

SELA subscale	Testing time 1		Testing time 2		T1 & T2 scores
	M	SD	M	SD	Correlation
Self-awareness	3.29	0.97	3.55	0.98	0.704[a]
Self-management	3.23	1.15	3.44	1.10	0.759[a]
Social awareness	3.65	1.04	3.83	1.04	0.788[a]
Relationship skills	3.41	1.11	3.62	1.05	0.796[a]
Decision making	3.60	1.07	3.78	1.08	0.872[a]
SE composite	17.16	4.89	18.23	4.67	0.887[a]
Motivation to learn	3.46	1.25	3.65	1.15	0.857[a]
Reading	3.10	1.25	3.33	1.19	0.870[a]
Mathematics	3.05	1.17	3.29	1.17	0.835[a]
AF composite	9.61	3.11	10.27	3.04	0.910[a]

Note. These data are based on 12 teachers' ratings of 266 students in grades Prep through 3 1-month apart. These students were a different, but overlapping, sample of students rated in the rest of the study.

[a] Correlation is significant at the 0.01 level (2-tailed test).

Table 7
SELA ROC area under the curve summary for predictions of risk.

SELA scale	Socially at-risk	Academic at-risk		Students with additional needs
		Reading	Math	
Self-awareness	0.83	0.75	0.72	0.70
Self-management	0.83	0.71	0.70	0.69
Social awareness	0.82	0.72	0.70	0.66
Relationship skills	0.84	0.76	0.74	0.68
Decision making	0.86	0.76	0.74	0.65
Motivation to learn	0.88	0.81	0.81	0.65
Reading	0.88	0.90	0.88	0.73
Mathematics	0.89	0.88	0.88	0.69
Social composite	0.90	0.78	0.76	0.70
Academic composite	0.92	0.91	0.90	0.71

Kettler, 2014). The SELA SE Composite score, not surprisingly, was the most effective for decisions about social emotional skills; however, the SELA AF Composite score by comparison produced even better sensitivity and specificity results for identifying students at-risk academically. In summary, the results show the SELA to have promise as a screening measure for the purpose of identifying students in need of additional services for social emotional and academic difficulties.

Table 8
Screening efficiency metrics for students at-risk.

Student population	Sensitivity	Specificity	Positive predictive power	Negative predictive power
Socially at-risk	60.47%	92.44%	60.47%	92.44%
Academically at-risk				
Reading	66.10%	97.13%	86.67%	91.03%
Math	51.35%	96.39%	84.44%	83.86%

The fifth and final notable outcome from the evidence collected was that nearly all teachers found both the paper and online versions of the SELA to be easy to use, time efficient, and of relevant to their efforts to improve children's social skills. It should be noted, that these results followed their use of other screening and behavior rating scales, so these qualitative comments are likely to be well grounded in experience. Future examination of the utility and feasibility of the SELA is needed, in particular, after it is used and a report of results is provided.

6.2. Implications for practice

The present findings, if replicated with more diverse and larger samples of teachers and students, would help establish the use of an efficient screening method that offers relevant information about the status of the students' social emotional functioning. This type of information characterized within the CASEL competency framework is valued by educators and could be used to drive universal or targeted interventions for a large number of students. Subsequent research is needed, but it is also likely the SELA given it psychometric and usability qualities could be used to monitor the effectiveness of interventions designed to improve SEL competencies like those advanced by CASEL.

6.3. Limitations and future research

Although the initial data for the SELA indicated that it is a promising screening tool for use by teachers, the data are limited in several ways that requires replication before generalized use can be recommended. The first and perhaps greatest limitation concerned the sample. Specifically, a larger sample of teachers and a more ethnically diverse sample with students ranging in age from preschool to grades 7 or 8 would greatly enhance the database and provide a stronger test of the existing psychometric claims. Another limitation concerned the design where the SELA was not available for use until the end of the project, thus creating a question regarding order-effects and negating opportunities to sample predictive powers against other widely used social emotional and academic screening measures. A final potential concern, if not limitation, is that the teachers in this study knew their students' social behavior functioning very well given they had all participated in a 10 to 12 week Classwide Intervention Program (CIP; Gresham & Elliott, 2008). The CIP involved pre- and post-assessments of all students. Thus, this level of knowledge about students' discrete social emotional skills is greater than most teachers would possess prior to a typical universal screening process and may have influenced the quality of data for making ratings on the SELA. In summary, the present findings about the SELA, although promising, are representative of only initial data in the typical test development process. With replication, more diverse samples, and designs that allow for repeated measures concurrent with other known measures of social emotional functioning such as the DESSA, the psychometric evidence needed to refine our understanding of SELA and its effective uses would be significantly advanced.

