Examining the Measurement Invariance and Validity of the SSIS SEL Brief + Mental Health Scales – Student Version in Austria and Germany

Christopher J. Anthony[a][1] , Sepideh Hassani[b], Susanne Schwab[bc], Abigail P. Howe[a], Michayla Yost[a], Stephen N. Elliott[d], Marwin Löper[e], Gamze Görel[e] and Frank Hellmich[e]

[a]*School of Special Education, School Psychology, and Early Childhood Studies, University of Florida,*

[b]*Center for Teacher Education, University of Vienna,*

[c]*Research Focus Area Optentia, North-West University, Vanderbijlpark, South Africa*

[d]*Sanford School of Social & Family Dynamics, Arizona State University,*

[e]*Institute of Educational Research, Paderborn University*

The SSIS SEL Brief + Mental Health Scales (SSIS SELb+MHS) are multi-informant assessments developed in the United States to assess the social and emotional learning (SEL) competencies and emotional behavior concerns (EBCs) of school-age youth. Although there are translations of the SEL items of the SSIS SELb+MHS available in other languages, a German translation has never been completed and validated, despite the growing need for SEL and mental health assessment in German-speaking countries. To address this need, this study's primary purpose was the examination of a German translation of the assessment with a specific focus on measurement invariance and concurrent validity invariance testing with 821 3rd through 6th-grade students in Austria and Germany. Results indicated that the SELb+MHS items clustered into 2 SEL factors and 2 EBC factors. With regard to measurement invariance, the SELb+MHS functioned similarly across both Austria and Germany and full scalar invariance was achieved. Additionally, the overall pattern of concurrent

---

[1] Corresponding author. Email address: cjanthony1@gmail.com

validity relationships was as expected and similar across countries. Implications and future directions are discussed.

Keywords: Social and Emotional Learning; Mental Health; SSIS SEL Brief + Mental Health Scales; Measurement Invariance; International Assessment.

Introduction

Children who can make and maintain friendships, manage their emotions and behaviors, and understand themselves and others tend to learn more in school, be more successful in their careers, and have better overall life outcomes (e.g., Durlak et al., 2011). Yet, for many children, developing these skills and competencies is not an easy process. Indeed, with a global prevalence of 10 to 20%, mental health issues in children and adolescents are a growing public health concern that impacts educational attainment and social development in early adulthood (WHO, 2018). Unfortunately, the effects of the COVID-19 pandemic have further negatively impacted children and adolescent's mental health and social development worldwide with reported increases in depressive symptoms, negative affect, loneliness, and lower academic scores, specifically for those that were already at risk prior to the pandemic (Branje et al., 2021).

One of the key settings in which challenges like these can be addressed via targeted intervention and support is schools. In the United States, the past two decades have seen immense growth in the development and provision of social, behavioral, and mental health services to promote the positive development of all children (e.g., Kim et al., 2022; Mahoney et al., 2021). Likewise, there have been several intervention programs developed across Europe, and specifically in German-speaking regions (Cefai et al., 2018). Yet, access to a key feature undergirding evidence-based school service delivery lags behind – assessment. Although there are some social-emotional and mental health assessments available in German, they have limited evidence for validity and are not optimized for applied practice. As such, the current study addressed this need by translating and providing preliminary validity evidence for a brief self-report measure of children's SEL and mental health, the SSIS SEL Brief + Mental Health Scales (SSISb+MHS; Elliott, et al., 2020) with a sample of German and Austrian youth.

Social-Emotional Skills and Mental Health: Conceptualization and Current Practice
Recently, there have been reported increases in psychosomatic issues, mental health problems, and higher anxiety levels in children and adolescents in Germany and Austria (eg., Ravens-Sieberer et al., 2023). To address the mental health needs of children and adolescents, there has been a considerable increase in the promotion of social-emotional and mental health-focused programs in schools across Europe (e.g., WHO, 2018). For example, a recent prominent report considering student social-emotional well-being in Europe

provided extensive recommendations for increased attention to understanding and promoting development of social-emotional well-being throughout European schools (Cefai et al., 2018) concluding that "social and emotional education should be recognized as a core curricular area in the education of children and young people, and one of the major constituents of quality education in Europe." (p. 12).

