Exploring the Dimensionality of the Social Skills Improvement System Using Exploratory Graph Analysis and Bifactor-(*S* − 1) Modeling

Margarita Panayiotou[1] ⬤, **João Santos**[1], **Louise Black**[1] ⬤, and **Neil Humphrey**[1]

Abstract

Since its development over a decade ago, the Social Skills Improvement System (SSIS) has been one of the most widely used measures of social skills in children. However, evidence of its structural validity has been scant. The current study examined the original seven-factor and more recent five-factor structure (SSIS-SEL) of the self-report SSIS in a sample of English elementary school students ($N = 3,331$) aged 8 to 10 years ($M = 8.66$, $SD = 0.59$). A problematic fit was found for both structures with poor discriminant validity. Using exploratory graph analysis and bifactor-(S − 1) modeling, we found support for a four-factor structure, the variation of which was captured by a general factor defined by "empathy and prosocial skills." Future researchers, particularly those interested in using specific domains of the SSIS, are urged to assess its structure in their studies, if their findings are to be theoretically meaningful.

Keywords

social skills improvement system, SSIS, exploratory graph analysis, network analysis, bifactor modeling

Social skills are behaviors needed for the successful completion of social tasks, such as making friends, playing a game, and initiating conversation (Gresham et al., 2010). Children and young people with social skills deficits are at increased risk of experiencing far-reaching difficulties in and out of school. Considerable evidence has shown that such deficits are a key characteristic of developmental and academic difficulties (Blair et al., 2015; Elias & Haynes, 2008; Malecki & Elliot, 2002; Montroy et al., 2014; Rabiner et al., 2015; Racz et al., 2017; Spence, 2003; Verboom et al., 2014; Wentzel, 1991), with their impact extending far into adulthood (Jones et al., 2015).

Social skills can be developed and improved (Gresham & Elliott, 2008) and explicit instruction of social skills is a core component of programs that focuses on the treatment of emotional, behavioral, and developmental disorders (Spence, 2003). Indeed, meta-analyses of multicomponent programs focused on enhancing—among others—the social skills of children and adolescents report reduced behavioral and emotional problems, and more favorable social behaviors, school bonding, and academic attainment (Durlak et al., 2010; Durlak et al., 2011; Taylor et al., 2017). Identifying children with or at risk of social skills deficits using appropriate instruments has thus become a key focus of intervention efforts, given that the assessment of such skills plays a crucial role in the design of appropriate interventions (Elliott et al., 2008).

The Social Skills Improvement System

The Social Skills Improvement System (SSIS; Gresham & Elliott, 2008) is one such instrument. The SSIS represents a comprehensive and improved revision of the widely used Social Skills Rating System (SSRS; for a detailed comparison see Gresham et al., 2011), and it is the first measure to directly link assessment to social skills interventions (Frey et al., 2011). The SSIS provides the opportunity to assess social skills, problematic behaviors and academic competence through ratings from teachers, carers, and students (Gresham & Elliott, 2008). The current study focuses exclusively on the social skills strand of the student self-report SSIS, which purports to assess skills of communication (e.g., *saying please*), cooperation (e.g., *paying attention when others talk*), empathy (e.g., *feeling bad when others are sad*), assertion (e.g., *asking for information*), responsibility (e.g., *careful with other people's belongings*), engagement (e.g., *getting along with others*), and self-control (e.g., *staying calm when teased*).

[1]University of Manchester, Manchester, UK

Corresponding Author:
Margarita Panayiotou, Manchester Institute of Education, University of Manchester, Ellen Wilkinson Building, Oxford Road, Office A5.1, Manchester M13 9PL, UK.
Email: margarita.panayiotou@manchester.ac.uk

The SSIS was shown to be psychometrically superior to the SSRS (Gresham et al., 2011). However, unlike the SSRS, the reduction of items and their assignment into each of the seven domains was not based on statistical methods such as exploratory factor analysis (EFA). Instead, its structure was driven by theory and empirical evidence (Frey et al., 2011). While EFA can be an important and informative step in scale development, it is not uncommon for confirmatory factor analysis (CFA) to be performed instead of EFA when prior theory exists about a measure's structure (Henson & Roberts, 2016). However, when a seven-factor CFA was conducted in the original validation study (Gresham & Elliott, 2008), it resulted in a poor model fit (CFI = mid80s) with modification indices suggesting multiple cross-loadings. Despite this, the structure was not revised or investigated further. The authors report that the "purpose of this analysis was not to test a factor model, but rather to identify possible beneficial changes in subscale composition" (Gresham & Elliott, 2008, p. 51). While such revisions were considered, the authors decided against them as this would have "reduced the number of items loading on each factor, which in turn would have reduced the reliability of the factor" (Gresham & Elliott, 2008, p. 51). While a few studies have explored the measure's reliability and validity (Cheung et al., 2017; Frey et al., 2014; Gamst-Klaussen et al., 2016; Gresham et al., 2010; Gresham et al., 2011; Sherbow et al., 2015), its structural validity remains largely underexplored, with the limited available evidence pointing to a weak structure.

Using a polytomous IRT model Anthony et al. (2016) found 19 items of the teacher-report SSIS to perform poorly. Similarly, while fit was acceptable according to some indices, the comparative fit and Tucker–Lewis indices were below recommended cutoffs in a Chinese sample. The seven subscales were also shown to be redundant, as they failed to contribute sufficient explanatory variance beyond the total score (Wu et al., 2019). In a recent effort to advance social and emotional learning (SEL) measurement using the original standardization data ($N = 224$), Gresham and Elliott (2018) reconfigured the 46-item SSIS into a five-factor structure representing the five competencies of CASEL's (Collaborative for Academic, Social, and Emotional Learning [CASEL], 2008) SEL framework: self-awareness, self-management, social awareness, relationship, and responsible decision making. Despite the promising applications of such a measure, Gresham et al. (2020) found inconsistent model fit (root mean square error of approximation [RMSEA] = .06, comparative fit index [CFI] = .83) and poor discriminant validity for the self-report version, with 8 out of 10-factor correlations exceeding .85. Similar findings were found by Panayiotou, Humphrey, and Wigelsworth (2019) in a sample of 7- to 10-year-old English students, with an inadmissible structure due to poor discriminant validity (factor correlations >1).

The Current Study

Since its development a decade ago, the SSIS has continued to receive increased attention and use in the field of SEL, with over 400 citations; however, the validity of its structure, especially for the self-report version, remains a neglected area of enquiry. This is a significant oversight, given the increased use of self-report assessment in research and clinical practice, in line with policy that focuses on the voice of the child (Deighton et al., 2014; Sturgess et al., 2002). Extending the work by Panayiotou et al. (2019), and as encouraged by the authors of the measure (Frey et al., 2011), the current study aims to examine the structure of the student-report SSIS using secondary analysis of a major data set of English students. The fit of the original seven-factor measure and the newly proposed five-factor SSIS-SEL are assessed, and the structure of the measure is further explored with the use of a new and powerful exploratory analytic technique (exploratory graph analysis [EGA]; Golino & Epskamp, 2017).

Method

Design

The current study is based on a secondary analysis of baseline data drawn from a major randomized trial of a school-based preventive intervention.

Participants

Given that the self-report SSIS was originally validated in a sample of children aged 8 years and older and its readability grade was shown to be 1.8 (Gresham & Elliott, 2008), we excluded any children that were in Grade 1 during the first year of data collection. Of the original sample ($N = 5,218$), the current study included 3,331 students (male; $n = 1,720$, 51.6%) aged 8 to 10 years ($M = 8.66$, $SD = 0.59$). Their characteristics mirrored those of students in state-funded English elementary schools, albeit with larger percentages of students eligible for free school meals (28.6%), speaking English as additional language (21%) and with special educational needs (20.7%; Department for Education, 2012, 2013). Similar to the national picture (Department for Education, 2012), 68.8% of the sample were Caucasian ($n = 2,292$), 11.3% Asian ($n = 376$), 7% Black ($n = 233$), 5.6% mixed ethnicity ($n = 187$), 2.4% other/unclassified ethnicity ($n = 80$), and 0.6% ($n = 20$) Chinese. Ethnic background data were not available for the remaining 143 (4.3%) students.

