Validation of the Elementary Social Behaviour Assessment: Teacher Ratings of Students' Social Skills Adapted to Norwegian, Grades 1-6

Anne Arnesen

Department of Special Needs Education, University of Oslo, Norway

Keith Smolkowski

Oregon Research Institute, Eugene, OR, USA

Terje Ogden

Norwegian Center for Child Behavior Development, Oslo, Norway

Monica Melby-Lervåg

Department of Special Needs Education, University of Oslo, Norway

Abstract

This study investigated the psychometric properties of the Elementary Social Behavior Assessment (ESBA), a teacher-report measure of students' social skills and academic engagement adapted for use in Norwegian schools. At two times, 8 weeks apart, 151 teachers rated 793 students in Grades 1-6 on the ESBA and the Social Skills Rating System Teacher's form (SSRS-T). Exploratory factor analysis with the first assessment suggested one or possibly two factors. Confirmatory factor analysis with the second assessment confirmed the one- or two-factor models but did not offer strong evidence for one over the other. The ESBA demonstrated good to excellent reliability. Strong concurrent and predictive correlations with the SSRS-T established criterion validity. The school and teacher levels explained only a limited amount of the total variance in ESBA scores, suggesting that teachers rate students quite similarly. The Norwegian-adapted ESBA paralleled the reliability and validity evidence from US samples.

Keywords: screeners, social behaviour, academic engagement, reliability, validity

Validation of the Elementary Social Behavior Assessment: Teacher Ratings of Students'

Social Skills Adapted to Norwegian, Grades 1 to 6

Most students behave as expected with social skills that contribute to positive learning processes throughout their time in school. For instance, listening to the teacher's instructions and following directions, working with effort, asking for assistance in an appropriate manner, getting along with peers, and behaving positively in the classroom and on the playground are social skills and classroom behaviours that are associated with experiencing success in elementary school. In contrast, some students for different reasons struggle in social functioning and do not adhere to expected social skills that are critical to school adjustment and academic learning.

To prevent difficulties and support students at risk for behavioural and academic problems, early identification that targets social skills interventions is important. When specific difficulties are identified early, less intensive support is needed (Elliott, Huai and Roach, 2007; Lane, Oakes and Menzies, 2010). In the US, elementary schools have access to a screening instrument, the Elementary Social Behavior Assessment (ESBA; Pennefather and Smolkowski, 2015), which teachers can use to efficiently screen all students on their performance in prosocial skills related to school success, identify students with skill deficits in order to match brief differentiated interventions, and monitor students' response to the given interventions (Marquez et al., 2014; Marquez, Yeaton and Vincent, 2013; Pennefather and Smolkowski, 2015; What Works Clearinghouse, 2015).

Social skills related to school success

There is compelling evidence that students struggling in social functioning at school are at risk of experiencing social difficulties related not only to friendship and peer interactions in general but also to academic performance (Algozzine, Wang and Violette, 2011; Rivera, Al-Otaiba and Koorland, 2006). In other words, students who struggle in one of these domains may

also be at risk for failure in the other domain. Welsh, Parke, Widaman, and O'Neil (2001) examined the direction of the presumed causal relationship between academic skills and social skills over time and found that the best fitting model was a reciprocal relationship between the two skills. In their model, students' academic achievement in Grades 1 to 3 had a direct impact on social competence, and social competence was related to academic proficiency from Grades 2 to 3. Research has also shown that social and behavioural challenges in early elementary school can endure through middle school but that interventions can diminish both immediate and long-term problems (Bradshaw, Zmuda, Kellam and Ialongo, 2009; Kellam, Ling, Merisca, Brown and Ialongo, 1998). In summary, these findings agree on the importance of promoting students' social and academic success in early grades, and to do so, early identification of difficulties is crucial.

Identifying students at risk for social behavioural challenges associated with academic learning and social interactions at an early age is vital for prevention and early intervention efforts. There are, however, previous European researches that have focused on teachers' and parents' assessments of children's social and academic characteristics in relation to preventative practice in preschool and kindergarten (e.g. Mooij, 2000; Mooij and Smeets, 2009). Still, to our knowledge, little tradition for the systematic use of educational assessment instruments that can lead to effective instructions for social skills to promote students' success in school (Arnesen, Braeken, Ogden, and Melby-Lervåg, 2017). For instance, in Norway, there are only a few validated instruments available for use in schools (e.g. SSRS; Social Skills Rating System, Gresham and Elliott, 1990; Ogden, 2003). Such instruments often take too much time to be of practical use for teachers in a general classroom setting. Many other instruments used for assessing students' social skills in schools do not have acceptable psychometric properties. Notably, most of the instruments that have been validated focus more on problem behaviour than prosocial skills or broad social skills that are not as relevant in a school context (e.g. Crowe,

Beauchamp, Catroppa and Anderson, 2011; Floyd et al., 2015).

Teachers most often rely on their own judgements without systematic descriptions, which can be haphazard and unreliable (Arnesen et al., 2017). Although most teachers have sufficient experiences to judge students' social functioning and classroom behaviour informally on a regular basis, these judgements might be inconsistent or inaccurate. Their judgements may often be too broad, such as describing a student's behaviour as simply 'poor', 'inappropriate', or 'troublemaking', rather than in terms of modifiable skills (Merrell, 2002). Because of external demands, teachers have limited time for systematic judgement of students in these domains, but at the same time, they need to be able to measure social skills reliably and validly in order to intervene or support their skill development. Therefore, access to efficient and valid skill-based assessment instruments can help teachers identify students and tailor instruction to accommodate students' social skill competencies and deficits. Furthermore, it is prudent to establish regular practices and systems for screening students' social skills that impact their academic engagement and peer relations in the school context to support data-based decisions for early intervention. Monitoring students' progress in social functioning over time can also help determine if interventions are having their intended impact.

The purpose of the present study is to examine the psychometric properties of the ESBA and its relationship with the teacher's form of the SSRS in a sample of Norwegian students. The ESBA was developed in the US for students in kindergarten to Grade 3 (age 5-8). Due to its strong theoretical foundation (Walker et al., 2015), the positively worded, skill-based items that make it simple and efficient for teachers, and the ease in which they can match brief classroom interventions to skills, we adapted the ESBA for Norwegian students in Grades 1-6 (age 6-12) to expand the literature and extend school practice to non-English speaking populations. To date, no studies have translated or adapted the ESBA for schools outside of the US. Also noteworthy,

there are only two previous validation studies of the ESBA in the US (Pennefather and

Smolkowski, 2015).

Elementary Social Behaviour Assessment

ESBA collects teacher ratings on 12 positively worded items based on several studies of

teacher-preferred student behaviours and behavioural expectations (e.g. Walker and McConnell,

1995; Walker and Severson, 1992; Walker et al., 2014). The items were designed to match social

skills that teachers usually perceived as those most important for students to succeed

academically and socially (Pennefather and Smolkowski, 2015). The ESBA consists of a 3-point

scale for universal screening: mastery, needs improvement, and cause for concern. The items

used for monitoring students' progress from social skills intervention are the same as for the

universal screening, but it uses a more sensitive 6-point scale, splitting each of the three screener

ratings into a high and low rating (e.g. high-skill mastery vs. low-skill mastery, both above higher

or low-'needs improvement').