7. Conclusions

This study was motivated by the need for a high-quality universal screening assessment for use with children with potential social emotional difficulties and to address fundamental questions about the SELA and its use by teachers to screen all their students. The resulting data provided a basis for positive preliminary answers to fundamental questions about the usability and technical soundness of a CASEL inspired, competency-aligned assessment. Specifically, with regard to our first question of "How well aligned is the content of the SELA with the CASEL framework for social emotional learning?", we found both experienced researchers and teachers provided evidence that the content of the five SELA social emotional rubrics was consistent with the descriptions of the corresponding CASEL five competencies. Thus, it seems reasonable to assert that SELA is a content valid measure of social emotional learning skills theorized as important to the development of children and youth. Because of the strong alignment between CASEL five competencies and the components of the Personal and Social capability of the Australian Curriculum, it also can be claimed that the SELA provides a content valid measure for Australian teachers wishing to screen their students on their Personal and Social capabilities, and to repeatedly monitor student progress on these capabilities.

With regard to our second question, which asked, "Is the SELA a usable and feasible universal screening measure of elementary students' social emotional learning skills?", we found clear evidence from teacher reports that the paper and online versions of SELA are easy and time efficient to complete in authentic school situations with very little training support. That is, most of the teachers completed the SELA for an entire class of students in < 25 min.

For our third question, we asked, "Does the SELA yield reliable scores?" Although our evidence was limited to estimates of internal consistency and test-retest reliability with elementary teachers and their students, we found that these estimates were all in the high to very high range and resulted in low standard errors of measurement for composite and subscale scores. This clearly suggested that there is relatively little error around any SE or AF subscale ratings. Offering test score precision is highly valued when making screening and intervention effectiveness decisions.

Our final two questions concerned whether the SELA could be meaningfully used to improve the identification of children at-risk for social or academic difficulties. Our specific questions were: "Can the SELA be used to identify students known to be at-risk for social emotional difficulties?" and "Can the SELA be used to identify students known to be at-risk for academic difficulties?" The evidence from our analysis was strong when risk status was determined by a previous SSIS PSG assessment, indicating that the SELA does a very good job relative to chance at identifying students in the early school years who are at-risk and in need of additional services for social emotional and/or academic difficulties.

In summary, the initial research of the practical and technical characteristics of SELA indicates it is a promising measure for elementary teachers for screening students for social emotional universal intervention programs. More research remains, however, to replicate and extend these findings to educators in other Australian schools and to the US and test its application with more diverse samples of students.

References

ACARA (2013). Personal and social capabilities. http://www.australiancurriculum.edu.au/GeneralCapabilities/Personal-and-social-capability/Introduction/Introduction.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington DC: American Educational Research Association.

Australian Council of Educational Research (ACER) (2014). *ACER Progressive Achievement Tests (PAT). 2013 National Update*. (ttps://www.acer.edu.au/files/PAT_Report_2013.pdf).

Collaborative for Academic, Social, and Emotional Learning. (2012). *2013 CASEL guide: Effective social and emotional learning programs—Preschool and elementary* (school ed.). Chicago, IL: Author.

Crowe, L. M., Beauchamp, M. H., Catroppa, C., & Anderson, V. (2011). Social function assessment tools for children and adolescents: A systematic review from 1988 to 2010. *Clinical Psychology Review*, *31*, 767–785.

Davies, M., Cooper, G., Kettler, R., & Elliott, S. N. (2015). Developing social skills to improve academics and classroom inclusion within the context of the Australian curriculum. *Australasian Journal of Special Education*, *39*(1), 37–55.

DiPerna, J. C., Volpe, R., & Elliott, S. N. (2002). A model of academic enablers and elementary reading/language arts achievement. *School Psychology Review*, *31*(3), 298–312.

Doll, E., & Elliott, S. N. (1994). Consistency of observations of preschoolers' social behavior. *Journal of Early Intervention*, *18*(2), 227–238.

Elias, M. J., & Moceri, D. (2012). Developing social and emotional aspects of learning: The American experience. *Research Papers in Education*, *27*(4), 423–434.

Elliott, S. N. (2016). *Social emotional learning assessment*. Scottsdale, AZ: Author.

Elliott, S. N., & Busse, R. T. (2004). Assessment and evaluation of students' behavior and intervention outcomes: The utility of rating scale methods. In R. B. Rutherford, M. M. Quinn, & S. R. Mathur (Vol. Eds.), *Handbook of research in behavioral disorders. Chapter 7. Handbook of research in behavioral disorders* (pp. 123–142). New York: Guilford Publications.

Elliott, S. N., Frey, J. R., & Davies, M. (2015). Systems for assessing and improving students' social skills to achieve academic competence. In J. Durlak, C. Domitrovich, R. Weissberg, & T. Gullotta (Eds.), *Handbook of social & emotional learning: Research and practice* (pp. 301–319). New York: Guilford Press978-1-4625-2015-2.

