



Measuring Social and Emotional Learning Skills of Preschool Children in Croatia: Initial Validation of the SSIS SEL Brief Scales

Sanja Tatalović Vorkapić¹, Christopher J. Anthony^b, Stephen N. Elliott^c, Ilaria Grazzani^d, Valeria Cavioni^e

^a *Faculty of Teacher Education in Rijeka, University of Rijeka, Rijeka, Croatia*

^b *University of Florida, Gainesville, USA*

^c *Arizona State University, Tempe, USA*

^d *University of Milano-Bicocca, Department of Human Sciences for Education 'Riccardo Massa', Italy*

^e *Department of Social Sciences, University of Foggia, Italy*

Although there is increased interest in social and emotional competence and mental health in Croatia, there are currently limited measurement options available for early childhood settings. Thus, the SSIS SEL Brief Scales (SSIS SEL*b*), an efficient measure of social and emotional learning competencies developed in the United States, was translated to Croatian and used by 49 early childhood educators to rate a sample of 685 children (average age 4.3 years) from several counties in Croatia. Regarding measurement invariance, the final model cohered substantially with a CASEL-inspired framework. Overall reliability was also high especially for the SEL Composite ($\alpha = .94$), with notably lower reliability for subscales. Regarding cross-group concurrent validity, concurrent coefficients were largely similar across age and gender, but there were regional differences in validity. Likewise, validity correlations were in line with expectations, with moderate relationships observed between the SSIS SEL Composite and Child and Youth Resilience Measure scores. In sum, the high level of reliability provided a foundation for applied and research usage of the Croatian SSIS SEL*b*, although further validation research will continue to be necessary before widescale implementation. Limitations to the study are discussed and also point to needed additional research before utilizing the Croatian translated SSIS SEL*b* for applied decisions with young children.

Keywords: Preschool Education; Social and Emotional Learning; Croatia; SSIS SEL Brief Scales; Validation

First submission 7th November 2023; Accepted for publication 8th March 2024.

1

Corresponding author. Email address: sanjatv@uniri.hr

<https://doi.org/10.56300/RRMN9191>

Introduction

Cefai and colleagues (2018) emphasize the importance of integrating social and emotional competencies as a key component in the education of children and youth. Although most educational systems focus on children's academic development, it has been shown that for children to grow into independent, autonomous, and active individuals, it is necessary to develop social and emotional competencies and skills. Numerous studies indicate the importance of promoting the development of social-emotional competencies from an early age (Grazzani et al., 2016), and various programs are being introduced in educational and pedagogical work with children (Boncu et al., 2017; Durlak et al., 2011). Much of these developments have coalesced around the term *social and emotional learning* (SEL).

SEL programs exist as interventions that demonstrate the importance of recognizing and managing emotions and develop social competence (CASEL, 2013). Poorer outcomes on measures of social and emotional competence are achieved by children who live in an environment that highlights their shortcomings, limits, and emphasizes limitations. Although there are various conceptualizations of individual domains of SEL competencies, perhaps the most prominent framework was developed by the Collaborative for Academic, Social, and Emotional Learning (CASEL; e.g., 2013), which includes five basic dimensions: self-awareness and self-management as intrapersonal skill domains, social awareness and relationship skills as interpersonal skill domains, and responsible decision making which includes both intrapersonal and interpersonal skills.

Early Childhood Education and SEL Assessment in Croatia

In the Republic of Croatia, educational, health care, and nourishment programs are available, but not compulsory from ages six months until school age (six or seven). Preschool education, the start of compulsory and free education in Croatia, occurs one year prior to enrollment in eight-year elementary education (typically at age of six). These early school years, spread in Croatia across different levels of education, are critical for positive SEL development. Through primary education, which is compulsory and free for children between the ages of six/seven and fifteen, the student acquires knowledge and skills for continuing education at the secondary school level. Despite the acknowledged importance of early childhood and school settings for promotion of positive SEL competencies and prevention of mental health difficulties, until recently only a small number of SEL-programs have been developed and evaluated in Croatia (e.g., Tatalović Vorkapić, 2019; Tatalović Vorkapić et al., 2023). However, SEL at the complete level is not fully integrated into the early and preschool curriculum in Croatia.

All quality educational work with children of preschool age, entails the application of valid and objective measures for their social and emotional competencies. There are several ways to measure children's SEL competencies, such as self-assessments, peer reports, direct behavior observation, and third-party behavior rating scales (Monnier, 2015). Rating scale for teachers can have a twofold purpose: obtaining objective, valid and reliable ratings for children's SEL and using those ratings as a guideline for creating activities, workshops, and programs to enhance children's SEL skills. Perhaps because of their utility and

flexibility, teacher rating scales are widely proliferating both in the U.S. and throughout Europe (e.g., Elliott et al., 2020; Anthony et al., 2022a). Yet, there remain very limited options for SEL assessment via teacher rating scales adapted for use in Croatian early childhood settings.

The SSIS-SEL Brief Scales

One brief measure, recently translated into Croatian, holds promise to promote strong assessment of SEL in Croatian preschools – the SSIS SEL Brief Scales (SSIS SEL*b*; Anthony et al., 2020; 2021; Elliott et al., 2020). The SSIS SEL*b* were developed in the United States via the utilization of Item Response Theory (IRT) to select maximally efficient sets of items from the SSIS SEL (Gresham & Elliott, 2017), a prominent, comprehensive, multi-rater assessment of children’s SEL utilizing the CASEL five-factor framework. The major goal of development was to develop an efficient measure well-adapted for use in schools’ multi-tiered support systems targeting students’ SEL development. Initial evidence has supported the psychometric quality of scores from the SSIS SEL*b* (Anthony et al., 2020a; 2020b; 2021; Elliott et al., 2021) and subsequent evidence has corroborated this conclusion (e.g., Anthony et al., 2021; Cavioni et al., 2023a).