Note that the social skills construct has been defined and operationalised in different ways

in the literature and there is little agreement among researchers and scholars on how it should be

defined and measured (Cordier et al., 2015; Stump, Ratliff, Wu and Hawley, 2009; Sutherland,

McLeod, Conroy and Cox, 2013). According to Missall and Hojnoski (2008), the relationship

between learning-relevant social skills and academic achievement is stronger than the association

between interpersonal skills and academic achievement. Because the construct social skills

covered in the ESBA is associated with teachers' expectations of students primarily in the

classroom context, it may be more narrowly defined than some other broad social skills

measures, such as the SSRS. The ESBA items have been defined to identify specific, observable

prosocial skills and behaviours that predict students' learning and school adjustment (Cummings,

Kaminski and Merrell, 2008; Gresham, Elliott, Cook, Vance and Kettler, 2010; Pennefather and

Smolkowski, 2015).

As mentioned earlier, two studies in the US have examined psychometric properties of the ESBA scale. These studies, reported in a single paper, showed high score reliabilities ($\alpha > .94$) and a high test–retest correlation ($r = .77$, $p < .001$; Pennefather and Smolkowski, 2015). The ESBA items fit a single factor, suggesting that the items in the US samples represented one general social skills measurement construct. Pennefather and Smolkowski also reported high criterion validity correlations with the Walker–McConnell Scale (WMS) of social competence and school adjustment–elementary version ($r = .84$; Walker and McConnell, 1995) in Study 1 ($n = 187$) and the Brief Behavior Rating Scale (BBRS; $r = .77$; Gresham, Cook, et al., 2010) in Study 2 ($n = 1616$). In summary, previous studies indicated that the ESBA scale adequately captured the constructs of students' classroom social skills and allowed the teachers to rate all their students in about 15-30 min per class (Marquez et al., 2014; Pennefather and Smolkowski, 2015).

Research objectives

The aim of the present study was to evaluate the psychometric properties of the Norwegian adaption of the ESBA. If successfully validated, the ESBA would meet the need for a valid instrument that targets social skills for elementary schools students that teachers can use easily and efficiently. Pennefather and Smolkowski (2015) described two studies used to validate the ESBA in the US schools with students in kindergarten to Grade 3, which demonstrated respectable psychometric properties. Suitable reliability and validity in an English-speaking sample from the US, however, does not mean that the ESBA will perform as well for different age groups or when used with students from a different country and culture who speak a different language. The present study also extended the investigation by Pennefather and Smolkowski (2015) by testing the psychometric properties for students Grades 1-6.

The adaption of a measure requires several carefully executed steps and does not guarantee it will provide the same information as the original. First, it requires an appropriate language translation and adjustment to culture and new context (International Test Commission, 2016). Although words may translate straightforwardly, their meaning in context can shift (e.g., ESBA's item 10: 'Can have "normal" conversations…' is translated as 'Can communicate properly and decent…', because 'normal' has a more normative meaning if translated directly to Norwegian). We know from several validation studies that translation itself can change both the factor structure and the concurrent and predictive validity of rating scales (e.g. Borsa, Damásio and Bandeira, 2012; van Widenfelt, Treffers, de Beurs, Siebelink and Koudijs, 2005). Second, the education for young students in Europe, and more specifically in Scandinavia, differs from American elementary schools in several ways. Scandinavian schools focus more on play and play-based learning than on formal instruction in preschool and the early grades. In Norway, children enter elementary school at the age of six and most of them have been to kindergarten from 1-year old until they begin Grade 1. In Norway, then, kindergarten combines what in the US is considered preschool and kindergarten and nearly all children attend. Factors such as these may affect the extent to which students develop social skills, their importance and relationship to learning in school, and how teachers perceive students' difficulties with social skills, which emphasises the need for this validation of the Norwegian adapation of the ESBA.

Our study will add to the literature and educational practice in many ways. The present study expands the use of the ESBA to Grades 1-6 (age 6-12) in a Norwegian context, whereas the previous studies of ESBA were carried out in the US with students in kindergarten through Grades 3 (age 5-8). Also, the previous validation studies compared the ESBA to the WMS (Walker and McConnell, 1995) and the BBRS (Gresham et al., 2010a), whereas the present study used the teacher version of the Social Skills Rating System (SSRS-T; Gresham and Elliott, 1990;

Ogden, 2003) to establish criterion validity. Note that we did not expect the ESBA to correlate with all SSRS-T subscales. Specifically, the ESBA does not contain items that specifically address assertion, internalising behaviours, or academic competence.

Taken together, this study aimed to evaluate the psychometric properties of the Norwegian adaption of the ESBA. More precisely, we intended to examine the (a) factor structure, (b) score and test–retest reliability, (c) criterion validity, and (d) whether the school or teacher levels could explain important variance on students' ESBA scores. We intended the latter aim to determine if differences between teachers and schools in the diverse Norwegian sample might explain notable variability in ESBA responses, which might imply subculture differences in the measure's functioning.

Method

We evaluated the Norwegian version of the ESBA screener with elementary classroom teachers and their students in spring 2015 using a test–retest design with a second assessment collected 8 weeks after the first.

Sample

From the official list of elementary schools, we invited a randomly selected pool of 100 schools to participate, of which 31 agreed. The study included 793 students in Grades 1 to 6 with parental consent, each rated by one of 151 teachers, and demographic information from 523 parents. The schools were located in both urban and rural districts across Norway, covered all regions of the country that were in the initial pool, and included students from a variety of socio–economic backgrounds. Table 1 reports gender, grade level, gender of students and teachers, and other demographic information for the students that participated. The students were predominantly Norwegian and only 8.4% of the students spoke Norwegian as a second language. A small number (5.4%) of the students received special education services. Parents ($n = 524$)

provided demographic information on 66.7 % of the students; 84.7 % were mothers, 14.3% fathers, and 1 % foster- or step-parents.

[Table 1 near here]

Measures

The ESBA is translated into Norwegian from the American original version and culturally adapted by an expert panel. The ESBA scale consists of 12 items that cover the social skills and behaviours that teachers associate with students' success in elementary grades (see Table 2). The scale was delivered to teachers via a web-based data collection system. Although the ESBA was designed to allow teachers to either rate all students on one item at a time or rate each student on all 12 items, for practical reasons the teachers in the present study rated one student at a time on all items. They were asked to score each item on a 3-point scale with one of three response options: (a) 'skill mastered' (almost always), coded 3; (b) 'needs improvement' (sometimes), coded 2; and (c) 'cause for concern' (rare or never), coded 1. The items were self-explanatory, although some items included short, additional descriptions. The items were averaged to create a mean score.

[Table 2 near here]

The Social Skills Rating System (Gresham and Elliott, 1990) is a social–emotional behaviour instrument for kindergarten, elementary, and middle school available in versions for self-reports from teachers, parents, and students from Grade 3. In this study, we evaluated the relation between the ESBA and the teachers' ratings (SSRS-T), which had been already adapted and validated for Norwegian (Ogden, 2003). The SSRS-T used is a 4-point scale from 1 (*never*) to 4 (*very often*) for ratings. The 30 items on social skills include subdomains of assertion, cooperation, and self-control, and the 18 items on problem behaviour include subdomains of externalising behaviour, internalising behaviour, and hyperactivity.

In addition, the SSRS-T included nine items about students' academic skills (such as overall performance, reading and math proficiency, motivation, parental encouragement, general cognitive function, and overall classroom behaviour) compared with other students at the grade level. Academic items are rated on a 5-point scale: 1 = lowest 10%, 2 = almost lowest 20%, 3 = average 40%, 4 = almost highest 20%, 5 = highest 10%. The Norwegian validation study of SSRS-T showed a mean score reliability of $\alpha = .91$ and a test–retest correlation of $r = .58$ (Ogden, 2003).