In response to the increased attention on SEL and mental health, there are an increasing number of programs targeting these domains across Europe and specifically in German-speaking countries. For example, in their comprehensive report on the state of social-emotional education in Europe, Cefai et al. (2018) note that several interventions have been translated and used in Germany including the Faustlos curriculum (Cierpka, 2001) and the ProAC+E program (Spröber et al., 2006). Additionally, in Austria, interventions such as "Starke Kinder - Gute Freunde" and " Eigenständig werden " are recommended and used in schools, although a systematic evaluation of these is still pending (Reicher & Matischek-Jauk, 2018). Despite the growing interest in SEL in these countries, systematic implementation of curricula and interventions in European schools also includes the challenge of assessing if the curricula or intervention had a positive impact on the SEL or mental health needs of the students. As Cefai and colleagues (2021) point out, in the European context there is a great need for a more rigorous assessment of SEL interventions to improve the evidence-base of the available interventions.

Assessing SEL and Mental Health in Austria and Germany

Several SEL and mental health assessments are available in German and used in Austria and Germany. First, one measure assessing SEL, available in German is the Behavioral, Emotional, and Social Skills Inventory (BESSI). The English version BESSI was introduced by Soto et al. (2022) and the German translation and validation were undertaken by Lechner et al. (2022) showing good psychometric properties. Another self-report assessment available in German is the Strengths and Difficulties Questionnaire (SDQ, Goodman, 1997), which includes five subscales assessing prosocial behavior, emotional symptoms, relationship problems with peers, conduct problems, and hyperactivity/attention. In a recent study, Becker et al. (2018) showed that the Germany SDQ self-report had moderate to good psychometric properties.

The Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001) is another tool similar to the SDQ. It assesses various areas of problem behavior and also has a self-report version available in German (YSR/11-18R; e.g., Döpfner et al, 2014). An SEL self-report assessment available in German for students aged 9 to 19 is the Schülereinschätzliste für Sozial- und Lernverhalten (SSL, Petermann & Petermann, 2014). This measure assesses several SEL domains such as empathy, self-management, and self-awareness. The SSL has demonstrated moderate to good validity; however, its content validity was found to be low according to Lohbeck et al. (2014). There is also a German version of the Trait Emotional Intelligence Questionnaire (TEIQue), a 153-item scale designed to capture behaviors and self-perceived strengths

related to emotional intelligence. This version demonstrated good reliability and construct validity when used with a German-speaking sample (Freudenthaler et al., 2008). Three additional measures of mental health that have been validated for use with German-speaking youth ages seven through 17 include the KIDSCREEN-10, the Screen for Child Anxiety Related Disorders (SCARED), and the Center for Epidemiological Studies Depression Scale for Children (CES-DC; Ravens-Sieberer et al., 2023).

Although all these assessments have positive qualities, they have limitations that affect SEL and mental health practices in schools. To function effectively, contemporary school-based service delivery systems require assessment tools that are well adapted to their particular function within a broader multi-tiered student support or MTSS system (e.g., Anthony & DiPerna, 2017; 2018). Measures that were developed for diagnostic or research purposes often lack the content and practical features needed for implementation of SEL or mental health-focused MTSS in schools. Of the assessments noted and available in German, most are long and require significant time to complete, restricted to a specific age group, and few concurrently examine both positive SEL skills and EBCs. This latter feature is particularly problematic as research has emphasized the importance of considering both construct domains (Elliott et al., 2023). Apart from the BESSI, none of the aforementioned assessments comprehensively examine both SEL and mental health. The range of the number of items (and time to complete them) on available assessments is broad, with the shortest scale having only 10 items (KIDSCREEN-10) and the longest scale having 192 items (BESSI). Median scale length for available measures is 40.5 items and only three have fewer than 30 items. Thus, there is a great need for further assessment options to support SEL and mental health-focused school-based practice in Germany and Austria.