Notably, the SSIS SEL*b* was recently translated into Croatian as part of a large-scale six-country project called Promoting Mental Health at Schools (Anthony et al., 2023; Cavioni et al., 2023b; Cefai et al., 2022a, 2022b). The thorough translation process was followed up by initial validation evidence exploring measurement invariance (Anthony et al., 2023) and cross-country validity (Anthony et al., 2022b) evidence for scores from the Croatian version of the measure. Evidence was positive, providing a strong foundation for further validation and expansion in Croatia and other European countries, such as Romania, Italy, Portugal, Latvia and Greece (Colomeischi et al., 2022; Martinsone et al., 2022). However, initial evidence did not evaluate score functioning *within* rather than between countries, including evaluation of the SSIS SEL*b* across relevant demographic groups within Croatia (e.g., gender, age, region). Furthermore, the initial sample mostly consisted of Primary Elementary and Secondary Elementary aged students and did not include students in the Early Elementary grades. Thus, there is need to build on the strong foundation of the Croatian translation of the SSIS SEL*b* to promote rigorous SEL assessment of young Croatia children as observed by early childhood educators.

Study Aim and Hypotheses

This study seeks to address the need for well-validated measures of young children’s SEL, particularly in Croatia. The main aim of this study was to validate and test the basic psychometric properties of the SSIS SEL*b* with a Croatian sample of early childhood educators’ ratings. The study sought to address several research questions with the following expectations.

1. What is the overall evidence for the reliability and validity of scores from the SSIS SEL*b* – Teacher version in Croatia?

2. Based on prior evidence from U.S. (e.g., Anthony et al., 2022a) and European (e.g., Anthony 2022b; Anthony et al., 2023) samples, we anticipated that reliability would be high for the overall composite from the SSIS SEL*b* and lower for SSIS SEL*b* subscales. We expected the same pattern for concurrent validity correlations with a moderate positive relationship between SSIS SEL*b* scores and concurrent validity measures of children's resilience.
3. What is the evidence for the measurement invariance of the SSIS SEL*b* and for concurrent validity invariance of scores from the SSIS SEL*b* – Teacher version for relevant subgroups in Croatia (age, gender, region)?

Based on prior European evidence (Anthony et al., 2022b; Anthony et al., 2023), we anticipated that the SSIS SEL*b* would show evidence of measurement and concurrent validity invariance, although these analyses were somewhat exploratory as within country analyses have never been completed with the SSIS SEL*b* outside of the U.S.

Methodology

Participants

In this study, the SSIS SEL*b* (Elliott et al., 2020) was administered to a multi-stage sample of 685 preschool children in Croatia. Their behavior was rated by 49 early childhood educators from educational groups they work. The data collection was conducted in six counties in Croatia (Istria County, Split-Dalmatia County, Karlovac County, Koprivnica-Križevci County, Zagreb County, and the City of Zagreb). At the initial stage of the project, an equally distributed geographic representation of Croatian counties was planned. Therefore, of the total of 21 counties in Croatia, the most populated counties from the west, south and north of Croatia were selected. In the second phase of the study, the call for collaboration in research was sent to each kindergarten in the selected counties, and data were collected at those kindergartens which positively replied to the request. Unfortunately, not all eligible kindergarten classrooms had principals who agreed to participate. This led to some differences in participation across region that highlight the need to examine region as a potentially important moderator variable. Therefore, the sample consisted of those kindergartens which were available for participating in this study regarding their consent to be included in the study. In the third phase, all children were rated in those kindergartens whose principals agreed to participate in this study. With the purpose of keeping the identity of early childhood educators and kindergartens anonymous, the information on the names of kindergartens is not presented. The average age of educators who participated in this study was 33.59 (SD = 9.88) and ranges from 19 to 59 years. Of the total number of educators, 98.0% (48) were female, while 2.00% (1) were male. The average work experience of educators was 9.39 years (SD = 9.66), ranging from 1 to 36 years of work experience. Demographic characteristics of students is presented in Table I.

Table I.*Demographic Characteristics (Percentages) of Participants (N = 685)*

Characteristic	%
Gender	
Boy	52.4
Girl	47.6
Location	
Zagreb City	36.8
Zagreb County	28.3
Other	34.9
Age	
3-year-olds	30.1
4-year-olds	26.9
5-year-olds	28.2
6-7-year-olds	14.3
Child has Special Needs	3.2
<i>Note.</i> Percentages do not sum to 100 in all cases due to rounding and/or missing data	

Measures

SSIS SEL Brief Scales. The SSIS SELb Teacher version was administered, which was translated using the back-translation method to the Croatian language within the PROMEHS project (Anthony et al., 2022b; Anthony et al., 2023). The SSIS SELb is a 20-item, multi-person assessment scale that focuses exclusively on the assessment of children’s social-emotional learning skills. Each item is rated on a 4-point Likert scale from 0 (*Never*) to 3 (*Almost Always*). Completion of the scale takes roughly 5 minutes. With regard to evidence of reliability and validity, Elliott et al. (2020) report reliability coefficients of .95, .78, and .65 for the Cronbach’s α , Test-Retest Reliability, and Interrater reliability of the teacher form of the SSIS SELb in the standardization sample. Elliott et al., also reported concurrent validity coefficients in line with expectations for a variety of constructs. Such initial validity evidence has also been corroborated by subsequent research in samples both in the United States (Anthony et al., 2022a) and European Countries (e.g., Anthony et al., 2022b; Anthony et al., 2023). The test authors recommend that the SEL Composite is used for decision making purposes, and thus this score (which is the sum of all 20 items) was utilized as the focal score of this analysis.

Child and Youth Resilience Measure. To gather evidence of concurrent validity and compare that evidence across demographic groups, we utilized the Child and Youth Resilience Measure (CYRM; Ungar & Liebenberg, 2011). Resilience refers to a dynamic process of encompassing positive adaptation within the context of significant adversity (Luthar et al., 2000). Starting from this resilience definition, the CYRM was created as a measure of the resources (individual, relational, communal and cultural) that are available to individuals that may bolster their resilience (Ungar & Liebenberg, 2011). For the current study, we used a short form (12 items; Ungar, 2016) of the original 28-item measure. For this study, the scale went through a

back-translation procedure, and applied for validation study purposes, on the sample of preschool aged children. The assessment of the resilience of each child in the educational group is rated on a scale of 1 (*No*) to 3 (*Yes*). The sum of all evaluations on this scale, same as the mean value present the total resilience of each child. The reliability coefficient (Cronbach α) obtained in Croatian validation study for CYRM-12 was .80 (Tatalović Vorkapić, 2021), which is consistent with the results of the reliability analyzes from the original study ($\alpha = .84$; Liebenberg et al., 2013). In addition to reliability, there is extensive published evidence on the validity of scores from this measure in samples across the world (e.g., Russell et al., 2021).