Demographic information was collected from participants. Parents completed a 15-item questionnaire about student and family demographic background (such as numbers of family members, living area, language minority, nationality, siblings, parent status, residence, period of residence, age, household income, and parents' education). Income was reported in Norwegian krone (NOK). In 2015, one NOK equaled approximately €0.11 Euro or $0.12 US, so 300,000 NOK equals about €33,000 or $36,000.

Data collection procedure

Teachers rated participating students on the ESBA and SSRS-T during each of two assessment windows that were about 8 weeks apart. Teachers completed a web-based questionnaire form about different aspects of student's social skills and academic abilities. Each teacher rated 5-7 students on average, chosen at random, to reduce the workload for the participating teachers. The teachers required 1-2 min to rate each student on the ESBA scale and 15-20 min on the SSRS-T scale on average. The parents completed their questionnaires during the second assessment window at the end of the school year in 2015. We offered parents a paper-and-pencil or an electronic online version. The vast majority (71%) of the parents completed the online form.

Statistical analysis

Means, variances, and item-total correlations were computed for each item on both the ESBA and SSRS-T social skills scale (assertion, cooperation, and self-control). Average scores were calculated for each student on the EBSA at pre-test and post-test as well as the SSRS-T. Similar to Pennefather and Smolkowski (2015), we examined construct validity with an exploratory factor analysis (EFA) on ESBA pretest items to identify the factors underlying the Norwegian adapted version and a confirmatory factor analysis (CFA) on the posttest ESBA items to validate the factor structure.

The ESBA items have only three ordinal values, so we could not assume they were normally distributed. Hence, we conducted a factor analysis that assumed ordered categorical data with M*plus* (Muthén and Muthén, 1998-2012). Maximum-likelihood estimation with robust standard errors produces unbiased estimates for models with categorical outcomes (Brown, 2006). Model fit indices, however, are unavailable from maximum likelihood methods used with categorical outcomes. For both EFA and CFA, we estimated model parameters with robust maximum likelihood methods and the tests of model fit with robust weighted least squares. As in Pennefather and Smolkowski, the overall pattern of model parameters (e.g., structure matrix, loadings, factor correlations) differed only marginally across the two methods and did not affect conclusions.

The EFA on the pretest ESBA used the oblique rotation, geomin, which allows correlations between the factors (Browne, 2001). With model fit criteria unavailable for EFA with categorical data, we adopted criteria for confirmatory factor analyses of categorical data from Schreiber, Nora, Stage, Barlow, and King (2006). Schreiber et al. recommend comparative fit index (CFI) and Tucker–Lewis index (TLI) values equal to or greater than .95. They did not suggest criteria value for the standardised root mean square residual (SRMR) for categorical data,

although they generally recommended SRMR values below .08. To validate the factor structure suggested from pretest data and Pennefather and Smolkowski (2015), we fit the post-test ESBA data to one- and two-factor models with a CFA.

We estimated score reliability with Guttman's (1945) λ_2 and λ_3 as well as 8-week test–retest reliability. Guttman's λ_3 or coefficient α, often incorrectly attributed to Cronbach (1951; Sijtsma, 2009), has advantages in ease of estimation (Guttman, 1945) but does not perform as well compared with λ_2 (Bendermacher, 2010; Guttman, 1945; Sijtsma, 2009). We reported λ_3 (coefficient α), however, due to its widespread use. Criterion validity, both predictive and concurrent, was estimated with correlation analyses. We also ran partial correlation analysis to explore whether validity correlations differ across variation in demographic factors (e.g. student or teacher gender, receipt of special education services, socioeconomic characteristics). Finally, we estimated intraclass correlations to examine the extent of school- or teacher-level variance in teachers' ratings with the ESBA.

We were missing data for 19 students (2.4%) on the pretest and 91 students (11.5 %) at post-test. This was mainly because they were absent from school at the time of assessment or they had moved from the school. Therefore, we consider them missing at random, albeit not missing completely at random (Graham, 2009). We excluded students with missing data from the analysis at each assessment and given the limited number of missing cases and the most likely mechanisms, we do not expect missingness to bias the results (Graham, 2009; Schafer and Graham, 2002).

Results

The mean score on the 12 ESBA items ranged between 2.6 and 2.8 at both pre-test and post-test (see Table 2). Thus, on the 3-point mastery scale the teachers general evaluation was that most children had mastered the specific skills measured by each item, which was as expected

and aligned with the results of Pennefather and Smolkowski (2015) from the US schools. For every item at both time points, however, at least one student scored a 1, indicating cause for concern, and several scored a 2, representing a need for improvement, which allows teachers to identify students or skills that require additional instruction.

Factor analysis

The results of the exploratory factor analyses (EFA) at pretest are presented in Table 3. The fit statistics supported multiple solutions, although eigenvalues of 9.4, 1.1, 0.4, and 0.3 for the one- to four-factor solutions suggested a one- or two-factor model using the eigenvalue-greater-than-one rule (Kaiser criterion) or subjective scree test. The principle of parsimony also supported a one-factor or two-factor model. The three-factor model had better fit to the data than the two-factor model (difference $\chi^2 = 182.994$, $df = 11$, $p < .001$), but the solution extracted a third factor on which no items loaded most highly. That is, in the three-factor model, all items loaded on the first two factors, which formed factors identical to those of the two-factor model. The four-factor solution similarly included a factor on which no items loaded most highly.

[Table 3 near here]

The distinction between the one- and two-factor solutions, however, is less clear, as the data fit the two models similarly according to the CFI and TLI. The two-factor solution had better fit than the one factor solution according to SRMR and χ^2 difference test ($\chi^2 = 372.8$, $df = 10$, $p < .001$). Nonetheless, the two factors correlate highly, $r = .72$, at pretest.

To confirm the factor structure and further clarify whether the data fit a one- or two-factor solution, we conducted a CFA with post-test assessments. Table 4 shows the standardised factor loadings and fit statistics. We did not test a three- or four-factor confirmatory model because the EFA solutions did not define three or four factors, respectively. The CFI and TLI fit statistics again indicated the data adequately fit both the one- and two-factor models (all > .975), but a

statistically significant difference χ^2 suggested the two-factor model was more appropriate ($\chi^2 =$ 130.9, $df = 17$, $p < .001$). Table 4 also shows that the single-factor model and both factors in the two-factor model included standardised loadings greater than .78.

[Table 4 near here]

Thus, the results of the factor analyses suggest that users of the measure could split the scale into two measures or use a single scale with all items. In the two-factor model, Factor 1 includes the social skills generally important in structured, instructional settings, such as when teachers are leading instruction to promote students' academic engagement, so we label Factor 1 Academic Engagement factor. Factor 2 includes interpersonal social skills generally experienced when students are working or playing independently or in less-structured environments, so we refer to Factor 2 as Peer Social Relations factor. The factors were correlated highly in both the pretest EFA, $r = .73$, and post-test CFA, $r = .84$.

We next created subscales by taking the mean of the items across each factor and a total scale from the mean of all items. The Academic Engagement and Peer Social Relations subscales also correlated highly at pre-test, $r = .72$, and post-test, $r = .76$, but not quite as highly as the correlation among latent factors in the CFA at post-test. See Table 5 for the correlations between subscales at pre-test and post-test separately for Grades 1-3 and for Grades 4-6 as well as across all six grades.