Measures

The self-report SSIS for ages 8 to 12 years was used in the current study (Gresham & Elliott, 2008). Students are asked

to indicate how true a statement is for them using a 4-point scale ("never," "seldom," "often," "almost always") and the 46 items are summed to represent a total social skills score or seven individual domains of social skills (communication, cooperation, empathy, assertion, responsibility, engagement, and self-control). This version was originally shown to have acceptable internal consistency and test-rest reliability for the overall scale (α = .94, r = .80) and for the seven subscales (α range .72-.80; r range = .58-.79; Gresham & Elliott, 2008).

Procedure

Following approval from the authors' host institution ethics committee, written consent was sought from schools' head teachers. Opt-out consent was sought from parents, and opt-in assent from participating students. Data collection took place between May and July (summer term) of 2012. Data were collected electronically via a secure online survey site. School staff supported any students with literacy difficulties to enable them to complete the measure.

Data Analysis

Existing Structures. The original SSIS structure and SSIS-SEL were tested in M*plus* 8.3 using the weighted least squares with mean and variance adjusted (WLSMV) estimator, as this is optimal with ordinal data using large number of latent factors and large sample sizes (Muthén et al., 2015). Model fit was assessed using multiple indices as generally recommended; specifically the CFI, Tucker–Lewis index (TLI), RMSEA; including 90% confidence intervals [CIs]), and standardized root squared mean residual (SRMR). TLI and CFI values above .95, RMSEA values below .06, and SRMR values below .08 were considered to indicate good model fit (Hu & Bentler, 1999). Modification indices and the residual correlation matrix were also assessed for areas of misfit. Given that students were nested within schools (n = 45; intracluster correlation coefficients = .004-.063), goodness-fit-statistics and standard errors of the parameter estimates were adjusted to account for the dependency in the data (using Type = complex).

New Structure. EGA (Golino et al., 2020; Golino & Epskamp, 2017) is part of network psychometrics, a rapidly developing field that estimates the relationships between observed variables rather than treating them as functions of latent variables (Epskamp et al., 2016). In these models, nodes (circles) represent the observed variables, and edges (lines) represent their level of connection, as partial correlations, after conditioning on all other variables in the model (Epskamp & Fried, 2018). EGA first applies a Gaussian Graphical Model (Lauritzen, 1996) to estimate the network by modelling the inverse of the variance–covariance matrix (Epskamp et al., 2017). Then, using penalized maximum

likelihood estimation (graphical least absolute shrinkage and selection operator [glasso]), the model structure and parameters of a sparse inverse variance–covariance matrix are obtained (Golino & Epskamp, 2017). glasso uses a tuning parameter to minimize the extended Bayesian information criterion, thus estimating the most optimal model fit (Epskamp & Fried, 2018). glasso is useful in avoiding overfitting, by shrinking small partial correlation coefficients, and can therefore also increase the interpretability of network structures (Epskamp et al., 2016). Finally, the walktrap algorithm (Pons & Latapy, 2006) is used to find how many dense subgraphs (clusters) exist in the data. These clusters are considered to be mathematically equivalent to latent variables (Golino & Epskamp, 2017). As with other traditional exploratory factor analytic methods, EGA is data driven and does not rely on the researcher's a priori assumptions or beliefs, making it an ideal technique for testing or reevaluating the structure of a measure (Christensen et al., 2019). Traditional factor analytic methods follow a two-step approach, where the number of dimensions is estimated first, for instance through parallel analysis, followed by EFA with the number of dimensions found in Step 1. EGA on the other hand, offers an advantage over traditional methods, as it follows a single-step approach, thus reducing the number of researcher degrees of freedom and the potential for bias and error (Golino et al., 2020). It was also shown to outperform parallel analysis and minimum average partial procedure, especially in models with highly correlated factors, such as the SSIS (Golino et al., 2020; Golino & Epskamp, 2017).

To examine the structure of the SSIS, the sample was split into two random halves, one for EGA (n = 1,666) and one for CFA (n = 1,665). The EGA was performed using the R package *EGAnet* (version 0.9.3; Golino & Christensen, 2019), which makes use of the *qgraph* package to estimate the polychoric correlations and glasso. The network model and CFA models were visualized using the R packages *qgraph* (version 1.6.3; Epskamp et al., 2012) and *semPlot* (version, 1.1.2; Epskamp et al., 2019), respectively.

CFA, which was used to evaluate the EGA structure found in previous steps, was estimated using M*plus* 8.3. Data were analyzed using WLSMV with pairwise presence, and accounting for the school clustering. Model fit was determined based on the criteria aforementioned. The code for all analyses is provided in the supplementary material.

Results

Seven-Factor SSIS Structure

The correlated seven-factor structure was shown to have acceptable model fit in the current sample (N = 3,331), χ^2 (968) = 2479.394, p < .001; RMSEA = .023 (90% CI [.021, .024]), CFI = .928, TLI = .924; SRMR = .043. While the CFI and TLI values were somewhat below the

Table I. Factor Correlations for the Correlated (Left Diagonal) and Higher Order (Right Diagonal) Seven-Factor Models.

	1	2	3	4	5	6	7
1. Communication	—	.848	.793	.950	.823	.770	.786
2. Cooperation	.901	—	.744	.892	.772	.722	.737
3. Assertion	.745	.639	—	.834	.722	.676	.690
4. Responsibility	.945	1.002	.791	—	.866	.810	.826
5. Empathy	.867	.755	.757	.824	—	.701	716
6. Engagement	.732	.627	.835	.736	.735	—	.669
7. Self-control	.751	.719	.729	.854	.681	.715	—

Note. All correlations were significant at the .001 level.

acceptable thresholds, it is known that these indices can be worsened by having a large number of indicators (Kenny & McCoach, 2003). Acceptable factor loadings (λ range = .42-.76) were observed across most SSIS items. However, a correlation of 1 between responsibility and cooperation resulted in a nonpositive psi matrix. Furthermore, as shown in Table 1, seven pairs of factor correlations indicated poor discriminant validity with $r >$ .80 (Brown, 2015). A nonpositive psi matrix was also observed for the second-order structure due to a negative but small residual variance for the responsibility factor ($\xi = -.006$) and a factor correlation >1 between responsibility and the second-order factor. While this was resolved when the residual variance was fixed to zero, as with the lower order model, eight pairs of factor correlations were shown to be substantially large (see Table 1). The fit of this model, $\chi^2(983) = 2848.616$, $p < .001$; RMSEA = .025 (90% CI [.024, .026]), CFI = .912, TLI = .907, SRMR = .048, and the clear evidence for poor discriminant validity, pointed to a possible misspecified solution.

SSIS-SEL Structure

As a next step, the structure of the newly proposed SSIS-SEL was considered in the same sample. Results pointed to two issues. First, as with the original study (Gresham et al., 2020), this model was shown to have inconsistent fit, with respect to CFI and RMSEA, $\chi^2(979) = 3355.421$, $p < .001$; RMSEA = .028 (90% CI [.027, .029]), $p > .05$; CFI = .888, TLI = .881, SRMR = 053. Strictly following the minimum acceptable cutoffs, it appeared to fit the data well in terms of RMSEA, but poorly when CFI was considered. Given the limitations of fit indices (McNeish et al., 2018), the aim is not to blindly disregard the model, but try to explain why such discrepancy might have occurred. When investigating the residual correlation matrix in the current data, only a relatively small percentage of correlations (9.7%) were equal or greater than .10 (Kline, 2016). This, along with the SRMR index, indicates that the level of misfit is low. In this case, some suggest

that inconsistent model fit may be due to high measurement accuracy (i.e., low unique variances; Browne et al., 2002). However, communalities ranged between .14 and .58 within our data, with the majority (63%) being below .40 (Costello & Osborne, 2005). Accordingly, the unique variance was high for most items (ε = .42-.86) indicating that while the level of misfit is low, many of the items are not meaningfully related to their respective factors. In such instances, further exploration of the factor structure is warranted (Costello & Osborne, 2005).

The second issue was that the psi matrix was again nonpositive definite, caused by the substantially high correlations between the five factors, ranging between .74 (social awareness × self-management) and 1.01 (self-awareness × responsible decision making), with 9 out of 10 correlations exceeding .80. Therefore, the degree to which the five SEL factors represent distinct constructs is questionable. While it would be possible to collapse the highly correlated factors in an effort to improve fit, the model misspecifications noted above could arise from an improper number of factors (Brown, 2015). While CFA relies on a strong theoretical background, as was the case for the development of the SSIS, factor misspecifications "should be unlikely when the proper groundwork for CFA has been conducted" (Brown, 2015, p. 141). Given the lack of exploratory techniques during the initial stages of the SSIS development, we thus sought to further explore the structure of the measure within our data.