Procedure

The research was conducted in Croatian kindergartens, on a multi-phase sample of young children and early childhood educators. In the first phase, the largest counties from three geographical parts of Croatia (South, West and North) were chosen to participate in the study. The principals of all kindergartens from the different counties received a formal letter from the Faculty of Teacher Education in Rijeka explaining the goal of the study, and method of the research, as a request for participating in this study. Ethical approval was provided by the Ministry of Education in Croatia. After their formal permission and consent to participate, all early childhood educators from kindergartens who agreed to participate evaluated their children's social and emotional competencies using the Croatian SSIS SEL*b* version. Their participation was voluntary and anonymous. The ratings are completely anonymous, since only the educators who assessed the children from the educational groups in which they worked knew the identity of the children. After kindergarten's principal permission, researchers sent a link to online questionnaire to be distributed to early childhood educators in kindergarten by their principals. Early childhood educators rated each child in their educational group, giving each child a code (only educators knew the codes for each child in a group).

Data Analysis

Data analysis proceeded in several stages with two primary analyses to evaluate the measurement invariance and invariance of validity coefficients of the SSIS SEL*b* in Croatia.

Overall Reliability and Validity Analysis

First, we completed a set of basic reliability analyses on the SSIS SEL*b* SEL Composite and subscale scores. Specifically, we computed Cronbach's α and concurrent validity correlations with the CYRM.

Measurement Invariance Testing

Next, a set of analyses were undertaken to evaluate the measurement invariance of the SSIS SEL*b* across age (3–4-year-olds vs. 5–7-year-olds), gender (boy vs. girl), and region (Zagreb vs. other regions in Croatia). We conducted preliminary analyses to evaluate data quality and consider missingness. Because the SSIS SEL*b* only includes 4 response options (*Never, Sometimes, Often, Almost Always*) we treated data as ordinal for all measurement invariance testing and utilized a robust weighted least squares estimator for all analyses. With

regard to missing data, no data was missing for 19 of the 20 items and missingness was only 0.6% for the remaining item. Thus, we utilized standard case wise deletion procedures to address missing data.

In addition to considering missing data, measurement invariance testing with ordinal indicators requires that items have the same number of response options across groups. When data are sparse (e.g., if *Never* is not endorsed for any individual in one group of a multi-group analyses) problems can arise. To address this issue, we evaluated our data and collapsed categories (as suggested by Liu et al., 2017) where necessary. Fortunately, collapsing data was only necessary in one instance as the item *Acts responsibly when with others* did not have any *Never* responses indicated for children in regions of Croatia outside Zagreb.

After preliminary data evaluation, we proceeded with our primary measurement invariance analyses. We followed procedures for measurement invariance testing as guided by Wu and Estabrook (2016) who specified an approach and process of model comparison that did not require assuming invariance of some items to identify multi-group models (a distinct advantage over other approaches which require an analyst to assume measurement invariance of at least one item to identify models and proceed with analyses). Throughout testing, we followed guidance of Svetina et al. (2020). Model syntax was generated via implementation of the `meas.Eq.syntax` function of the `semTools` package (Jorgensen et al., 2022) and models were fit in the `lavaan` package (Rosseel, 2012) of the R statistical program (R Core Team, 2013). As noted above, we utilized a robust least squares estimator, the WLSMV implemented by `lavaan`, as recommended for ordinal data with fewer than 5 categories (Rhemtulla et al., 2012).

We first tested a configural invariance model, which imposes the same basic pattern of latent variables and loadings across groups but makes no constraints on the values of any estimates. This model was evaluated against standard fit indices suggested by Hu and Bentler (1999). Specifically, we considered CFI/TLI values greater than or equal to .95, RMSEA values less than or equal to around .06, and SRMR values less than or equal to .08 as indicative of adequate fit. In instances of poor fit of the configural model, we evaluated resulting model estimates to evaluate potential reasons for poor fit.

As will be fully detailed, standard confirmatory configural models did not produce adequate fit to the data and reasonable minor model modifications were not able to be found. As such, progressing with standard confirmatory factor analytic approaches to measurement invariance testing were not possible. Thus, we utilized Exploratory Structural Equation Modeling (ESEM; van Zyl & Klooster, 2022) as an alternate methodology for considering the measurement properties of the SSIS SELb in the sample. In contrast to standard CFA approaches, which constrain all cross-loadings for items not indicating a particular factor to 0, ESEM allows small cross-loadings to be incorporated in otherwise confirmatory models. These models allow some of the flexibility of exploratory factor analysis to be combined with the analytic advantages and procedures possible in confirmatory factor analysis (such as measurement invariance). Procedures based in ESEM have recently grown in prominence due to analytic advantages for achieving superior fit, obtaining less biased estimates of interfactor correlations, and more realistic measurement assumptions (Marsh et al., 2014).

Based on recommendations to consider theoretical and empirical models (van Zyl & Klooster, 2022), we considered three models when employing ESEM. First, we evaluated the standard five-factor CASEL

model – the model that guided the development of the SSIS SELb (Anthony et al., 2020a; 2021). Next, we considered the empirical model suggested by recent large-scale analyses of the SSIS SELb in 6 European countries conducted by Anthony et al., (2023) in which the Self-Awareness, Self-Management, Responsible Decision-Making, and Relationship Skills merged to form a single factor and the Social Awareness factor remained distinct. Finally, we considered a three-factor model suggested by CASEL (2013) in which Self-Awareness and Self-Management comprise an Intrapersonal Skills domain; Social Awareness and Relationship Skills comprise an Interpersonal Skills domain; and Responsible Decision-Making remains its own domain comprised of both Intra- and Interpersonal Skills. We based retention of these models based on overall fit with the same criteria as listed above as well as theoretical meaningfulness.