[Table 5 near here]

Reliability

To support the use of the single scale suggested by Pennefather and Smolkowski (2015), we estimated two types of reliability: score reliability and test–retest reliability. We estimated score reliability for the full ESBA scale with λ_3 (coefficient α) and Guttman's λ_2, which produced values of .94 ($n = 774$) and .95 ($n = 702$) at pre-test and post-test, respectively. Guttman's λ_2 and

λ_3 (α) differed by only .001 for this scale. All 12 ESBA items correlated with the total scale score

greater than $r = .73$ at pre-test and $r = .74$ at post-test.

Score reliability was also high for the ESBA Academic Engagement and Peer Social

Relations subscales, with reliabilities of .92 and .91 at pre-test, respectively, and .93 and .90 at

post-test. High score reliability values were consistent across grades and at both time points. Note

that we also found high score reliabilities for the SSRS-T's scales for social skills (assertion,

cooperation, and self-control) at both pre-test and post-test, λ_3 from .94 to .96, in line with Ogden

(2003).

To examine the test–retest reliability, we correlated ESBA scores from the two time

points. As expected the teachers' ratings of students' social skills remained stable over the period

of 8 weeks, $r = .78$. We estimated test–retest reliabilities of .78 and .71, respectively, for the

ESBA Academic Engagement and Peer Social Relations subscales.

Criterion validity

We next assessed criterion validity of the ESBA through correlations with the SSRS-T.

Table 6 presents the concurrent and predictive validity correlations for the ESBA total score and

the two subscales and all SSRS-T scales and subscales. Due to the broad age span, we provide

correlations for early elementary students in Grades 1-3 and late elementary students in Grades 4-

6 to examine invariance of correlations across age ranges as well as all students.

[Table 6 near here]

Across both assessment times and for all students, the ESBA and its subscales

correlated .63 to.70 concurrently with the SSRS-T Social Skills total scale, -.71 to -.77 with the

Problem Behavior scale, and .32 to .48 with the Academic scale. The concurrent validity

correlations for the ESBA scales were moderate to high, with at least 25% overlapping variance

($|r| \geq .50$), for all SSRS-T scales except for the Assertion, Internal Problem Behavior, and

Academic scales at pre-test and post-test and for early or late elementary grades or for the whole sample. The concurrent correlations within the three samples, early or late elementary or all students, agreed within .10 for all cases except one. The SSRS-T Assertion and ESBA Peer Social Relations scales were correlated .34 for early elementary, .48 for late elementary, and .41 for all students.

Eight-week predictive correlations were slightly lower than concurrent correlations for all students, with pretest ESBA scales correlating with the SSRS-T post-test scales .58 to .64 for Social Skills, -.65 to -.71 for Problem Behavior, and .31 to .45 for the Academic scale. The predictive correlations also agreed within .10 across the three samples, with some exceptions. The correlations between pretest SSRS-T and post-test ESBA scales in some cases differed between the early and late elementary samples. For example, the pre-test SSRS-T External Problems was a stronger predictor of post-test ESBA total scale in late elementary (e.g., $r = -.76$) than early elementary grades (e.g. $r = -.57$). See Table 6 for details.

Overall, the ESBA scales correlated higher with the SSRS-T scales of individual social skills and problem behaviour than the SSRS-T scale in which teachers rated students' academic functioning compared with other students in their class. The ESBA Academic Engagement scale correlated higher with the SSRS-T Academic than the ESBA Peer Social Relations scale. The ESBA scales also had lower correlations with the SSRS-T Assertion and External Problems scales than other social skills, with average correlations across all cells of .39 and.35, respectively, in absolute value.

The correlations between the ESBA scores and between the ESBA and the SSRS-T scales at pre-test and post-test were also consistent after controlling for background variables. The background variables were controlled for one at a time (i.e. grade levels, students' or teachers' gender, region of the country, students' receipt of special education, whether students had

siblings, whether students lived with one or both parents, how long students lived in the country, and parents' education, income, ethnicity). Notably, parents' levels of education are similar to the mean national percentage of adults in between 30-50 years who have attained tertiary education in Norway in 2015 (60% graduated after 4 years and 29% > 4 years) (Statistic Norway, 2016). This means that the correlations that partialed out the background variables were similar to the uncontrolled bivariate correlations.

Variances at school and teacher level

The students in our samples were nested in schools and classrooms. This violates the assumption of independency in the data. It was therefore important to examine whether our results were affected by this dependency in observations. We therefore estimated the percent of school- and teacher-level variance (i.e., intraclass correlations) to examine how this dependency affected our results. We estimated the intraclass correlations from a three-level model and decomposed the total variance into three independent variance components, one for each level (Hox, 2002). At pretest, we estimated 0% of the variance at the school level (very slightly negative) and 4.7% at the teacher level. At posttest, the school level accounted for 1.2% of the variance and the teacher level 5.1%. These intraclass correlations are low when compared to those found for most academic measures (e.g. Hedges and Hedberg, 2007) but similar to the classroom-level intraclass correlations reported for other social skills measures (e.g., Low, Cook, Smolkowski and Buntain-Ricklefs, 2015). Thus, it is not likely that the dependency in the data creates a bias in our findings.

Discussion

The purpose of this study was to evaluate the psychometric properties of the adaptation of the ESBA for Norwegian elementary students in Grades 1-6. The results demonstrated a factor structure with one or possibly two factors, excellent score and test–retest reliability, adequate

criterion validity, and that the school and teacher levels explained a limited amount of the total variance in ESBA scores. The majority of the results were remarkably similar to those of Pennefather and Smolkowski (2015), who evaluated the measure with students from the western states of the US. The EFA and CFA fit was similar as were the reliabilities. For example, we estimated test–retest reliability for the ESBA total of .78, whereas Pennefather and Smolkowski estimated test–retest reliability of .77.

This study differed from the previous study of the ESBA (Pennefather and Smolkowski, 2015) because it examined the use of the measure in the Norwegian culture and context. Our study also included a larger range in students' grade levels, 1 to 6, than the American study, kindergarten to Grade 3, and compared with another external instrument (SSRS) than those used in the American version. Nonetheless, we found that Norwegian teachers rated the students' mastery of social skills very similarly to American teachers.

Notably, the results differed between the two studies since our study suggested potentially two constructs within the ESBA scale. Although the results of Pennefather and Smolkowski also suggested two factors, the items that contributed to their two factors produced less interpretable scales than the items in our study. The present analysis, in contrast, suggested one factor or scale for social skills within an instructional setting (Academic Engagement) and a second factor for less structured settings (Peer Social Relations), and we estimated excellent score reliability for the subscales. The identification of the two subscales certainly extends the ESBA scale's content validity.

The two ESBA subscales were highly correlated (see Table 5) but address slightly different areas of social skills that may affect school success. Given the high correlations, teachers may opt to use the total scale for screening purposes. Nonetheless, the two subscales might allow for supplemental instruction in two groups with different foci corresponding to

students with low scores on different subscales. That is, students who scored poorly on the Academic Engagement subscale could receive supplemental supports or instruction on how to behave appropriately in the classroom, while students with inadequate scores on the Peer Social Relations subscale could receive supplemental supports to improve interactions with peers outside of structured classroom settings as well as how to recognise and control intense feelings. If teachers have access to just one intervention, however, that addresses both of these highly related sets of skills, then the total scale would be more appropriate.