New Structure. EGA of the 46 items ($n = 1,666$) resulted in a dense network with four clusters of partial correlations (see Figure 1). CFA of this four-factor EGA structure in the second random half ($n = 1,665$) showed acceptable model fit, notably better than the original seven-factor and five-factor SEL structure, $\chi^2(983) = 1601.943$, $p < .001$, RMSEA = .020, (90% CI [.019, .022]), CFI = .955, TLI = .953, SRMR = .043. The majority of residual correlations approached zero and only 38 (3.7%) were ≥ .10, indicating low misfit. It is worth noting, however, that as seen in Figure 2, 28 of the items had communalities below .40 (with λ < .63) and for seven of these this was below .25 (with λ < .50; Costello & Osborne, 2005). Finally, two-factor correlations were above .80, still indicating issues with discriminant validity. Specifically, F2 was shown to correlate most strongly with all other factors (ψ = .76-.88). The four factors were considered to represent empathy and prosocial skills ([F1] 17 items, ω = .85), engagement and relationship skills ([F2] 16 items, ω = .84), cooperation ([F3] 8 items; ω = .84), and self-control ([F4] 5 items; ω = .79). A comparison between this four-factor and existing seven- and five-factor structures can be found in the appendix.

Post hoc bifactor models. Given the persistent high factor correlations in the four-factor structure, a bifactor model
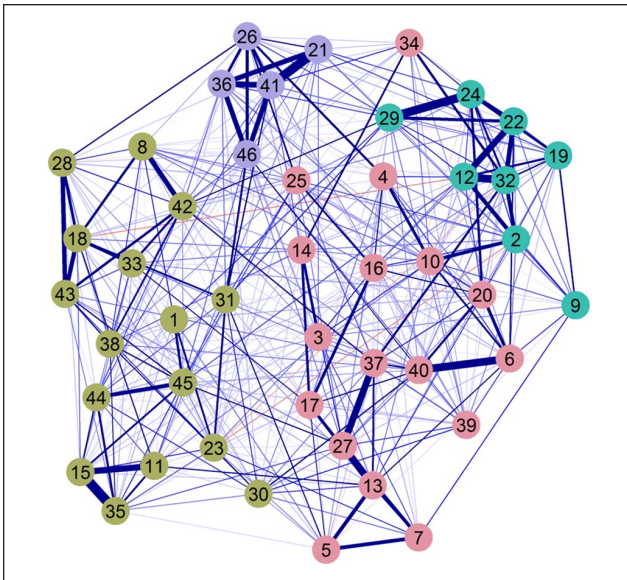
Figure 1. Exploratory graph analysis of the 46 SSIS items, with each color representing one cluster (latent variable).
Note. Nodes (circles) represent observed variables, and edges (lines) represent partial correlations. The magnitude of the partial correlation is represented by the thickness of the edges. SSIS = Social Skills Improvement System; blue edges = positive correlations; red edges = negative correlations.
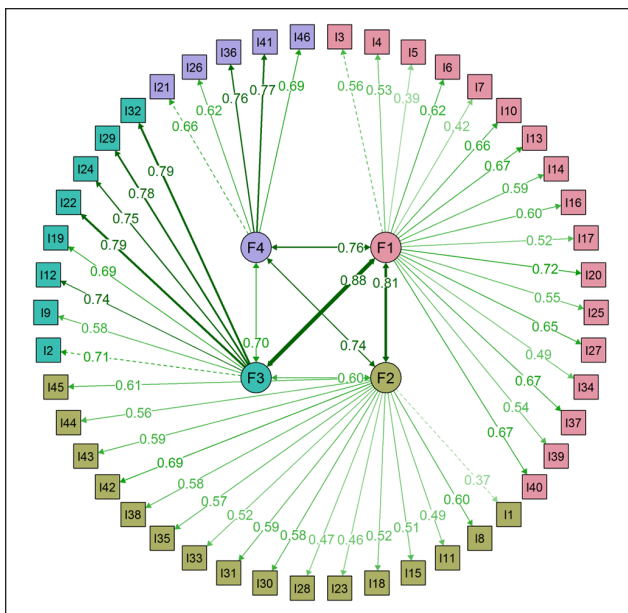

Figure 2. Confirmatory factor analysis of the four-factor EGA structure.
Note. EGA = exploratory graph analysis; F1 = empathy and prosocial skills; F2 = engagement and relationship skills; F3 = cooperation; F4 = self-control.

was examined (Reise et al., 2010). In this model, the covariance between the items can be accounted for by a common

general factor (g) and specific domain factors, alongside measurement error (Reise, 2012). This allowed further exploration of the dimensionality of the SSIS and estimation of the extent to which any differences in social skills are determined by a common factor or by specific components (Bornovalova et al., 2020; Eid et al., 2018). While this model was shown to have a good fit, $\chi^2(943) = 1477.129$, $p < .001$, RMSEA = .019, (90% CI [.017, .021]), CFI = .961, TLI = .957, SRMR = .040, this was not used to guide model selection, given the overfitting issues associated with bifactor modeling (Bonifay et al., 2017; Greene et al., 2019). Most items had moderate to strong factor loadings onto g (λ = .30, .73, $p < .001$, $M\lambda$ = .53) and all items, with the exception of i5, i11, i15, and i35, loaded more strongly onto the general than their respective factors (see Figure 3). This resulted in 4 weak specific factors that each mostly reflected a few indicators (Bornovalova et al., 2020). Once the general factor was accounted for, F1 (empathy and prosocial skills) resulted in an uninterpretable pattern of irregular loadings with a near-zero nonstatistically significant variance (ψ = .004, $p > .05$). This is a common problem occurring in bifactor modeling, whereby a factor vanishes and the general factor becomes specific to the set of items for which the factor vanished (Eid et al., 2017; Eid et al., 2018; Geiser et al., 2015). In such cases, applying a bifactor-$(S-1)$ model is recommended, as these models avoid estimation problems and provide a clear interpretation of the g factor (Eid et al., 2017). Bifactor-$(S-1)$ is a reconfiguration of the classical bifactor model, where one specific factor is omitted. In this model, the general factor is defined by the omitted (reference) factor and the specific factors capture variation in the items that are not accounted for by the general factor (Eid et al., 2018).