Once an appropriate model was identified, we utilized ESEM specific syntax generated via an online syntax generator (De Beer & Morin, 2022) and employed in MPlus (Muthén & Muthén, 2017). We fit and compared configural, weak invariance (loadings constrained to equality across groups), and strong invariance (loadings and thresholds constrained to equality across groups). Because strict invariance models are not required for cross-group comparisons (Svetina et al., 2020; Thompson & Green, 2006; Little, 2013), we did not proceed with testing strict invariance models after scalar or partial scalar invariance was established. The comparative fit of these models was evaluated based on established criteria for measurement invariance testing specified by Svetina et al. (2020). Specifically, we deemphasized $\Delta\chi^2$ values due to their sensitivity to sample size (Little et al., 2013) in favor of considering Δ RMSEA (desired decrement in model fit of less than .015; Chen, 2007) and Δ CFI (desired decrement in model fit of .01 or less; Chen, 2007; Cheung & Rensvold, 2002).

Validity Invariance Testing

Next, we undertook a series of analyses to determine if validity coefficients were equivalent across age, gender, and region. Our analyses were guided by Lautenschlager and Mendoza (1986) who specified particular sets of regression models to test and compare to determine test bias. We determined that consideration of the slope term for a regression model in which a focal predictor (i.e., CYRM scores) predicted the SEL Composite indicated concurrent validity and comparing a model with only the group variable and focal predictor with a model with a group variable, a focal predictor, and their interaction would indicate whether differential relationships between predictor scores and the SEL Composite existed. Moderation was evaluated via consideration of ΔR^2 values their statistical significance. There were 13 cases (1.9% of the sample) with out of bounds CYRM data (scores greater than 3) which were deleted for all validity correlation and invariance testing.

Results

Descriptive statistics for the sample are presented in Table II. With regard to overall reliability and validity evidence, Cronbach's α was high for the SEL Composite ($\alpha = .94$) and notably lower for subscales, ranging from .61 to .83 (Median = .76). Concurrent validity correlations were generally high and statistically significant with the CYRM. Full results are presented in Table III.

Table II*Descriptive Statistics for Analytic Variables Across Age, Gender, and Region*

	SEL Composite		CYRM	
	M	SD	M	SD
Age				
3-4	3.01	0.55	2.68	0.31
5-7	3.08	0.60	2.69	0.32
Gender				
Boy	2.99	0.57	2.66	0.30
Girl	3.09	0.56	2.72	0.32
Region				
Zagreb	2.93	0.57	2.67	0.33
Other Regions in Croatia	3.23	0.51	2.72	0.28

Note. Mean scores for all analytic variables presented. SSIS SEL Composite is rated on a scale of 0 (*Never*) to 3 (*Almost Always*). CYRM = Child and Youth Resilience Measure. CYRM is rated on a scale of 1 (*No*) to 3 (*Yes*).

Table III*Overall Reliability and Validity Evidence for Scores from SSIS SELb*

	Cronbach's α	CYRM Correlations
Self-Awareness	.76	.67
Self-Management	.61	.45
Social Awareness	.73	.72
Relationship Skills	.82	.73
Responsible Decision-Making	.83	.67
SEL Composite	.94	.73

Note. CYRM = Child and Youth Resilience Measure; all CYRM correlations were statistically significant ($p < .001$).

Measurement Invariance Testing

First, we conducted measurement invariance testing across age groups (3–4-year-olds vs. 5–7-year-olds). Fitting the configural model resulted in several estimation problems. Examination of latent variable correlation matrices indicated several outer bound correlations indicating very high relationships between latent constructs. Latent factors were collapsed, and models tested iteratively until model convergence was reached, which ultimately required collapsing all factors other than Self-Management to generate a two-factor model. This model evidenced poor fit to the data ($\chi^2 = 1,963.80$; $p < .001$; RMSEA = .127; CFI = .916; TLI = .905; SRMR = .093). The same pattern of collapsing was necessary for the gender model and resulted in a configural model with similarly poor fit ($\chi^2 = 1,935.27$; $p < .001$; RMSEA = .126; CFI = .901; TLI = .888; SRMR = .092).

Finally, for the country model, factor collapsing was also necessary, but resulted in a two-factor model in which Self-Awareness, Self-Management, and Responsible Decision-Making were collapsed and Social Awareness and Relationship Skills were collapsed. This model also showed poor fit to the data ($\chi^2 = 2,215.39$; $p < .001$; RMSEA = .131; CFI = .919; TLI = .909; SRMR = .107).

Based on these indications of poor fit, we proceeded to utilize ESEM to test alternate potential configural models. We started by evaluating competing ESEM models for the overall sample to determine the best model to use as a baseline for measurement invariance testing. The CASEL-based 5-factor model evidenced strong fit to the data ($\chi^2 = 320.39$; $p < .001$; RMSEA = .057; CFI = .989; TLI = .980; SRMR = .021). This model also had some theoretical shortcomings, however, as 3 of the 5 factors had 50% or less of their supposed indicators load more highly on their intended factor and 40% of all indicators loaded more highly on non-intended factor than an intended factor. Next, the 2-factor model based on Anthony et al., 2023 evidenced marginal fit to the data ($\chi^2 = 828.68$; $p < .001$; RMSEA = .081; CFI = .896; TLI = .870; SRMR = .040). This model also evidenced some poor coherence with a priori expectations as only one of the Social Awareness items loaded more highly on its factor than the other factor comprising remaining scales. Finally, the CASEL-based 3 factor model evidenced good fit to the data ($\chi^2 = 978.74$; $p < .001$; RMSEA = .089; CFI = .960; TLI = .949; SRMR = .042). Loadings for this model also largely aligned with theory as only 3 items on the Intrapersonal factor and 2 items on the Interpersonal factor had higher loadings on their non-specified factors. Furthermore, of the 4 items allowed to load equally on both Intra- and Inter-personal factors, 2 loaded higher on the Intrapersonal factor and 2 loaded higher on the Interpersonal factor. Based on this fit and coherence with theory, this model was retained for measurement invariance testing. Fit statistics for all measurement invariance models can be found in Table IV. Strong measurement invariance was reached across all analytic variables as all successively constrained models did not lead to unacceptable decrements in model fit. Unstandardized loadings of final strong invariance models can be found in Table V. Table V also includes interfactor correlations from standardized solutions of final strong invariance models.