The strong concurrent and predictive relationships between ESBA and the Norwegian SSRS-T (Ogden, 2003) showed respectable criterion validity for the ESBA measure. ESBA criterion validity was highest for SSRS-T scales of Social Skills and Problem Behavior. The ESBA was most strongly correlated with the Cooperation and Self-Control subscales of the Social Skills scale. Within the Problem Behavior scale, the ESBA was highly correlated with the External Problems and Hyperactivity subscales but less so with External Problems. It is unusual, however, for a positively worded scale that focuses on skill mastery to correlate highly with negatively worded problem behaviour scales (e.g. Boles, Biglan and Smolkowski, 2006). This suggests that the ESBA can help teachers identify struggling students who exhibit problem behaviours with a set of positively worded, mastery-focused items rather than a scale the focuses on negatively worded problem behaviours. The ESBA scales was, as expected, not a strong predictor of academic competence on SSRS-T, although that association was stronger than the correlation between the ESBA and Assertion and Internal Problems subscales of the SSRS-T.

Compared with the US schools, Norwegian schools and teachers follow similar public policies and instructional practices across schools. We therefore expected little variation at either school or teacher level. The intraclass correlations showed a low proportion of school-level variance, which also indicates limited variability across schools in different regions in Norway.

Also, there were variations in socio-economic levels of the participating schools. However, it is not likely that this has affected the intra class correlation since the schools were not overrepresented in areas characterised by higher level of parents' education and income, the intraclass heterogeneity may not be reduced. Although the small number of schools (31) may produce unstable variance estimates (Feng, Diehr, Peterson and McLerran, 2001; Hox and Maas, 2001), given the estimates of 0%-1% of the total variance, estimates from a larger sample are unlikely to be large.

Teachers accounted for about 5% of the total variance from both assessments. Although important for statistical modeling, this level of variability has limited practical importance for teachers who wish to screen their classrooms. The intraclass correlations for teachers did not include school-level variability, so the 5% of the total variance represented the association of teacher reports within schools. That is, teacher ratings of their students within classrooms were more similar than across classrooms within schools but not between schools. The intraclass correlations might suggest differences in teachers' expectations about how students should be able to perform within schools, but it is also important to recognise that we cannot distinguish variability in teacher ratings of students from actual differences in class composition and student behaviour, which may explain some of that 5%. It therefore appears that teachers rate students with the ESBA quite similarly.

In summary, the results indicate that the ESBA scale works as well in Norway as in the US (Pennefather and Smolkowski, 2015) and may work similarly well to screen students in other languages or cultures. The Norwegian adapted version of the ESBA shows respectable psychometric properties and seems to work nicely across Grades 1 to 6 regardless of schools' region, students' and teachers' gender, socio–economic status, and other background characteristics. The adjustment and translation of the instrument to the Norwegian language and

school culture using an expert panel maintained the quality of the original instrument. Thus, the present findings highlight the importance of not only translating languages back and forth but also the need for careful attention to cultural adjustment of the language used (International Test Commission, 2016).

Limitations

This study offers no information about the participating teachers' experiences using the ESBA. The sample of 151 teachers were not provided with the opportunity to offer their perspective on the ESBA measure in terms of content or administration procedures. The small random sample of 31 participating schools also limits generalizability to other Norwegian elementary schools because we have no information about why the nonparticipating schools chose not to provide ESBA ratings of their students. The teachers and schools that did participate, however, hailed from every region of the country, and the Norwegian version of the ESBA performed similarly to the English-language measure even through translation and adaptation for a different part of the world.

Implications for practice and further research

The ESBA is a reliable and valid measure for screening students' social skills by classroom teachers and extends the pool of assessment instruments available. The results provide evidence supporting the reliability and validity of the Norwegian version of ESBA similar to the previous US study. Because the ESBA is an efficient, time-saving instrument, it allows teachers to identify students for brief social skills interventions in the classroom (see Marquez et al., 2014). Although many teachers bemoan behaviour and social-skills assessments due to the tedium of their length, the ESBA allows for early identification and monitoring of the development of students who face difficulties in social skills with considerably less effort. Additional research, however, should be dedicated towards the original and Norwegian adapted

ESBA scale for progress monitoring, which uses a 6-point scale to allow for more sensitivity when rating students who are participating in active interventions or receiving other behavioural supports.

Acknowledgements

References

Algozzine, B., Wang, C. and Violette, A. S. 2011. Reexamining the relationship between academic achievement and social behavior. *Journal of Positive Behavior Interventions, 13,* 3–16. doi:10.1177/1098300709359084

Arnesen, A., Braeken, J., Ogden, T., and Melby-Lervåg, M. 2017. *Assessing children's social functioning and reading proficiency: A systematic review and psychometric evaluation of assessment instruments used in Norwegian elementary schools.* Manuscript submitted for publication.

Bendermacher, N. 2010. Beyond alpha: Lower bounds for the reliability of tests. *Journal of Modern Applied Statistical Methods, 9,* 95–102.
http://digitalcommons.wayne.edu/jmasm/vol9/iss1/11

Boles, S., Biglan, A. and Smolkowski, K. 2006. Relationships among negative and positive behaviours in adolescence. *Journal of Adolescence*, *29*(1), 33–52. doi: 10.1016/j.adolescence.2005.01.007

Borsa, J. C., Damásio, B. F., and Bandeira, D. R. 2012. Cross-cultural adaptation and validation of psychological instruments: Some considerations. *Paidéia (Ribeirão Preto), 22,* 423–432. doi:10.1590/1982-43272253201314

Bradshaw, C. P., Zmuda, J. H., Kellam, S. G. and Ialongo, N. S. 2009. Longitudinal impact of two universal preventive interventions in first grade on educational outcomes in high school. *Journal of Educational Psychology, 101,* 926–937. doi:10.1037/a0016586

Brown, T. 2006. *Confirmatory factor analysis for applied research.* New York, NY: Guildford.

Browne, M. W. 2001. An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36,* 111–150. doi:10.1207/S15327906MBR3601_05

Cordier, R., Speyer, R., Chen, Y., Wilkes-Gillan, S., Brown, T., Bourke-Taylor, H., . . . Leicht, A. 2015. Evaluating the psychometric quality of social skills measures: A systematic review. *PLoS ONE, 10*(7): e0132299. doi:10.1371/journal.pone.0132299

Cronbach, L. J. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334. doi:10.1007/BF02310555

Crowe, L. M., Beauchamp, M. H., Catroppa, C. and Anderson, V. 2011. Social function assessment tools for children and adolescents: A systematic review from 1988 to 2010. *Clinical Psychology Review, 31*(5), 767–785. doi:10.1016/j.cpr.2011.03.008

Cummings, K. D., Kaminski, R. A. and Merrell, K. W. 2008. Advances in the assessment of social competence: Findings from a preliminary investigation of a general outcome measure for social behavior. *Psychology in the Schools, 45,* 930–946. doi: 10.1002/pits.20343

Elliott, S. N., Huai, N. and Roach, A. T. 2007. Universal and early screening for educational difficulties: Current and future approaches. *Journal of School Psychology, 45,* 137–161. doi:10.1016/j.jsp.2006.11.002

Feng, Z., Diehr, P., Peterson, A. and McLerran, D. 2001. Selected statistical issues in group randomized trials. *Annual Review of Public Health, 22,* 167–187. doi:10.1146/annurev.publhealth.22.1.167

Floyd, R. G., Shands, E. I., Alfonso, V. C, Phillips, J. F., Autry, B. K., Mosteller, J. A., , … Irby, S. 2015. A systematic review and evaluation of the technical characteristics of adaptive behavior scales and recommendations for practice. *Journal of Applied School Psychology, 31,* 83–113. doi:10.1080/15377903.2014.979384

Graham, J. W. 2009. Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60,* 549–576. doi:10.1146/annurev.psych.58.110405.085530

Gresham, F. M., Cook, C., Collins, T., Rasethwane, K., Truelson, E. and Grant, S. 2010a. Developing a change-sensitive brief behavior rating scale as a progress monitoring tool for social behavior: An example using the Social Skills Rating System–Teacher Form. *School Psychology Review, 39,* 364–379.