Elliott, S. N., & Gresham, F. M. (2007). *SSIS classwide intervention program Teacher's guide*. Minneapolis, MN: Pearson Assessment.

Elliott, S. N., Gresham, F. M., Freeman, T., & McCloskey, G. (1988). Teachers' and observers'ratings of children's social skills: Validation of the Social Skills Rating Scale. *Journal of Psychoeducational Assessment*, *6*, 152–161.

Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology, Psychiatry, and Allied Disciplines*, *38*(5), 581–586.

Gresham, F. M. (2002). Teaching social skills to high-risk children and youth: Preventive and remedial approaches. In M. Shinn, H. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches* (pp. 403–432). Bethesda, MD: National Association of School Psychologists.

Gresham, F. M., & Elliott, S. N. (1990). *Social skills rating system*. Circle Pines, MN: American Guidance Service.

Gresham, F. M., & Elliott, S. N. (2008). *Social skills improvement system (SSIS)*. Minneapolis, MN: Pearson Assessment.

Gresham, F. M., Elliott, S. N., & Kettler, R. J. (2010). Base rates of social skills acquisition/performance deficits, strengths, and problem behaviors: An analysis of the Social Skills Improvement System-Rating Scales. *Psychological Assessment*, *22*(4), 809–815.

Halle, T. G., & Darling-Churchill, K. E. (2016). *Review of measures of social and emotional development*[Journal of Applied Developmental Psychology]. http://dx.doi.org/10.1016/j.appdev.2016.02.003.

Humphrey, N., Kalambouka, A., Wigelsworth, M., Lendrum, A., Deighton, J., & Wolpert, M. (2011). Measures of social and emotional skills for children and young people: A

systematic review. *Educational and Psychological Measurement*, *71*(4), 617–637.
Humphrey, N., & Wigelsworth, M. (2016). Making the case for universal school-based mental health screening. *Emotional and Behavioral Difficulties*, *21*(1), 22–42.
Jenkins, L. N., Demaray, M. K., Wren, N. S., Secord, S. M., Lyell, K. M., Magers, A. M., ... Tennant, J. (2014). A critical review of five commonly used social-emotional and behavioral screeners for elementary or secondary schools. *Contemporary School Psychology*. http://dx.doi.org/10.1007/s40688-014-0026-6.
Kamphaus, R. W., & Reynolds, C. R. (2007). *BASC-2 Behavioral and Emotional Screening System manual*. Bloomington, MN: Pearson.
Kettler, R. J., Glover, T. A., Albers, C. A., & Feeney-Kettler, K. A. (2014). An introduction to universal screening in educational settings. In R. J. Kettler, T. A. Glover, C. A. Albers, & K. A. Feeney-Kettler (Eds.), *Universal screening in educational settings: Evidence-based decision making for schools*. Washington, D.C.: American Psychological Association.
Kilgus, S. P., & von der Embse, N. P. (2015). *Social, Academic, and Emotional Behavior Risk Screener (SAEBRS)*. Minneapolis, MN: Theodore J. Christ & Colleagues.
LeBuffe, P. A., Shapiro, V. B., & Naglieri, J. A. (2009/2014). *Devereux Student Strengths Assessment*. Lewisville, NC: Kaplan.
McDougal, J. L., Bardos, A. N., & Meier, S. T. (2011). *Behavior Intervention Monitoring Assessment System technical manual*. Multi-Health Stystems: Toronto.
Miller, F. G., Cohen, D., Chafouleas, S. M., Riley-Tillman, T. C., Welsh, M. E., & Fabiano, G. A. (2015). A comparison of measures to screen for social, emotional, and behavioral risk. *School Psychology Quarterly*, *30*(2), 184–196.
Miyamoto, K., Huerta, M. C., & Kubacka, K. (2015). Fostering social and emotional skills for well-being and social progress. *European Journal of Education*, *50*(2), 147–159.
Naglieri, J. A., LeBuffe, P., & Shapiro, V. B. (2011). Universal screening for social-emotional competencies: A study of the reliability and validity of the DESSA-mini. *Psychology in the Schools*, *48*(7), 660–671.
Nickerson, A. B., & Fishman, C. (2009). Convergent and divergent validity of the Devereux Student Strengths Assessment. *School Psychology Quarterly*, *24*(1), 48–59.
Nielsen, L., Meilstrup, C., Nelausen, M. K., Koushede, V., & Holstein, B. E. (2015). Promotion of social and emotional competence: Experiences from a mental health intervention applying a whole school approach. *Health Education*, *115*(3/4), 339–356.
Walker, H. M., & Severson, H. (1992). *Systematic screening for behavior disorders: Technical manual*. Austin, TX: Pro-Ed.
Weissberg, R. P., & Cascarino, J. (2013). Academic learning + social-emotional learning = national priority. *Phi Delta Kappan*, *95*(2), 8–13.