Validity Invariance Testing

Next, we undertook our concurrent validity invariance testing via series of hierarchical regression. All models, interaction terms, and p -values are presented in Table VI. For Age and Gender based invariance, hierarchical regression testing indicated that SSIS SELb SEL Composite scores were roughly equally related to CYRM scores. For Regional Invariance, however, differential relationships were found ($\Delta R^2 = 0.02$, $p < .001$).

Table IV*ESEM Measurement Invariance Fit Statistics Tested Models*

	χ^2	df	<i>p</i>	RMSEA	SRMR	CFI	TLI	$\Delta \chi^2$	df	<i>p</i>	Δ RMSEA	Δ CFI
Age												
Configural	1,174.59	302	<.001	.092	.049	.965	.956	-	-	-	-	-
Weak	1,387.70	338	<.001	.095	.060	.957	.952	271.42	36	<.001	.003	.008
Strong	1,464.55	376	<.001	.092	.062	.956	.955	157.89	38	<.001	-.003	.001
Gender												
Configural	1,158.93	302	<.001	.091	.048	.960	.950	-	-	-	-	-
Weak	1,196.53	338	<.001	.086	.055	.960	.955	158.84	36	<.001	-.005	<.001
Strong	1,205.17	376	<.001	.080	.056	.962	.961	80.94	38	<.001	-.006	-.002
Region												
Configural	1,266.80	302	<.001	.097	.055	.962	.953	-	-	-	-	-
Weak	1,279.43	338	<.001	.090	.062	.963	.959	173.25	36	<.001	-.007	-.001
Strong	1,411.78	375	<.001	.090	.066	.959	.959	218.22	37	<.001	<.001	.004
<i>Note.</i> Negative Δ CFI and Δ RMSEA values indicate increases in these indices relative to previous model.												

Table V

Unstandardized Item Loadings from Final Strong Invariance Models

Intended Domain	Item	Age		Gender		Region	
		Intrapersonal	Interpersonal	Intrapersonal	Interpersonal	Intrapersonal	Interpersonal
Intrapersonal	1	-0.27	1.19	-0.44	1.34	-0.39	1.57
	6	0.38	0.77	0.42	0.77	0.51	0.96
	11	0.62	0.48	1.32	0.45	1.26	0.45
	16	0.78	0.59	1.09	0.65	1.09	0.66
	2	0.37	0.59	0.31	0.72	0.29	0.78
	7	0.50	0.38	0.78	0.27	0.44	0.29
	12	0.72	-0.51	1.08	-0.76	0.87	-0.74
	17	0.74	0.56	0.85	0.57	0.70	0.50
Both	5	0.62	0.91	0.63	1.09	0.84	1.34
	10	0.81	0.59	1.17	0.50	0.73	0.50
	15	0.95	0.40	1.26	0.39	1.24	0.37
	20	0.74	1.08	0.72	1.29	0.74	1.32
Interpersonal	3	0.35	1.11	0.20	1.43	0.23	1.47
	8	1.06	0.27	1.05	0.22	0.83	0.17
	13	0.35	0.59	0.37	0.79	0.35	0.74
	18	-0.13	1.38	-0.28	1.66	-0.15	1.99
	4	0.39	1.33	0.16	2.06	0.30	2.37
	9	0.21	1.18	0.12	1.28	0.19	1.56
	14	-0.13	1.49	-0.23	1.27	-0.21	1.49
	19	0.97	0.68	1.37	0.64	1.71	0.78
Interfactor Correlations		3-4 = .68	5-7 = .14	Boys = .53	Girls = .55	Zagreb = .57	Other = .63
<i>Note.</i> Unstandardized loadings presented because these are constrained to equality across groups. Correlations reported from standardized solution and reported for each group because strong invariance models do not constrain them to equality.							

Table VI

Hierarchical Regression Models for CYRM Concurrent Validity Comparisons across Age, Gender, and Region

	Model 1		Model 2	
	B	p	B	p
Age Group (3-4 year olds vs. 5-7 year olds)				
Intercept	-0.68	<.001	-0.90	.166
CYRM	1.32	<.001	1.41	<.001
Age Group	0.06	.042	0.15	.556
Interaction	-	-	-0.03	.723
R ²	0.54	<.001	0.54	<.001
ΔR ²	-	-	<0.001	.723
Gender (Boy vs. Girl)				
Intercept	-0.54	<.001	-0.90	.029
CYRM	1.33	<.001	1.46	<.001
Gender	0.01	.729	0.25	.341
Interaction	-	-	-0.09	.359
R ²	0.53	<.001	0.53	<.001
ΔR ²	-	-	<0.001	.359
Region (Zagreb vs. Other Regions in Croatia)				
Intercept	-0.76	<.001	-2.56	<.001
CYRM	1.30	<.001	1.97	<.001
Region	0.22	<.001	1.63	<.001
Interaction	-	-	-0.52	<.001
R ²	0.57	<.001	0.58	<.001
ΔR ²	-	-	0.02	<.001
<i>Note. 5-7 Year Olds, Boys, and Other Regions in Croatia coded 1; Model 1 = Model without Interaction Term; Model 2 = Model with Interaction Term</i>				

Discussion

SEL competencies continually develop throughout the lives of children and youth. To advance understanding of this early developmental period and the emergence of key SEL competencies of boys and girls, sound assessments that can efficiently capture teachers' and parents' observations of young children are needed. Thus, the aim of this study was to contribute to the growing knowledge base about the SSIS SEL*b*, an assessment originally developed in the United States as a universal screening tool. This brief SEL assessment has recently been translated into eight common languages throughout Europe and used as a key descriptive and outcome variable for several social-emotional intervention projects. As such, this study's main contribution was validity evidence about SSIS SEL*b* scores of early childhood educators to document young children's SEL competencies.