Gresham, F. M. and Elliott, S. N. 1990. *Social Skills Rating System.* Circle Pines, MN: American Guidance Service.

Gresham, F. M., Elliott, S. N., Cook, C. R., Vance, M. J. and Kettler, R. 2010b. Cross-informant agreement for ratings for social skill and problem behavior ratings: An investigation of the Social Skills Improvement System—Rating Scales. *Psychological Assessment, 22,* 157–166. doi:10.1037/a0018124

Guttman, L. 1945. A basis for analyzing test-retest reliability. *Psychometrika*, *10*, 255–282. doi:10.1007/BF02288892

Hedges, L. V. and Hedberg, E. C. 2007. Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis, 29,* 60–87. doi:10.3102/0162373707299706

Hox, J. J. 2002. *Multilevel analysis: Techniques and applications.* Mahwah, NJ: Erlbaum.

Hox, J. J. and Maas, C. J. M. 2001. The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling, 8,* 157–174. doi:10.1207/S15328007SEM0802_1

International Test Commission. 2016. *International test commission guidelines for translating and adapting tests* (2nd ed.). Retrieved from http://www.psyssa.com/wp-content/uploads/2015/11/ITC-Guidelines-Translating-and-Adapting-Tests-v2-3.pdf

Kellam, S. G., Ling, X., Merisca, R., Brown, C. H. and Ialongo, N. 1998. The effect of the level of aggression in the first grade classroom on the course and malleability of aggressive

behavior into middle school. *Development and Psychopathology, 10,* 165–185.

    doi:10.1017/S0954579498001564

Lane, K. L., Oakes, W. and Menzies, H. 2010. Systematic screenings to prevent the development

    of learning and behavior problems: Considerations for practitioners, researchers, and

    policy makers. *Journal of Disability Policy Studies, 21,* 160–172. doi:

    10.1177/1044207310379123

Low, S., Cook, C. R., Smolkowski, K. and Buntain-Ricklefs, J. 2015. Promoting social-emotional

    competence: An evaluation of the elementary version of Second Step®. *Journal of School*

    Psychology, 53, 463–477. doi:10.1016/j.jsp.2015.09.002

Marquez, B., Yeaton, P. and Vincent, C. 2013. Delivering quick, efficient, and accurate

    behavioral universal screening and progress monitoring assessments using web-based,

    electronic technology. In H. M. Walker and F. Gresham (Eds.), *Handbook of evidence-*

    based practices for emotional and behavioral disorders (pp. 192–210) New York, NY:

    Guilford.

Marquez, B., Marquez, J., Vincent, C. G., Pennefather, J., Sprague, J. R., Smolkowski, K. and

    Yeaton, P. 2014. The iterative development and initial evaluation of We Have Skills!, an

    innovative approach to teaching social skills to elementary students. *Education and*

    Treatment of Children, 37, 137–161. doi:10.1353/etc.2014.0000

Merrell, K. W. 2002. *School Social Behavior Scales*. Baltimore, MD: Brookes.

Missall, K. N. and Hojnoski, R. L. 2008. The critical nature of young children's emerging peer

    related social competence for transition to school. In W. H. Brown, S. L. Odom and S. R.

    McConnell (Eds.), S*ocial competence of young children: Risk, disability, and intervention*

    (pp. 117–137). Baltimore, MD: Brookes.

Mooij, T. 2000. Screening children's entry characteristics in kindergarten. *Early Child Development and Care, 165,* 23-40. Doi: 10.1080/0300443001650103

Mooij, T., and Smeets, E. 2009. Towards systemic support of pupils with emotional and behavioural disorders. *International Journal of Inclusive Education, 13(*6), 597-616. doi: 10.1080/13603110802047978

Muthén, L. K. and Muthén, B. O. 1998-2012. *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén and Muthén.

Ogden, T. 2003. The validity of teacher ratings of adolescents' social skills. *Scandinavian Journal of Educational Research, 47,* 63–76. doi:10.1080/00313830308605

Pennefather, J. and Smolkowski, K. 2015. Validation of the Elementary Social Behavior Assessment: A measure of student prosocial school behaviors. *Assessment for Effective Intervention, 40*, 143–154. doi:10.1177/1534508414557562

Rivera, M., Al-Otaiba, S. and Koorland, M. 2006. Reading instruction for students with emotional and behavioral disorders and at risk of antisocial behaviors in primary grades: A literature review. *Behavioral Disorders, 31,* 323–339.

Schafer, J. L. and Graham, J. W. 2002. Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147–177. doi:10.1037/1082-989X.7.2.147

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A. and King, J. 2006. Reporting structural equation modeling and confirmatory factor analysis results: A review. *Journal of Educational Research, 99,* 323–337. doi:10.3200/JOER.99.6.323-338

Sijtsma, K. 2009. On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74,* 107–120. doi:10.1007/S11336-008-9101-0

Statistics Norway. 2016. *Population´s level of education, 1 October 2015.* Retrieved from Statistics Norway's website: https://www.ssb.no/en/utdanning/statistikker/utniv

Stump, K., Ratliff, J., Wu, Y. and Hawley, P. 2009. Theories of social competence from the top-down to the bottom-up: A case for considering foundational human needs. In J. L. Matson (Ed.), *Social behavior and skills in children* (pp. 23–37): New York, NY: Springer.

Sutherland, K. S., McLeod, B. D., Conroy, M. A. and Cox, J. R. 2013. Measuring implementation of evidence-based programs targeting young children at risk for emotional/behavioral disorders: Conceptual issues and recommendations. *Journal of Early Intervention, 35,* 129–149. doi:10.1177/1053815113515025

van Widenfelt, B. M., Treffers, P. D. A., de Beurs, E., Siebelink, B. M. and Koudijs, E. 2005. Translation and cross-cultural adaptation of assessment instruments used in psychological research with children and families. *Clinical Child and Family Psychology Review, 8,* 135–147. doi:10.1007/s10567-005-4752-1

Walker, H. M., Marquez, B., Yeaton, P., Pennefather, J., Forness, S. R. and Vincent, C. G. 2015. Teacher judgment in assessing students' social behavior within a response-to-intervention framework: Using what teachers know. *Education and Treatment of Children, 38,* 363–382. doi:10.1353/etc.2015.0019

Walker, H. M. and McConnell, S. R. 1995. *The Walker-McConnell scale of social competence and school adjustment, Elementary version.* San Diego, CA: Singular Publishing Group.

Walker, H. M. and Severson, H. H. 1992. *Systematic screening for behavior disorders (SSBD): User's guide and administration manual.* Longmont, CO: Sporis West.