A bifactor-$(S-1)$ with empathy and prosocial skills (F1) as the general reference factor resulted in acceptable fit $\chi^2(960) = 1583.320$, $p < .001$, RMSEA = .021, (90% CI [.019, .022]), CFI = .955, TLI = .951, SRMR = .043 (see Figure 3). The variances of the three factors were positive and statistically significant, though very small variances were observed for F2 (ψ = .05) and F3 (ψ = .08), due to the small factor loadings of the reference indicators. It is important to note that while for simplicity the first indicator is typically used for the identification of the latent factor, in theory, the variance of F2 and F3 could take any value between .02 to .22, and .05 to .27, respectively, depending on the choice of reference indicator. Out of the 29 items in the three specific factors, 26 (90%) loaded more strongly on the general empathy and prosocial skills factor (GEP). Most items had factor loadings above the minimum threshold of .30-.40 (Brown, 2015) on the specific factors, though only 11 were shown to exceed .40 (see Table 2). Thus, it is noteworthy, that while all factor loadings were, as expected, positive and statistically significant, small factor loadings were observed for some items on all three specific factors.
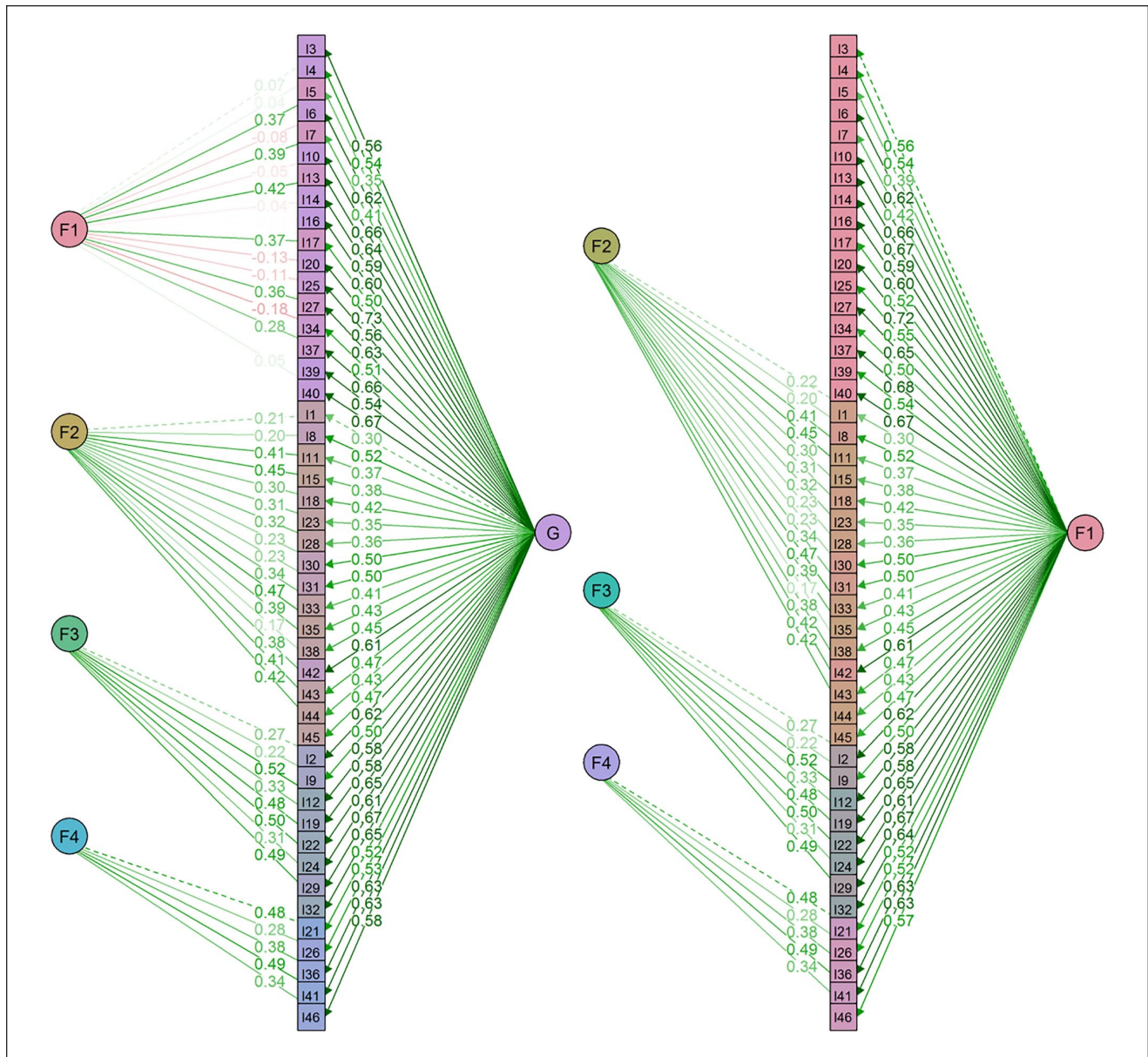
Figure 3. Classical bifactor and bifactor-(S − 1) models of the SSIS.
Note. SSIS = Social Skills Improvement System.

Key bifactor indices were also computed, using the R package *BifactorIndicesCalculator* (version 0.2.0; Dueber, 2020). Given that these were originally developed for the classical bifactor model, only the following that could be extended to bifactor-(S − 1) were considered (Rodriguez et al., 2016). The explained common variance represents the proportion of common variance that is due to g (Rodriguez et al., 2016). For the specific factors, this is interpreted as the proportion of common variance of the items in *a specific factor* that is due to g. Omega (ω) internal consistency estimates the *combined* reliability of g and specific factors, while *hierarchical* omega (ω_h) represents

the reliability of a factor after controlling for the variance due to g (Rodriguez et al., 2016). Construct reliability (H) provides the variance of the factor that is accounted for by the items and can also be interpreted as a reliability coefficient (Hancock & Mueller, 2001). Factor determinacy (FD) represents the correlation between factor scores and the factors, with values closer to 1 indicating better determinacy (Grice, 2001). Finally, the consistency and specificity of the bifactor-(S − 1) items were considered. Consistency provides the proportion of the true score variance of a nonreference item that is determined by the reference factor, while the sensitivity of an item estimates its true score variance

Table 2. SSIS Bifactor-(S − 1) Factor Loadings and Indices.

	λ_G	λ_S	Con	Spe	ECV	ω	ω_h	H	FD
Reference g: Empathy and prosocial ($\psi = .31$)					.77	.96	.90	.95	.97
i3: Forgive	.56								
i4: Careful with other's belongings	.54								
i5: Stand up for others	.39								
i6: Say please	.62								
i7: Feel bad for others	.42								
i10: Take turns	.66								
i13: Make others feel better	.67								
i14: Do my part	.59								
i16: Look at people	.60								
i17: Help friends	.52								
i20: Polite	.72								
i25: Self-awareness	.55								
i27: Think of other's feelings	.65								
i34: Do homework	.50								
i37: Nice to others	.68								
i39: Keep promises	.54								
i40: Say thank you	.67								
F2: Engagement and relationship skills ($\psi = .05$)					.63	.87	.32	.69	.83
i1: Ask for things	.30	.22	.92	.08					
i8: Get along	.52	.20	.95	.05					
i11: Show how I feel	.37	**.41**	.56	.44					
<u>i15: Say I have a problem</u>	.38	**.45**	.54	.46					
i18: Make friends	.42	.30	.83	.17					
i23: Invite others	.36	**.31**	.77	.23					
i28: Talk to new friends	.36	**.32**	.75	.25					
i30: Smile or wave	.50	.23	.93	.07					
i31: End disagreement	.50	.23	.92	.08					
i33: Play games	.41	**.34**	.77	.23					
<u>i35: Say when not treated well</u>	.43	**.47**	.53	.47					
i38: Ask to join	.45	**.39**	.70	.30					
i42: Work well with others	.62	.17	.97	.03					
i43: Make new friends	.47	**.38**	.72	.28					
i44: Tell people-mistakes	.44	**.42**	.64	.36					
i45: Ask for help	.47	**.42**	.69	.31					
F3: Cooperation ($\psi = .08$)									
i2: Listen to others	.62	.27	.91	.09	.69	.90	.26	.63	.83
i9: Ignore naughty	.50	.22	.95	.05					
i12: Do what teacher asks	.58	**.52**	.62	.38					
i19: Do work	.58	**.33**	.86	.14					
i22: Follow school rules	.65	**.48**	.69	.31					
i24: Behaved	.61	**.50**	.66	.34					
i29: Do right thing	.67	**.31**	.90	.10					
i32: Listen to teacher	.65	**.49**	.68	.32					
F4: Self-control ($\psi = .23$)					.67	.83	.27	.50	.77
i21: Stay calm-teased	.52	**.48**	.50	.50					
i26: Stay calm-mistakes	.53	.28	.87	.13					
i36: Stay calm-problems	.63	**.38**	.73	.27					
i41: Stay calm-bothered	.63	**.49**	.56	.44					
i46: Stay calm-disagreement	.58	**.34**	.79	.21					

Note. The item numbering corresponds to the SSIS Rating Scales, and descriptions reflect the content of the items but are abbreviated to avoid copyright violations. In bold are items with factor loadings >.30. Underlined are items that load more strongly onto their own domain. SSIS = Social Skills Improvement System; λ_G and λ_S = factor loadings on the general and specific factors, respectively; Con = consistency; Spe = specificity; ECV = explained common variance; ω_h = hierarchical omega reliability, H = construct reliability; FD = factor determinacy. All factor loadings were statistically significant ($p < .001$).

that is not determined by the reference factor (1-consistency; Eid et al., 2017).

Omega and H values >.70 were considered acceptable (Hancock & Mueller, 2001; Rodriguez et al., 2016), and only factors with FD > .90 were considered reliable for using factor scores as their proxy (Gorsuch, 1983). Results are summarized in Table 2. The explained common variance and omega reliability coefficients indicated that the majority of the variance in the SSIS was explained by GEP. Overall, GEP met the recommended thresholds for omega reliability, FD and H in our sample, but this was not the case for the three specific factors, which were shown to explain very little variance. Notably, while the omega reliability coefficient was shown to be high for the specific factors (ω = .83-.90), this was substantially lower once the variance associated with GEP was partitioned out (ω_h = .26-.32). Results were further supported by the consistency and specificity of the nonreference items, which showed that 50% to 97% of their variance was accounted for by GEP.