Our findings addressed research questions about the reliability and validity of SEL scores for the Croatian translation of the SSIS SEL*b* for a robust sample of young children. Specifically, we investigated (a) the measurement invariance of the assessment for subgroups of students who differed by age, gender, and

region and (b) the concurrent validity of the SEL scores with a known measure, the CYRM. Overall, evidence was generally positive, with some mixed findings for cross-group comparisons. Overall reliability was high especially for the SEL Composite ($\alpha = .94$). This high level of reliability provides a critical foundation for the potential for applied and research usage of the Croatian SSIS SELb. Yet, individual subscales had lower levels of reliability, especially the self-management subscale. This could have arisen for several reasons including particular cultural and/or linguistic differences between the U.S./American English and Croatia/Croatian for these items. Additionally, it could have arisen because the underlying behaviors implicated in this subscale are themselves more variable in early childhood settings in Croatia (or in general). Further research is needed to evaluate these possibilities, however, any future applied work with the SSIS SELb should focus on the SEL composite as recommended by the authors in other countries (Elliott et al., 2020). Similar to overall reliability findings, validity correlations were in line with expectations, with moderate relationships observed between the SEL Composite and CYRM scores ($r = .73$), which also provides positive evidence. More evidence, however, is certainly needed before utilizing the Croatian SSIS SELb for individual applied decisions.

Regarding measurement invariance testing, the ultimate supported structure cohered somewhat with the CASEL framework of prior investigations but differed from prior investigations (e.g., Anthony et al., 2023) in the overall lack of fit for the standard configural models. Once an adequate configural invariance model was identified, a high degree of measurement invariance was supported. The final model actually cohered substantially with a CASEL-inspired framework with the five factors distributed amongst Intrapersonal competencies (self-awareness and self-management); Interpersonal competencies (relationship skills and social awareness) and responsible decision making, which incorporates both intra and interpersonal competencies. Both intra and interpersonal domains were indicated by a majority of their intended indicators (63% and 75% respectively) and responsible decision making had 50% of its items load on to both factors, in line with theory.

Although a different analytical approach was used in comparison to prior investigations, in some ways it was both a more accurate reflection of the CASEL model and a more realistic approach to factor models of complex, interrelated social and emotional behaviors. Specifically, the CASEL framework itself was not developed according to rigorous psychometric criteria (Frye et al., in press), but was rather a conceptual synthesis of prior research. Thus, it is perhaps unrealistic to expect each of the five factors to completely separate in empirical analyses as we observed in the current study. Relatedly, it is well known that all human raters bring their own biases, perspectives, and idiosyncrasies to the rating process (Tatalović Vorkapić & Žagar, 2017), which introduces further dependencies and blurry lines between factors. Finally, although SSIS SELb items have a long pedigree, they represent a molar behavior that likely represents multiple competencies especially for children at a more basic level of development. For example, the item *Asks for help when needed* is intended to be an intrapersonal item (on the self-awareness scale) but loaded more highly on the interpersonal skills factor for all models. It is certain that this item recruits some level of self-awareness – one must be self-aware to understand one's needs – yet *expression* of this item also requires adequate interpersonal skills. Molar behaviors requiring a complex set of skills is likely the rule rather than the exception for social behaviors as

would be indicated by the factorial complexity of most items in our analysis. Indeed, the more realistic nature of ESEM approaches is considered one of their major advantages over standard CFA methodology (Marsh et al., 2014). Thus, although the current analysis differs from prior ones in important ways, it perhaps illustrates a better approach to evaluating the factor structure of SEL behavior rating scales that could be emulated in other samples in the future.

Another interesting finding emerging from the measurement invariance models was the interfactor correlations, which were much lower than in prior research. For example, the interfactor correlations found by Anthony et al., (2022a) when utilizing a similar 2 factor model but utilizing standard CFA methods were very high ($r = .89$). In contrast, the observed interfactor correlations of the current study were much lower (mean $r = .52$). Such a finding is plausibly a result of the use of ESEM methodology, which has shown to produce more accurate correlations in factor analytic work relative to the typically inflated correlations from CFA approaches (Shao et al., 2022). It is possible that high interfactor correlations from prior research were influenced by utilized methodology and future research should use ESEM to evaluate this possibility.

Regarding cross-group concurrent validity, concurrent coefficients were largely similar across age and gender, but there were regional differences in validity between Zagreb and Other Regions in Croatia. Validity was similar across age and gender, but a statistically significant difference was observed across region. This may indicate that although when considered at a broad level the SSIS SEL Composite functions the same across groups, there are important regional differences to consider. As noted before, there is a possibility of regional differences due to different type of kindergartens' founders what influenced their working conditions; different initial education of kindergarten teachers what resulted with differences in their educational work; different interpretation of National curriculum for an Early and Preschool Care and Education (2015) what results in different approaches in educational work; and different overall SEL-competencies of kindergarten teachers who participated in the research. It is important to note, however, that this difference may also be due to regional differences in the CYRM, which seems more likely given the positive measurement invariance findings for the SSIS SEL*b* across region in Croatia.

Limitations and Future Research

Although this study broke new ground empirically regarding the assessment of preschool children's social and emotional skills, it has some limitations which concern the sample and the need for additional reliable evidence. First, it is important to note that the current sample was not a random sample of Croatian children. Even though all kindergartens within those counties were contacted, there were numerous of them that did not agree to participate in this study, which may be due to systematic differences between these kindergartens. Additionally, due to limitations in the number of cases from various counties other than Zagreb, we had to collapse other Croatian counties into a single "Other" category for most analyses. There may be important sociocultural differences that were obscured by this conflation and future research with larger samples should explore cross-county differences more closely. Ultimately, the current sample should not be viewed as a true

representation of the Croatian population of 3–7-year-olds. Future research and standardization with larger samples should be undertaken to fully substantiate the SSIS SEL*b* for use in Croatia.

Limitations were also observed regarding the validity measures used and the grouping variables available. Specifically, although the CYRM was an adequate measure to use for initial validation purposes, it is indirect and shares informant variance with the SSIS SEL*b*. Future research should be conducted with measures of a more varied set of constructs (e.g., mental health; problem behaviors; academic achievement) and with a wider variety of measurement approaches (e.g., direct assessment; direct observation; sociometric techniques). Additionally, work should be undertaken to further validate the parent and student forms of the SSIS SEL*b* to enable use of multiple informants in research and practice. Similarly, future studies evaluating the structure and concurrent validity of scores across a more varied set of demographic variables (e.g., special education status; socioeconomic status) would be beneficial for supporting the use of the scale in Croatia.