Walker, H. M., Small, J. W., Severson, H. H., Seeley, J. R. and Feil, E. G. 2014. Multiple-gating approaches in universal screening within school and community settings. In R. J. Kettler, T. A. Glover, C. A. Albers and K. A. Feeney-Kettler (Eds.), *Universal screening in educational settings: Evidence-based decision making for schools* (pp. 47–75). Washington, DC: American Psychological Association.

Welsh, M., Parke, R., Widaman, K. and O'Neil, R. 2001. Linkages between children's social and academic competence: A longitudinal analysis. *Journal of School Psychology, 39,* 463–482. doi:10.1016/S0022-4405(01)00084-X

What Works Clearinghouse. 2015. *WWC review of the report: The iterative development and initial evaluation of We Have Skills!, an innovative approach to teaching social skills to elementary students.* Retrieved from ERIC website: http://files.eric.ed.gov/fulltext/ED556128.pdf

Table 1. Demographic overview for students from teacher and parent reports.

Measure		n	(%)
Teacher Reports on Students (N = 793)			
Gender	Girls	403	(51)
Grade	First	162	(20)
	Second	147	(19)
	Third	120	(15)
	Fourth	119	(15)
	Fifth	108	(14)
	Sixth	137	(17)
Special Ed, IEP		42	(5)
Parent Reports on Students and their Families (N = 523)			
Gender	Girls	264	(50.5)
Child lives with	Both parents	425	(81)
	Single parent mother	66	(13)
	Single parent father	38	(7)
	One parent and a stepparent	27	(5)
	Foster home or adoptive parents	9	(2)
Family income in NOK per year	< 300 000	11	(2)
	300 000-500 000	41	(8)
	500 000-750 000	78	(15)
	750 000-1 000 000	149	(28)
	1 000 000-1 500 000	131	(25)
	1 500 000-2 500 000	60	(11)
	> 2 500'	14	(3)
Bilingual	Home	48	(9)
	Mother	78	(15)
	Father	75	(14)
Parents age	≤ 30	12	(2)
	31-40	254	(49)
	41-50	234	(45)
	≥ 50	22	(4)
Parents education	None	2	(0.4)
	Primary school	14	(3)
	1-2 years of high school	65	(13)
	Graduated high school	61	(12)
	1-2 years in college or university	48	(9)
	Graduated college or university	332	(63)
Siblings	None	30	(6)
	1	229	(44)
	2	185	(35)
	3	47	(9)
	≥ 4	31	(6)

Note. Data for the 793 students were obtained from 151 teachers in 31 schools and 523 parents. IEP = Individual education plan. NOK = Norwegian krone. In 2015 NOK 1.00 ≈ €0.11 or $0.12.

Table 2. Overview of the Elementary Social Behavior Assessment items and scale.

	Pretest		Posttest	
Skill	M	SD	M	SD
1 Listens to and respects the teacher	2.72	0.51	2.72	0.53
2 Follows the teacher's directions	2.69	0.56	2.72	0.54
3 Works with effort	2.71	0.53	2.72	0.53
4 Does seatwork assignments as directed	2.76	0.51	2.78	0.49
5 Makes assistance needs known in an appropriate manner	2.74	0.53	2.76	0.51
6 Follows rules	2.78	0.48	2.79	0.48
7 Avoids breaking rules even when encouraged by a peer	2.69	0.54	2.70	0.54
8 Behaves appropriately outside the classroom	2.82	0.43	2.81	0.47
9 Works out strong feelings appropriately	2.72	0.54	2.74	0.53
10 Can have "normal" conversations without becoming hostile	2.74	0.51	2.76	0.5
11 Gets along with peers	2.77	0.47	2.79	0.46
12 Resolves peer conflicts adequately without teacher assistance	2.59	0.61	2.64	0.59
Total Score (mean)	2.73	0.41	2.74	0.41

Note. Teachers rated each item as 3 = skill mastered (almost always), 2 = needs improvement (sometime), and 1 = cause for concern (rare or never). Pretest data were available for 774 students and post-test for 702 students.

Table 3. Exploratory factor analysis structure and fit statistics for pretest assessment data.

Item or Statistic	Single Factor	Two Factors F1	F2	Three Factors F1	F2	F3
1 Listens to and respects the teacher	.83	**.91**	.61	**.97**	.67	-.26
2 Follows the teacher's directions	.84	**.93**	.68	**.94**	.69	-.08
3 Works with effort	.83	**.91**	.67	**.93**	.67	.27
4 Does seatwork assignments as directed	.87	**.94**	.70	**.94**	.70	.16
5 Makes assistance needs known in an appropriate manner	.79	**.82**	.70	**.83**	.70	.11
6 Follows rules	.91	**.90**	.81	**.89**	.81	.00
7 Avoids breaking rules even when encouraged by a peer	.80	**.80**	.76	**.79**	.76	-.03
8 Behaves appropriately outside the classroom	.90	.79	**.93**	.76	**.93**	.06
9 Works out strong feelings appropriately	.83	.70	**.94**	.67	**.93**	.09
10 Can have "normal" conversations without becoming hostile	.88	.75	**.97**	.71	**.96**	-.06
11 Gets along with peers	.79	.64	**.90**	.61	**.91**	.06
12 Resolves peer conflicts adequately without teacher assistance	.73	.61	**.87**	.58	**.87**	-.01
χ^2	508.5	142.2		73.4		
Degrees of freedom	54	43		33		
p-value	< .001	< .001		< .001		
CFI	.979	.995		.998		
TLI	.974	.993		.996		
SRMR	.083	.020		.016		

Note. This table reports the factor structure—the correlations between items and factors—for one-, two-, and three-factor solutions from an exploratory factor analysis. F1-F3 label the factors from each model. With ordered categorical data, we used maximum likelihood to obtain unbiased factor structure estimates, but fit statistics were unavailable. Fit statistics were therefore obtained from weighted least squares methods. Bolded items represent those that loaded on the factor. For categorical data, Schreiber et al. (2006) suggest that the comparative fit index (CFI) should exceed .95 and the Tucker–Lewis index (TLI) should exceed .96. As a general rule, the standardised root mean square residual (SRMR) should not exceed .08.

Table 4. Confirmatory factor analysis standardised loadings and fit statistics for post-test assessment.

Item or Statistic	Single Factor	Two Factors	
		Academic Engagement (F1)	Peer Social Relations (F2)
1 Listens to and respects the teacher	.89	.92	
2 Follows the teacher's directions	.91	.93	
3 Works with effort	.88	.91	
4 Does seatwork assignments as directed	.90	.93	
5 Makes assistance needs known in an appropriate manner	.85	.85	
6 Follows rules	.96	.96	
7 Avoids breaking rules even when encouraged by a peer	.86	.84	
8 Behaves appropriately outside the classroom	.93		.90
9 Works out strong feelings appropriately	.84		.89
10 Can have "normal" conversations without becoming hostile	.87		.95
11 Gets along with peers	.83		.91
12 Resolves peer conflicts adequately without teacher assistance	.78		.88
χ^2	445.1	314.2	
Degrees of freedom	54	53	
p-value	< .001	< .001	
CFI	.980	.987	
TLI	.976	.984	

Note. This table reports the factor loadings for one- and two-factor confirmatory models. The first factor (F1) describes social skills in instructional settings and the second factor (F2) describes social skills that generally occur in unstructured settings. With ordered categorical data, we used maximum likelihood to obtain unbiased factor structure estimates, but fit statistics were unavailable. Fit statistics were therefore obtained from weighted least squares methods. For categorical data, Schreiber et al. (2006) suggest that the comparative fit index (CFI) should exceed .95 and the Tucker–Lewis index (TLI) should exceed .96. The standardised root mean square residual (SRMR), provided for EFA, was unavailable for CFA.