Discussion

The aim of the current study was to assess, for the first time, the dimensionality of the self-report SSIS in a sample of English elementary school students. Both the current and original studies, as well as that for the revised SSIS-SEL, found a poor fitting model with low CFI values (Gresham & Elliott, 2008; Gresham et al., 2020; Panayiotou et al., 2019). While, in the current study, this might have been the result of low communalities, our evaluation of the SSIS structure did not rely exclusively on fit index cutoffs, as good practice suggests (McNeish et al., 2018). In fact, the most important finding of the original studies, as well as the current one, is the measure's poor discriminant validity. In many instances, the correlations between factors neared or exceeded 1, indicating that they generally fail to assess distinct constructs (Brown & Moore, 2012). This was in line with findings by Wu et al. (2019), who found that the seven factors were redundant and did not explain much variance beyond the total score. This issue is of relevance particularly when researchers are interested in specific sets of skills, and are using the SSIS factors as independent stand-alone scales. In fact, the findings reported here challenge the idea that these factors can be used to measure seven distinct constructs.

From a statistical perspective, excessively high factor correlations might be the result of overfactoring (Brown & Moore, 2012). This was supported in our sample of English students, where robust analyses pointed to a four-factor solution. These findings are, to a great extent, unsurprising, given that exploratory analysis of the SSIS and SSIS-SEL was not conducted during development. While it is true that EGA is a data-driven approach and, as the authors note, "CFA is most appropriate for theory testing rather than for theory generation as is done in exploratory factor analysis" (Gresham et al., 2020, p. 197), their stance can become highly problematic under certain conditions. When said CFA techniques result in poor model fit and statistically indistinguishable factors, this might be the result of improper number of factors due to the lack of robust exploratory groundwork (Brown, 2015). At this stage, model revision based on robust exploratory methods might be more appropriate.

Additionally, premature use of CFA might be inappropriate with measures that are based on inconsistent conceptualization of the construct under study. As many researchers have noted, the area of social and emotional development suffers from "jingle and jangle fallacies," where different definitions are used to assess the same skills, resulting in great measurement challenges (Abrahams et al., 2019; Jones et al., 2016). Social functioning, for instance has been considered an umbrella term for "social competence" and "social skills," but other times these have been used interchangeably (Cordier et al., 2015). Future researchers are therefore urged to take such challenges into consideration and test the structure of the measure in their own sample prior to testing structural differences. In the current sample, most items of the original "engagement" and "assertion" domains clustered together and were considered to represent "engagement and relationship skills." Its factor loadings were, however, varied (λ = .37-.70; h^2 = .22-.49), suggesting that not all items are well explained by this factor, potentially also leading to issues with sum scores (McNeish & Wolf, 2020). The remaining three factors in our four-factor structure were considered to represent "self-control" (e.g., *stay calm when teased*), "cooperation" (e.g., *well behaved*), and "empathy and prosocial skills" (e.g., *make others feel better*). While the four-factor structure was shown to have a better model fit than the original one, issues with discriminant validity remained. Specifically, the empathy and prosocial skills factor was shown to correlate very highly with all domains (ρ = .76-.88), further questioning the dimensionality of this construct.

Though the aim of the current study was not to revise the self-report SSIS, EGA provides a unique account of the relationships between the 46 items and allows for a deeper understanding of possible problematic areas in the structure. While the dense network of the current study (Figure 1) resulted in four clusters, a few of the strong correlations between items of different domains (e.g., i2 × i10), and items that deviate from their cluster (e.g., i30, i34), could explain the poor discriminant validity observed in the current structure (the full partial correlation matrix is provided in Table S1, Supplementary material, available online). EGA can thus provide a detailed representation of how the items from each cluster relate to one another, but also how these clusters are placed with each other in multidimensional space (Christensen et al., 2019).

Concerns about the measure's unclear dimensionality were confirmed by the post hoc bifactor findings. While bifactor models have received increased attention within psychology in the past decade, their application within social skills has been scant. Our analyses were therefore necessarily exploratory. Consistent with the greater literature, the classical bifactor structure for the SSIS resulted in inadmissible results (Eid et al., 2017). Specifically, in the presence of g, the factor loadings of F1 were uninterpretable and irregular, causing the specific factor to vanish. Within stochastic measurement theory such findings are expected when classical bifactor models are applied to structures with noninterchangeable (fixed) domains, such as those studied here (Eid et al., 2017). Bifactor-$(S - 1)$, on the other hand, has proved promising, and been applied to many psychological constructs, without the problems inherent to classical bifactor models (e.g., Black et al., 2019; Burns et al., 2020; Eid, 2020; Heinrich et al., 2018). While it is advised that the choice of reference factor in bifactor-$(S - 1)$ is based on theory and/or ease of interpretation (Burns et al., 2020; Eid, 2020), our decision was empirically driven: F1 was shown to drive the high factor correlations in the correlated-factor model and it vanished in the classical bifactor model. Given that the meaning of g varies depending on the choice of reference factor (Burke & Johnston, 2020; Burns et al., 2020), results should be interpreted with caution, as generalizability cannot be assumed.

The empirical collapse of F1 in this study fits with the work by Zachrisson et al. (2018) who treated "prosocial behavior" as their reference factor in the Lamer Social Competence in Preschool scale, under the assumption that this is the most aligned domain with overall social competence. Within our sample, social skills were captured by GEP and the three specific factors were considered to represent systematic variation among items that cannot be captured by the reference factor (Eid et al., 2018). It is therefore important to remember that neither the reference nor the specific factors represent overall social skills, as this study initially set out to explore. Rather, the specific factors indicate that a child exhibits more or less self-control, cooperation, and engagement and relationship skills, than one would expect given his or her levels of GEP. This aligns with theoretical and empirical evidence that suggests empathy is a key driver of prosocial behavior (Decety et al., 2016), and that prosocial behavior is directly related to self-regulatory behaviors, social interactions and general social competence (e.g., Andrade et al., 2014; Dewall et al., 2008; Spinard & Eisenberg, 2009). Indeed, in the current sample 50% to 97% of the variance in the items was captured by GEP, while the specific factors explained very little variance and were shown to be psychometrically unfit. Additionally, it is important to note that given the small factor loadings, the meaning captured by the specific factors becomes fundamentally different (Bornovalova et al., 2020). F3 for example, no longer captures general cooperation skills, but cooperation within school and classroom, if one were to use a threshold of $>.40$ for meaningful factor loadings.

Overall, our findings and those reported elsewhere (Gresham & Elliott, 2008; Gresham et al., 2020; Panayiotou et al., 2019) suggest that currently the SSIS is unable to meaningfully capture distinct domains of social skills (or social–emotional competence in the case of SSIS-SEL). This does not necessarily mean, however, that the poor discriminant validity of the measure is the result of a common general factor. Positive manifolds between symptoms and behaviors mathematically fit bifactor models but might represent processes other than a common cause (van Bork et al., 2017). For instance, the positive manifold observed in the current study might be the result of measurement problems (van der Maas et al., 2006). Based on sampling theory (Bartholomew et al., 2009), it is possible that it may be very difficult to obtain independent measures of different groups of social skills, as these rely on the same underlying behaviors (van Bork et al., 2017; van der Maas et al., 2006). This has been true for all the SSIS structures examined in the current sample, as it was proven difficult to obtain structurally independent domains of social skills. This overlap could be because different sets of skills are tapping onto the same underlying processes, in this case empathy and prosocial behavior. Given that the SSIS was developed through focus groups with professionals and teachers (Gresham & Elliott, 2008), one must also consider whether this overlap is caused by a gap between young children and scale developers in their ability to differentiate between such highly overlapped items. Thus, before concluding that empathy and prosocial skills can sufficiently explain the covariance between the SSIS items, more work is needed to understand how to disentangle these overlapping skills both conceptually and statistically. Indeed, some of the difficulties reported in the current study are consistent with general problems in the conceptualization of social–emotional learning. This suffers from inconsistent and variable scope and psychometric properties (Humphrey et al., 2011; Wigelsworth et al., 2010). However, as this is the first study to explore the structure of the self-report SSIS, more work is needed to consolidate the results reported herein. Until said work is carried out, for researchers using the self-report SSIS that are interested in specific domains of social skills, our findings suggest that a four-factor structure (see Figure 2), might be more appropriate than the original seven-factor structure. However, given that in our study two of the domains (F1 × F3) indicated substantial overlap, we urge researchers to explore the factor intercorrelations in their own data before accurate conclusions can be drawn.