Future research is also needed to further establish the reliability, validity, and utility of scores of the SSIS SEL*b* in Croatia. First, it would be useful to run test-retest and inter-rater reliability analyses in future research studies to augment the internal consistency evidence reported in this study. Moreover, another particularly profitable source of evidence might be conditional probability statistics and predictive validity studies evaluating how well the SSIS SEL*b* predicts valued educational and social outcomes in Croatia. Such analyses is needed to support use of this measure in kindergartens.

Conclusion

The findings from this study contribute to the growing knowledge base about the psychometric qualities and utility of the SSIS SEL*b* for assessing preschool Croatian children's social-emotional competences. Such a study provides initial but solid evidence about the reliability and validity of this assessment's scores. More research is certainly needed prior to applied use, but the findings from this study help guide it and help Croatian early childhood educators measure their efforts to affect preschool children's social-emotional health.

Declaration of Conflicting Interests

Christopher J. Anthony and Stephen N. Elliott are authors of the SSIS SEL Brief Scales and receive royalties for its sale in the United States. Other authors declared no potential conflicts of interest.

Acknowledgement

This study was co-funded by the European Commission within the Erasmus+ KA3 research project *PROMEHS - Promoting mental health at schools* (606689-EPP_2018-2-IT-PI-POLICY) and by University of Rijeka, Croatia within the national scientific university project *Children's well-being in transition periods: The empirical validation of Ecological-Dynamic model* (uniri-drustv-18-11).

References

- Anthony, C. J., Elliott, S. N., DiPerna, J. C., & Lei, P. W. (2020). The SSIS SEL Brief Scales—Student Form: Initial development and validation. *School Psychology, 35*(4), 277
- Anthony, C. J., Elliott, S. N., DiPerna, J. C., & Lei, P. W. (2021). Initial Development and Validation of the Social Skills Improvement System –Social and Emotional Learning Brief Scales-Teacher Form. *Journal of Psychoeducational Assessment 39*(2), 166-181
- Anthony, C. J., Brann, K., Elliott, S. N., & Garis, E. J. (2022a). Examining the structural validity of the SSIS SEL Brief Scales – Teacher and Student Forms. *Psychology in the Schools, 59*, 260-280.
<https://doi.org/10.1002/pits.22607>
- Anthony, C. J., Elliott, S. N., Yost, M., Lei, P. W., DiPerna, J. C., Cefai, C... Colomeischi, A. A. (2022b). Multi-informant validity evidence for the SSIS SEL Brief Scales across six European countries. *Frontiers in Psychology, 13*, 928189. <https://10.3389/fpsyg.2022.928189>
- Anthony, C. J., Lei, P. W., Elliott, S. N, DiPerna, J. C., Cefai, C., Bartolo, P. A.... Colomeischi, A. A. (2023). Measurement Invariance of Children's SEL Competencies. *European Journal of Psychological Assessment*. <https://doi.org/10.1027/1015-5759/a000753>
- Boncu, A., Costea, I. & Minulescu, M. (2017). A meta-analytic study investigating the efficiency of socio-emotional learning programs on the development of children and adolescents. *Romanian Journal of Applied Psychology, 19*(2), 35-41. <https://psycnet.apa.org/doi/10.24913/rjap.19.2.02>
- CASEL: The Collaborative for Academic, Social, and Emotional Learning Guide (2013). *Effective Social and Emotional Learning Programs: Preschool and Elementary School Edition*.
<https://files.eric.ed.gov/fulltext/ED581699.pdf>
- Cavioni, V., Conte, E., Grazzani, I., Ornaghi, V., Cefai C., Anthony C., ... & Pepe, A. (2023a). Validation of Italian students' self-ratings on the SSIS SEL brief scales. *Frontiers in Psychology, 14*, 1229653.
<https://doi.org/10.3389/fpsyg.2023.1229653>
- Cavioni, V., Grazzani, I., Orhaghi, V., Agliati, A., Gandellini, S., Cefai, C.... Conte, E. (2023b). A multi-component curriculum to promote teachers' mental health: Findings from the PROMEHS program. *International Journal of Emotional Education, 15*(1), 34-52. <https://doi.org/10.56300/KFNZ2526>
- Cefai, C., Bartolo P. A., Cavioni V., & Downes, P. (2018). *Strengthening Social and Emotional Education as a core curricular area across the EU. A review of the international evidence*, NESET II report. Publications Office of the European Union
- Cefai, C., Camilleri, L., Bartolo, P. A., Grazzani, I., Cavioni, V., Conte, E.... Colomeischi, A. A., (2022a). *Evaluating the effectiveness of the PROMEHS Programme in improving students' and teachers' social and emotional competence, resilience and mental health*. Centre for Resilience and Socio-Emotional Health, University of Malta