Table 4. Confirmatory factor analysis standardised loadings and fit statistics for post-test assessment.

Item or Statistic	Single Factor	Two Factors	
		Academic Engagement (F1)	Peer Social Relations (F2)
1 Listens to and respects the teacher	.89	.92	
2 Follows the teacher's directions	.91	.93	
3 Works with effort	.88	.91	
4 Does seatwork assignments as directed	.90	.93	
5 Makes assistance needs known in an appropriate manner	.85	.85	
6 Follows rules	.96	.96	
7 Avoids breaking rules even when encouraged by a peer	.86	.84	
8 Behaves appropriately outside the classroom	.93		.90
9 Works out strong feelings appropriately	.84		.89
10 Can have "normal" conversations without becoming hostile	.87		.95
11 Gets along with peers	.83		.91
12 Resolves peer conflicts adequately without teacher assistance	.78		.88
χ^2	445.1	314.2	
Degrees of freedom	54	53	
p-value	< .001	< .001	
CFI	.980	.987	
TLI	.976	.984	

Note. This table reports the factor loadings for one- and two-factor confirmatory models. The first factor (F1) describes social skills in instructional settings and the second factor (F2) describes social skills that generally occur in unstructured settings. With ordered categorical data, we used maximum likelihood to obtain unbiased factor structure estiamtes, but fit statistics were unavailable. Fit statistics were therefore obtained from weighted least squares methods. For categorical data, Schreiber et al. (2006) suggest that the comparative fit index (CFI) should exceed .95 and the Tucker–Lewis index (TLI) should exceed .96. The standardised root mean square residual (SRMR), provided for EFA, was unavailable for CFA.

Table 5. Correlations among ESBA Factors 1 and 2 at pre-test and post-test separately for upper and lower elementary school and for all students.

Time of Assess-ment	ESBA Factor	Grades 1 to 3 Above Diagonal & Grades 4 to 6 Below Diagonal				All Students in Grades 1 to 6		
		1	2	3	4	1	2	3
1 Pretest	Instructional Setting Scale (F1)		.71	.77	.58			
2	Peer Social Relations Scale (F2)	.73		.51	.66	.72		
3 Post-test	Instrucional Setting Scale (F1)	.80	.63		.73	.78	.57	
4	Peer Social Relations Scale (F2)	.66	.77	.79		.62	.71	.76

Note. The Academic Engagement scale was the mean of items in Factor 1 (F1), Skills 1-7, and the Peer Social Relations scale the mean of items in Factor 2 (F2), Skills 8-12. All coefficients were statistically significant, $p < .0001$. Shaded cells show correlations between Factors 1 and 2 within the same assessment time.

Table 6. Concurrent and predictive relationship between ESBA and SSRS-T across grade levels.

| | | Pretest ESBA | | | | | | | | | Posttest ESBA | | | | | | | | |
| | | Total | | | Academic Engagement (F1) | | | Peer Social Relations (F2) | | | Total | | | Academic Engagement (F1) | | | Peer Social Relations (F2) | | |
SSRS-T Measure	Time	Early	Late	All	Early	Late	All	Early	Late	All	Early	Late	All	Early	Late	All	Early	Late	All
							SSRS-T Social Skills												
Total	Pre	**.73**	**.68**	**.70**	.68	.64	.66	.66	.62	.64	**.62**	**.67**	**.64**	.59	.63	.61	.56	.63	.60
	Post	**.62**	**.66**	**.64**	.60	.61	.60	.54	.62	.58	**.65**	**.70**	**.68**	.62	.66	.64	.58	.68	.63
Assertion	Pre	.43	.42	.43	.39	.37	.38	.43	.42	.43	.33	.42	.38	.30	.39	.35	.32	.41	.37
	Post	.38	.42	.40	.35	.36	.35	.36	.44	.40	.37	.46	.41	.34	.41	.37	.34	.48	.41
Cooperation	Pre	.75	.68	.72	.78	.71	.75	.58	.53	.56	.67	.69	.68	.69	.69	.69	.53	.58	.56
	Post	.64	.69	.66	.68	.70	.69	.47	.56	.51	.69	.73	.71	.72	.73	.73	.54	.62	.58
Self-Control	Pre	.70	.68	.69	.59	.59	.59	.72	.70	.71	.60	.69	.63	.52	.59	.56	.61	.69	.65
	Post	.60	.65	.62	.53	.57	.54	.59	.66	.62	.64	.69	.66	.56	.60	.58	.64	.71	.68
							SSRS-T Problem Behavior												
Total	Pre	**-.79**	**-.75**	**-.77**	-.73	-.68	-.70	-.74	-.73	-.73	**-.67**	**-.79**	**-.73**	-.63	-.72	-.67	-.63	-.77	-.70
	Post	**-.71**	**-.71**	**-.71**	-.68	-.65	-.66	-.64	-.67	-.65	**-.77**	**-.76**	**-.77**	-.72	-.69	-.71	-.71	-.76	-.74
External	Pre	-.69	-.71	-.69	-.58	-.61	-.59	-.72	-.73	-.72	-.57	-.76	-.66	-.48	-.67	-.57	-.60	-.78	-.68
	Post	-.63	-.71	-.67	-.56	-.63	-.59	-.62	-.72	-.67	-.69	-.77	-.73	-.60	-.68	-.64	-.71	-.80	-.75
Internal	Pre	-.39	-.36	-.37	-.34	-.26	-.30	-.39	-.44	-.41	-.31	-.37	-.34	-.28	-.29	-.28	-.30	-.43	-.37
	Post	-.35	-.35	-.35	-.33	-.27	-.29	-.33	-.40	-.37	-.37	-.39	-.38	-.34	-.31	-.32	-.35	-.45	-.40
Hyperactivity	Pre	-.75	-.70	-.72	-.74	-.71	-.72	-.63	-.57	-.59	-.65	-.72	-.68	-.65	-.72	-.68	-.55	-.62	-.59
	Post	-.69	-.66	-.67	-.69	-.68	-.68	-.57	-.53	-.55	-.75	-.71	-.72	-.74	-.70	-.72	-.64	-.63	-.63
							SSRS-T Academic												
Total	Pre	**.46**	**.43**	**.45**	.49	.46	.48	.33	.32	.32	**.38**	**.43**	**.40**	.40	.46	.42	.30	.34	.32
	Post	**.40**	**.46**	**.43**	.43	.48	.45	.28	.35	.31	**.39**	**.44**	**.42**	.42	.45	.43	.30	.37	.34

Note. The Academic Engagement scale was the mean of items in Factor 1 (F1), Skills 1 to 7, and the Peer Social Relations scale the mean of items in Factor 2 (F2), Skills 8 to 12. Early and late refer to early elementary Grades 1 to 3 ($n = 426$) and late elementary Grades 4 to 6 ($n = 359$); all refers to all students ($N = 785$). For all students, the sample included 774 students at pretest, 702 at posttest, and 691 with both pretest and posttest data. Cells with correlations between ESBA total or major SSRS domains are presented in bold. All coefficients were statistically significant, $p < .0001$.