Strengths and Limitations

Despite being a very widely used measure, the psychometric properties of the SSIS have been a neglected area of inquiry. The current study is the first to explore the structure of the self-report SSIS since its development a decade ago. While the aim of this study was not to revise the SSIS, the application of robust analyses such as the EGA and bifactor-$(S-1)$ allowed us not only to explore the psychometric performance of the SSIS but also possibly shed light on our understanding of social skills more generally. Results from the current study provide robust evidence that currently the SSIS is not fit for the assessment of distinct domains of social skills. Although it is possible that results are culturally specific, the poor psychometric evidence for the seven-factor structure in the current study matched that of the original U.S. standardization sample. Our study is the first to validate the SSIS in a sample of English children, making a significant contribution on the replicability of its structure, further suggesting that even in English-speaking countries, the cultural transferability of the SSIS cannot be assumed (Humphrey et al., 2011). Given, however, that the current study relied exclusively on self-report data, findings cannot be generalized to other SSIS informant types. Future work is thus urgently needed to replicate the results of the current study to the parent and teacher forms and in different cultures and samples. Additionally, given recent findings that glasso might be less powerful in reducing false-positive rates (Williams & Rast, 2020), future work should consider adapting EGA with other nonregularized methods. Finally, any conclusions drawn are limited to the current sample, given that the classical bifactor model is sensitive to sampling (Bornovalova et al., 2020) and the meaning of bifactor-$(S-1)$ is conditional on the chosen reference factor. Nevertheless, findings from attention-deficit/hyperactivity disorder and oppositional defiant disorder research suggest that consistent results are a possibility within bifactor-$(S-1)$ (Eid, 2020). Therefore, more work is needed to identify whether there is a consistently outstanding general domain of social skills that could be used within clinical assessment.

Conclusion

The findings of the current study add new and robust evidence about the psychometric quality and, specifically, the structural validity of the SSIS measure. In the first validation study of the SSIS in an English sample of elementary school students, the current study rigorously demonstrated that the proposed seven- and five-factor structures of the SSIS are problematic and the 46 items are better represented by a four-factor structure, that are captured through a general reference factor of empathy and prosocial skills. Future researchers, especially those interested in using distinct domains of the SSIS, should consider using the four-factor structure found here, but are also urged to confirm this structure in their own sample, if their findings are to be theoretically meaningful. A better structure on the SSIS could improve the assessment and monitoring of children's social skills and deficits, and "ultimately contribute to their well-being, resiliency, and achievement of adaptive outcomes" (Abrahams et al., 2019, p. 468).

Appendix

Table A1. The Existing and Newly Proposed Structures of the SSIS.

Items	SSIS factors		
	EGA four-factor	Seven-factor	Five-factor SEL
i1	Engagement and relationship skills	Assertion	Relationship skills
i2	Cooperation	Cooperation	Self-awareness
i3	Empathy and prosocial	Empathy	Relationship skills
i4	Empathy and prosocial	Responsibility	Responsible decision making
i5	Empathy and prosocial	Assertion	Social awareness
i6	Empathy and prosocial	Communication	Self-awareness
i7	Empathy and prosocial	Empathy	Social awareness
i8	Engagement and relationship skills	Engagement	Relationship skills
i9	Cooperation	Cooperation	Self-management
i10	Empathy and prosocial	Communication	Relationship skills
i11	Engagement and relationship skills	Assertion	Social awareness
i12	Cooperation	Cooperation	Relationship skills
i13	Empathy and prosocial	Empathy	Social awareness
i14	Empathy and prosocial	Responsibility	Self-awareness

(continued)

Table A1. (continued)

	SSIS factors		
Items	EGA four-factor	Seven-factor	Five-factor SEL
i15	Engagement and relationship skills	Assertion	Self-awareness
i16	Empathy and prosocial	Communication	Relationship skills
i17	Empathy and prosocial	Empathy	Social awareness
i18	Engagement and relationship skills	Engagement	Relationship skills
i19	Cooperation	Cooperation	Self-management
i20	Empathy and prosocial	Communication	Self-awareness
i21	Self-control	Self-control	Self-management
i22	Cooperation	Cooperation	Responsible decision making
i23	Engagement and relationship skills	Engagement	Relationship skills
i24	Cooperation	Responsibility	Self-awareness
i25	Empathy and prosocial	Assertion	Self-awareness
i26	Self-control	Self-control	Self-management
i27	Empathy and prosocial	Empathy	Social awareness
i28	Engagement and relationship skills	Engagement	Relationship skills
i29	Cooperation	Responsibility	Responsible decision making
i30	Engagement and relationship skills	Communication	Relationship skills
i31	Engagement and relationship skills	Self-control	Self-management
i32	Cooperation	Cooperation	Self-management
i33	Engagement and relationship skills	Engagement	Relationship skills
i34	Empathy and prosocial	Responsibility	Responsible decision making
i35	Engagement and relationship skills	Assertion	Self-awareness
i36	Self-control	Self-control	Self-management
i37	Empathy and prosocial	Empathy	Social awareness
i38	Engagement and relationship skills	Engagement	Relationship skills
i39	Empathy and prosocial	Responsibility	Responsible decision making
i40	Empathy and prosocial	Communication	Relationship skills
i41	Self-control	Self-control	Self-management
i42	Engagement and relationship skills	Cooperation	Relationship skills
i43	Engagement and relationship skills	Engagement	Relationship skills
i44	Engagement and relationship skills	Responsibility	Responsible decision making
i45	Engagement and relationship skills	Assertion	Self-awareness
i46	Self-control	Self-control	Self-management

Note. The item numbering corresponds to the SSIS Rating Scales, and descriptions reflect the content of the items but are abbreviated to avoid copyright violations. SSIS = Social Skills Improvement System; EGA = exploratory graph analysis; SEL = social and emotional learning.

ORCID iDs

Margarita Panayiotou https://orcid.org/0000-0002-6023-7961
Louise Black https://orcid.org/0000-0001-8140-3343

Supplemental Material

Supplemental material for this article is available online.

References

Abrahams, L., Pancorbo, G., Primi, R., Santos, D., Kyllonen, P., John, O. P., & De Fruyt, F. (2019). Social-emotional skill assessment in children and adolescents: Advances and challenges in personality, clinical, and educational contexts. *Psychological Assessment*, *31*(4), 460-473. http://dx.doi.org/10.1037/pas0000591

Andrade, B. F., Browne, D. T., & Tannock, R. (2014). Prosocial skills may be necessary for better peer functioning in children with symptoms of disruptive behavior disorders. *PeerJ*, *2*, Article e487. http://dx.doi.org/10.7717/peerj.487

Anthony, C. J., DiPerna, J. C., & Lei, P. W. (2016). Maximizing measurement efficiency of behavior rating scales using item response theory: An example with the Social Skills Improvement System—Teacher rating scale. *Journal of School Psychology*, *55*(April), 57-69. http://dx.doi.org/10.1016/j.jsp.2015.12.005

Bartholomew, D. J., Deary, I. J., & Lawn, M. (2009). A new lease of life for Thomson's bonds model of intelligence. *Psychological Review*, *116*, 567-579. http://dx.doi.org/10.1037/a0016262

Black, L., Panayiotou, M., & Humphrey, N. (2019). The dimensionality and latent structure of mental health difficulties and wellbeing in early adolescence. *PLOS ONE*, *14*(2), Article e0213018. http://dx.doi.org/10.1371/journal.pone.0213018

Blair, B. L., Perry, N. B., O'Brien, M., Calkins, S. D., Keane, S. P., & Shanahan, L. (2015). Identifying developmental cascades among differentiated dimensions of social competence and emotion regulation. *Developmental Psychology*, *51*(8), 1062-1073. http://dx.doi.org/10.1037/a0039472

Bonifay, W., Lane, S. P., & Reise, S. P. (2017). Three concerns with applying a bifactor model as a structure of psychopathology. *Clinical Psychological Science*, *5*(1), 184-186. http://dx.doi.org/10.1177/2167702616657069

Bornovalova, M. A., Choate, A. M., Fatimah, H., Petersen, K. J., & Wiernik, B. M. (2020). Appropriate use of bifactor analysis in psychopathology research: Appreciating benefits and limitations. *Biological Psychiatry*, *88*(1), 18-27. http://dx.doi.org/10.1016/j.biopsych.2020.01.013

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press.

Brown, T. A., & Moore, M. T. (2012). Confirmatory factor analysis. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 361-379). Guilford Press.