- Cefai, C., Camilleri, L., Bartolo, P. A., Grazzani, I., Cavioni, V., Conte, E., ... Colomeischi, A. A. (2022b). The effectiveness of a school-based, universal mental health programme in six European countries. *Frontiers in Psychology, 13*, 925614. <https://doi.org/10.3389/fpsyg.2022.925614>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464-504
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255
- Colomeischi, A. A., Duca, D. S., Bujor, L., Rusu, P. P., Grazzani, I., & Cavioni, V. (2022). Impact of a School Mental Health Program on Children's and Adolescents' Socio-Emotional Skills and Psychosocial Difficulties. *Children, 9*, 1661. <https://doi.org/10.3390/children9111661>
- De Beer, L.T., & Morin, A.J.S (2022). (B)ESEM invariance syntax generator for Mplus. https://stattools.app/b_eseem/
- 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
- Elliott, S. N., Anthony, C. J., DiPerna, J. C., & Lei, P. (2020). *SSIS SEL Brief + Mental Health Scales User Guide and Technical Manual*. SAIL Collaborative.
- Grazzani, I., Ornaghi, V., Agliati, A., & Brazzelli, E. (2016). How to foster toddlers' mental state talk, emotion understanding and prosocial behavior: A conversation-based intervention at nursery school. *Infancy, 21*(2), 199-227
- Gresham, F. M., & Elliott, S. N. (2017). *Social skills improvement system social emotional learning edition rating forms*. Pearson Assessments
- 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
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2022). *semTools: Useful tools for structural equation modeling. R package version 0.5-6*. <https://CRAN.R-project.org/package=semTools>
- Lautenschlager, G. J., & Mendoza, J. L. (1986). A step-down hierarchical multiple regression analysis for examining hypotheses about test bias in prediction. *Applied Psychological Measurement, 10*(2), 133-139
- Liebenberg, L., Ungar, M. & LeBlanc, J. C. (2013). The CYRM-12: A brief measure of resilience. *Canadian Journal of Public Health, 104*(2), 131-135. <https://doi.org/10.1007/bf03405676>
- Little, T. D. (2013). *Longitudinal structural equation modeling*. Guilford Press
- Liu, S., Wang, X., Liu, M., & Zhu, J. (2017). Towards better analysis of machine learning models: A visual analytics perspective. *Visual Informatics, 1*(1), 48-56

- Luthar, S. S., Cicchetti, D., & Becker, B. (2000). The Construct of Resilience: A Critical Evaluation and Guidelines for Future Work. *Child Development*, 71(3), 543-562
- Martinone, B, Supe, I., Stokenberga, I., Damberg, I., Cefai, C., Camilleri, L.... Grazzani, I. (2022). Social Emotional Competence, Learning Outcomes, Emotional and Behavioral Difficulties of Preschool Children: Parent and Teacher Evaluations. *Frontiers in Psychology*, 3, 12, 760782. <https://doi.org/10.3389/fpsyg.2021.760782>
- Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, 10, 85-110. <https://doi.org/10.1146/annurev-clinpsy-032813-153700>
- Monnier, M. (2015). Difficulties in defining social-emotional intelligence, competencies and skills – a theoretical analysis and structural suggestion. *International Journal for Research in Vocational Education and Training*, 2(1), 59-84
- Muthén, B. O., Muthén, L. K., & Asparouhov, T. (2017). *Regression and mediation analysis using Mplus*. Muthén & Muthén
- National Curriculum for an Early and Preschool Care and Education (2015). Ministry of Science, education, and Sport, Zagreb, Croatia. <https://mzo.gov.hr/istaknute-teme/odgoj-i-obrazovanje/nacionalni-kurikulum/nacionalni-kurikulum-za-rani-i-predskolski-odgoj-i-obrazovanje/3478>
- R Core Team, R. (2013). *R: A language and environment for statistical computing*. Vienna <http://www.R-project.org/>
- Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 17(3), 354
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48, 1-36
- Russell, B. S., Collins, C. M., Tomkunas, A. J., & Hutchison, M. (2021). Exploring the factor structure of the child and youth resilience measure (CYRM-12) for young children in a disadvantaged community. *Children and Youth Services Review*, 120, 105746 <https://doi.org/10.1016/j.childyouth.2020.105746>
- Shao, K., Elahi Shirvan, M., & Alamer, A. (2022). How accurate is your correlation? Different methods derive different results and different interpretations. *Frontiers in Psychology*, 13, 901412. <https://doi.org/10.3389/fpsyg.2022.901412>
- Svetina, D., Rutkowski, L., & Rutkowski, D. (2020). Multiple-group invariance with categorical outcomes using updated guidelines: an illustration using M plus and the lavaan/semtools packages. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(1), 111-130
- Tatalović Vorkapić, S. (2021). The CYRM-12 Validation On Preschool Sample In Croatia. In L. Gómez Chova, A. López Martínez, & I. Candel Torres (Eds.), *ICERI2021 Proceedings, 8th-9th November*, Valencia, Spain (pp. 4762-4769)

- Tatalović Vorkapić, S. (2019). Podrška socijalno-emocionalnoj dobrobiti djece u Hrvatskoj: Prikaz tri znanstveno-stručna projekta. In M. Orel (Ed.), *EDUvision 2019 – Sodobni pristopi poučevanja prihajajočih generacij* (pp. 69-83). Ljubljana, Slovenija, EDUvision, Stanislav Jurjevčič s.p.
- Tatalović Vorkapić, S., & Žagar, J. (2017). Is the evaluation of children's temperament independent from pre-school teachers' personality? In Z. Bekirogullari, M. Y. Minas & R. X. Thambusamy (Eds.), *The European Proceedings of Social & Behavioural Sciences (EpSBS), XX*, (pp. 84-96). 5th icCSBs 2017, Future Academy Conferences, Brno, 9-11. January
<http://dx.doi.org/10.15405/epsbs.2017.01.02.10>
- Tatalović Vorkapić, S., Colomeischi, A. A. & Ornaghi, V. (2023). PROMEHS Teachers' Training: The effect of PROMEHS Curriculum on Teachers' Mental Health in Croatia and Romania. *Revista Românească pentru Educație Multidimensională*, 15(3), 400-418
<https://doi.org/10.18662/rrem/15.3/773>
- Thompson, M. S., & Green, S. B. (2006). Evaluating between-group differences in latent variable means. In G. R. Hancock & R. O. Mueller (Eds.), *Structural Equation Modeling: A second course* (1st ed., pp. 119–169). IAP.
- Ungar, M., & Liebenberg, L. (2011). Assessing resilience across cultures using mixed methods: Construction of the Child and Youth Resilience Measure. *Journal of Multiple Methods Research*, 5(2), 126-149.
- Ungar, M. (2016). *The Child and Youth Resilience Measure (CYRM), Youth Version: Usar's Manual: Research, May 2016*. Canada: Resilience Research Centre, 2016.
<https://cyrm.resilienceresearch.org/how-to-use/>
- van Zyl, L. E., & Ten Klooster, P. M. (2022). Exploratory structural equation modeling: Practical guidelines and tutorial with a convenient online tool for Mplus. *Frontiers in Psychiatry*, 12, 795672.
- Wu, H., & Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika*, 81(4), 1014-1045