Browne, M. W., MacCallum, R. C., Kim, C.-T., Andersen, B. L., & Glaser, R. (2002). When fit indices and residuals are incompatible. *Psychological Methods*, *7*(4), 403-421. http://dx.doi.org/10.1037/1082-989X.7.4.403

Burke, J. D., & Johnston, O. G. (2020). The bifactor S-1 model: A psychometrically sounder alternative to test the structure of ADHD and ODD? *Journal of Abnormal Child Psychology*, *48*(7), 911-915. http://dx.doi.org/10.1007/s10802-020-00645-4

Burns, G. L., Geiser, C., Servera, M., Becker, S. P., & Beauchaine, T. P. (2020). Application of the Bifactor S − 1 model to multisource ratings of ADHD/ODD symptoms: An appropriate bifactor model for symptom ratings. *Journal of Abnormal Child Psychology*, *48*(7), 881-894. http://dx.doi.org/10.1007/s10802-019-00608-4

Cheung, P. P., Siu, A. M., & Brown, T. (2017). Measuring social skills of children and adolescents in a Chinese population: Preliminary evidence on the reliability and validity of the translated Chinese version of the Social Skills Improvement System-Rating Scales (SSIS-RS-C). *Research in Developmental Disabilities*, *60*(January), 187-197. http://dx.doi.org/10.1016/j.ridd.2016.11.019

Christensen, A. P., Gross, G. M., Golino, H. F., Silvia, P. J., & Kwapil, T. R. (2019). Exploratory graph analysis of the multidimensional schizotypy scale. *Schizophrenia Research*, *206*(1), 43-51. http://dx.doi.org/10.1016/j.schres.2018.12.018

Collaborative for Academic, Social, and Emotional Learning [CASEL] (2008). Social and emotional learning (SEL) and student benefits: Implications for the safe schools/healthy students core elements. Retrieved from: https://eric.ed.gov/?id=ED505369

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

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, *10*(7), 1-9. https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1156&context=pare

Decety, J., Bartal, I. B., Uzefovsky, F., & Knafo-Noam, A. (2016). Empathy as a driver of prosocial behaviour: Highly conserved neurobehavioural mechanisms across species. *Philosophical Transactions of the Royal Society B*, *371*(1686), 1-11. http://dx.doi.org/10.1098/rstb.2015.0077

Deighton, J., Croudace, T., Fonagy, P., Brown, J., Patalay, P., & Wolpert, M. (2014). Measuring mental health and wellbeing outcomes for children and adolescents to inform practice and policy: A review of child self-report measures. *Child and Adolescent Psychiatry and Mental Health*, *8*, Article 14. http://dx.doi.org/10.1186/1753-2000-8-14

Department for Education. (2012). *Schools, pupils, and their characteristics, January 2012 (SFR10/2012)*. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/219260/sfr10-2012.pdf

Department for Education. (2013). *Special educational needs in England, January 2013 (SFR 30/2013)*. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/225699/SFR30-2013_Text.pdf

Dewall, C. N., Baumeister, R. F., Gailliot, M. T., & Maner, J. K. (2008). Depletion makes the heart grow less helpful: Helping as a function of self-regulatory energy and genetic relatedness. *Personality and Social Psychology Bulleting*, *34*(12), 1653-1662. http://dx.doi.org/10.1177/0146167208323981

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*(1), 405-432. http://dx.doi.org/10.1111/j.1467-8624.2010.01564.x

Durlak, J. A., Weissberg, R. P., & Pachan, M. (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. *American Journal of Community Psychology*, *45*(3-4), 294-309. http://dx.doi.org/10.1007/s10464-010-9300-6

Eid, M. (2020). Multi-faceted constructs in abnormal psychology: Implications of the bifactor S − 1 model for individual clinical assessment. *Journal of Abnormal Child Psychology*, *48*(7), 895-900. http://dx.doi.org/10.1007/s10802-020-00624-9

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods*, *22*(3), 541-562. http://dx.doi.org/10.1037/met0000083

Eid, M., Krumm, S., Koch, T., & Schulze, J. (2018). Bifactor models for predicting criteria by general and specific factors: Problems of nonidentifiability and alternative solutions. *Journal*

of Intelligence, *6*(3), Article 42. http://dx.doi.org/10.3390/jintelligence6030042

Elias, M. J., & Haynes, N. M. (2008). Social competence, social support, and academic achievement in minority, low-income, urban elementary school children. *School Psychology Quarterly*, *23*(4), 474-495. http://dx.doi.org/10.1037/1045-3830.23.4.474

Elliott, S. N., Gresham, F. M., Frank, J. L., & Beddow, P. A., III. (2008). Intervention validity of social behavior rating scales: Features of assessments that link results to treatment plans. *Assessment for Effective Intervention*, *34*(1), 15-24. http://dx.doi.org/10.1177/1534508408314111

Epskamp, S., Cramer, A. O., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, *48*(4), 1-8. http://dx.doi.org/10.18637/jss.v048.i04

Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, *23*(4), 617-634. http://dx.doi.org/10.1037/met0000167

Epskamp, S., Maris, G. K. J., Waldorp, L. J., & Borsboom, D. (2016). Network psychometrics. In P. Irwing, D. Hughes, & T. Booth (Eds.), *Handbook of psychometrics* (pp. 953-986). Wiley.

Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika*, *82*(4), 904-927. http://dx.doi.org/10.1007/s11336-017-9557-x

Epskamp, S., Stuber, S., Nak, J., Veenman, M., & Jorgensen, T. D. (2019). *semPlot: Path diagrams and visual analysis of various SEM packages' output*. https://CRAN.R-project.org/package=semPlot

Frey, J. R., Elliott, S. N., & Gresham, F. M. (2011). Preschoolers' social skills: Advances in assessment for intervention using social behavior ratings. *School Mental Health*, *3*(4), 179-190. https://doi.org/10.1007/s12310-011-9060-y

Frey, J. R., Elliott, S. N., & Kaiser, A. P. (2014). Social skills intervention planning for preschoolers: Using the SSiS-Rating Scales to identify target behaviors valued by parents and teachers. *Assessment for Effective Intervention*, *39*(3), 182-192. https://doi.org/10.1177/1534508413488415

Gamst-Klaussen, T., Rasmussen, L.-M. P., Svartdal, F., & Stromgren, B. (2016). Comparability of the social skills improvement system to the social skills rating system: A Norwegian study. *Scandinavian Journal of Educational Research*, *60*(1), 20-31. http://dx.doi.org/10.1080/00313831.2014.971864

Geiser, C., Bishop, J., & Lockhart, G. (2015). Collapsing factors in multitrait-multimethod models: Examining consequences of a mismatch between measurement design and model. *Frontiers in Psychology*, *6*, Article 946. http://dx.doi.org/10.3389/fpsyg.2015.00946

Golino, H., & Christensen, A. P. (2019). *EGAnet: Exploratory graph analysis: A framework for estimating the number of dimensions in multivariate data using network psychometrics*. https://cran.r-project.org/web/packages/EGAnet/index.html

Golino, H., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Sadana, R., Thiyagarajan, J. A., & Martinez-Molina, A. (2020). Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *Psychological Methods*, *25*(3), 292-320. http://dx.doi.org/10.1037/met0000255

Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PLOS ONE*, *12*(6), Article e0174035. http://dx.doi.org/10.1371/journal.pone.0174035

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Lawrence Erlbaum.

Greene, A. L., Eaton, N. R., Li, K., Forbes, M. K., Krueger, R. F., Markon, K. E., Waldman, I. D., Cicero, D. C., Conway, C. C., Docherty, A. R., Fried, E. I., Ivanova, M. Y., Jonas, K. G., Latzman, R. D., Patrick, C. J., Reininghaus, U., Tackett, J. L., Wright, A. G. C., & Kotov, R. (2019). Are fit indices used to test psychopathology structure biased? A simulation study. *Journal of Abnormal Psychology*, *128*(7), 740-764. http://dx.doi.org/10.1037/abn0000434

Gresham, F. M., & Elliott, S. N. (2008). *Social Skills Improvement System (SSIS) rating scales manual*. Pearson.

Gresham, F. M., & Elliott, S. N. (2018). *Social skills improvement system social-emotional learning edition manual*. Pearson.

Gresham, F. M., Elliott, S. N., Cook, C. R., Vance, M. J., & Kettler, R. (2010). 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*(1), 157-166. http://dx.doi.org/10.1037/a0018124

Gresham, F. M., Elliott, S. N., Metallo, S., Byrd, S., Wilson, E., Erickson, M., Cassidy, K., & Altman, R. (2020). Psychometric fundamentals of the social skills improvement system: Social–emotional learning edition rating forms. *Assessment for Effective Intervention*, *45*(3), 194-209. http://dx.doi.org/10.1177/1534508418808598

Gresham, F. M., Elliott, S. N., Vance, M. J., & Cook, C. R. (2011). Comparability of the Social Skills Rating System to the Social Skills Improvement System: Content and psychometric comparisons across elementary and secondary age levels. *School Psychology Quarterly*, *26*(1), 27-44. http://dx.doi.org/10.1037/a0022662

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods*, *6*(4), 430-450. http://dx.doi.org/10.1037/1082-989X.6.4.430

Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. d. Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future: A Festschrift in honor of Karl Jöreskog* (pp. 195-216). Scientific Software International.

Heinrich, M., Zagorscak, P., Eid, M., & Knaevelsrud, C. (2018). Giving g a meaning: An application of the bifactor-(S-1) approach to realize a more symptom-oriented modeling of the Beck Depression Inventory-II. *Assessment*, *27*(7), 1429-1447. http://dx.doi.org/10.1177/1073191118803738

Henson, R. K., & Roberts, J. K. (2016). Use of exploratory factor analysis in published research. *Educational and Psychological Measurement*, *66*(3), 393-416. http://dx.doi.org/10.1177/0013164405282485

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

Humphrey, N., Hennessey, A., Lendrum, A., Wigelsworth, M., Turner, A., Panayiotou, M.,... Calam, R. (2018). The PATHS curriculum for promoting social and emotional well-being among children aged 7 9 years: A cluster RCT. *Public Health Research*, *6*(10), 1–116. https://doi.org/10.3310/phr06100

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. https://doi.org/10.1177/0013164410382896

Jones, D. E., Greenberg, M., & Crowley, M. (2015). Early social-emotional functioning and public health: The relationship between kindergarten social competence and future wellness. *American Journal of Public Health*, *105*(11), 2283-2290. http://dx.doi.org/10.2105/AJPH.2015.302630

Jones, S. M., Zaslow, M., Darling-Churchill, K. E., & Halle, T. G. (2016). Assessing early childhood social and emotional development: Key conceptual and measurement issues. *Journal of Applied Developmental Psychology*, *45*(1), 42-48. http://dx.doi.org/10.1016/j.appdev.2016.02.008

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *10*(3), 333-351. http://dx.doi.org/10.1207/s15328007sem1003_1

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Lauritzen, S. L. (1996). *Graphical models* (Vol. 17). Clarendon Press.

Malecki, C. K., & Elliot, S. N. (2002). Children's social behaviors as predictors of academic achievement: A longitudinal analysis. *School Psychology Quarterly*, *17*(1), 1-23. http://dx.doi.org/10.1521/scpq.17.1.1.19902

McNeish, D., An, J., & Hancock, G. R. (2018). The thorny relation between measurement quality and fit index cutoffs in latent variable models. *Journal of Personal Assessment*, *100*(1), 43-52. http://dx.doi.org/10.1080/00223891.2017.1281286

McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*. Advance online publication. http://dx.doi.org/10.3758/s13428-020-01398-0

Montroy, J. J., Bowles, R. P., Skibbe, L. E., & Foster, T. D. (2014). Social skills and problem behaviors as mediators of the relationship between behavioral self-regulation and academic achievement. *Early Childhood Research Quarterly*, *29*(3), 298-309. http://dx.doi.org/10.1016/j.ecresq.2014.03.002

Muthén, B., Muthén, L., & Asparouhov, T. (2015, March 24). *Estimator choices with categorical outcomes* [Web log post]. http://www.statmodel.com/download/EstimatorChoices.pdf

Panayiotou, M., Humphrey, N., & Wigelsworth, M. (2019). An empirical basis for linking social and emotional learning to academic performance. *Contemporary Educational Psychology*, *56*, 193-204. https://doi.org/10.1016/j.cedpsych.2019.01.009

Pons, P., & Latapy, M. (2006). Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications*, *10*(2), 191-218. http://dx.doi.org/10.7155/jgaa.00124

Rabiner, D. L., Godwin, J., & Dodge, K. A. (2015). Predicting academic achievement and attainment: The contribution of early academic skills, attention difficulties, and social competence. *School Psychology Review*, *45*(2), 250-267. http://dx.doi.org/10.17105/SPR45-2.250-267

Racz, S. J., Putnick, D. L., Suwalsky, J. T. D., Hendricks, C., & Bornstein, M. H. (2017). Cognitive abilities, social adaptation, and externalizing behavior problems in childhood and adolescence: Specific cascade effects across development. *Journal of Youth and Adolescence*, *46*(8), 1688-1701. http://dx.doi.org/10.1007/s10964-016-0602-3

Reise, S. P. (2012). Invited paper: The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(6), 667-696. http://dx.doi.org/10.1080/00273171.2012.715555

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, *92*(6), 544-559. http://dx.doi.org/10.1080/00223891.2010.496477

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137-150. http://dx.doi.org/10.1037/met0000045

Sherbow, A., Kettler, R. J., Elliott, S. N., Davies, M., & Dembitzer, L. (2015). Using the SSIS assessments with Australian students: A comparative analysis of test psychometrics to the US normative sample. *School Psychology International*, *36*(3), 313-321. http://dx.doi.org/10.1177/0143034315574767

Spence, S. H. (2003). Social skills training with children and young people: Theory, evidence and practice. *Child and Adolescent Mental Health*, *8*(2), 84-96. https://doi.org/10.1111/1475-3588.00051

Spinard, T. L., & Eisenberg, N. (2009). Empathy, prosocial behavior, and positive development in schools. In R. Gilman, E. S. Huebner, & M. J. Furlong (Eds.), *Handbook of positive psychology in schools* (pp. 119-130). Routledge.

Sturgess, J., Rodger, S., & Ozanne, A. (2002). A review of the use of self-report assessment with young children. *British Journal of Occupational Therapy*, *65*(3), 108-116. http://dx.doi.org/10.1177/030802260206500302

Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development*, *88*(4), 1156-1171. http://dx.doi.org/10.1111/cdev.12864

van Bork, R., Epskamp, S., Rhemtulla, M., Borsboom, D., & van der Maas, H. L. J. (2017). What is the p-factor of psychopathology? Some risks of general factor modeling. *Theory & Psychology*, *27*(6), 759-773. http://dx.doi.org/10.1177/0959354317737185

van der Maas, H. L., Dolan, C. V., Grasman, R. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*(4), 842-861. http://dx.doi.org/10.1037/0033-295X.113.4.842

Verboom, C. E., Sijtsema, J. J., Verhulst, F. C., & Penninx, B. W. J. H. (2014). Longitudinal associations between depressive problems, academic performance, and social functioning in

adolescent boys and girls. *Developmental Psychology*, *50*(1), 247-257. http://dx.doi.org/10.1037/a0032547

Wentzel, K. R. (1991). Social competence at school: Relation between social responsibility and academic achievement. *Review of Educational Research*, *61*(1), 1-24. http://dx.doi.org/10.3102/00346543061001001

Wigelsworth, M., Humphrey, N., Kalambouka, A., & Lendrum, A. (2010). A review of key issues in the measurement of children's social and emotional skills. *Educational Psychology in Practice*, *26*(2), 173-186. http://dx.doi.org/10.1080/02667361003768526

Williams, D. R., & Rast, P. (2020). Back to the basics: Rethinking partial correlation network methodology. *British Journal of Mathematical and Statistical Psychology*, *73*(2), 187-212. http://dx.doi.org/10.1111/bmsp.12173

Wu, Z., Mak, M. C. K., Hu, B. Y., He, J., & Fan, X. (2019). A validation of the social skills domain of the Social Skills Improvement System-Rating Scales with Chinese preschoolers. *Psychology in the Schools*, *56*(1), 126-147. http://dx.doi.org/10.1002/pits.22193

Zachrisson, H. D., Janson, H., & Lamer, K. (2018). The Lamer Social Competence in Preschool (LSCIP) scale: Structural validity in a large Norwegian community sample. *Scandinavian Journal of Educational Research*, *63*(4), 551-565. http://dx.doi.org/10.1080/00313831.2017.1415963