The SSIS SELb+MHS Scales

One measure that holds promise to meet this need is the SSIS SEL Brief + Mental Health Scales (SSIS SEL*b*+MHS; Elliott et al., 2020). The SSIS SEL*b*+MHS were developed to meet the need for efficient assessment of SEL and mental health in the U.S. They were developed via the application of Item Response Theory (IRT) to the standardization sample of the SSIS SEL Edition (Gresham & Elliott, 2017). Evidence for reliability and validity is strong and indicates that use of both positive SEL skills and challenging mental health behaviors leads to the identification of students who would not otherwise be identified (e.g., Elliott et al., 2023). A criterion-referenced developmental framework for score interpretation was also developed for the assessment to facilitate communication of results about SEL and EBCs (Elliott et al., 2020).

The SSIS SEL*b*+MHS basic features and efficiency meet the needs for efficient SEL and mental health assessment, but specific research with the measure further enhances its promise to meet assessment needs in Germany and Austria. A series of recent studies (Anthony et al., 2022; 2023) has established that translations of the SSIS SEL*b* (the SEL portion of the SSIS SEL*b*+MHS) function well in several European

countries. Specifically, the SSIS SEL*b* was used as the primary outcome measure of the Promoting Mental Health at Schools (PROMEHS) Project, which evaluated a school-based SEL intervention in six European countries. The SSIS SEL*b* was translated into the official languages of these countries and measurement invariance (Anthony et al., 2023) and cross-country concurrent validity (Anthony et al., 2022) analyses were conducted. Results indicated that scores for the SSIS SEL*b* functioned similarly across these countries, providing a solid basis for further translation and application in other European countries.

Yet, these previous validation studies also had key limitations. Notably, the PROMEHS study was focused exclusively on the SSIS SEL*b*, which includes positive SEL skills only, and German, a language spoken by millions of children in several European countries, was not one of the languages for which translation was conducted. Likewise, no current evidence is available on the SSIS SEL*b*+MHS in the two most populous German-speaking nations, Germany and Austria. Despite sharing a common language, there are important differences between the cultures and educational systems of these countries that may lead to different assessment functioning. For example, Schwab et al. (2017) conducted a study to evaluate the psychometric properties of the Teacher Inclusive Education Self-Efficacy Scale (TIESES) among pre-service teachers in Austria and Germany. Their findings suggest that the scalar invariance model was only partially supported. Schwab et al. (2017) therefore caution that to conduct sound cross-cultural research, it is important to test for measurement invariance, even if the language is the same, and the environmental conditions appear similar.

Rationale, Validity Questions, and Expectations

In the current study, we addressed three validity-related research questions. For two of the questions, results were anticipated to be consistent with previous investigations of students' self-ratings of their social behavior. The third research question was exploratory. Specifically, our questions and anticipated findings were:

1. Is the factor structure of the SSIS SEL*b*+MHS Student invariant across the two included countries (Austria and Germany)? Based on prior work with the SSIS SELb in a host of other European countries (Anthony et al., 2022) we anticipated that the SSIS SEL*b*+MHS Student would evidence a structure similar to that observed in prior research and that this structure would hold across countries.

2. Are known group comparisons across gender (boy vs. girl) equivalent across the two countries? Considering the substantial evidence of gender differences in social behavior and mental health concerns (Eagly, 2013), we anticipated that group differences would be similar across countries and expected that girls would evidence higher levels of positive SEL skills and higher levels of

internalizing EBCs, whereas boys would evidence higher levels of externalizing EBCs in both countries.

3. Are validity coefficients with social inclusion equivalent across the two countries? This exploratory question is based on research showing that social inclusion is significantly associated with both academic achievement and mental well-being for included students (Bücker et al., 2018). Yet, little evidence has evaluated the SEL and mental health of students based on their willingness to include students with disabilities. Thus, we examined concurrent validity correlations between students' SEL/mental health scores and their attitudes towards peers with disabilities.

Methodology

Participants

Participants were 821 3rd through 6th grade (age approx. 8 - 12 years) students in Austria ($N = 268$) and Germany ($N = 553$). Austrian students attended public primary and lower secondary schools in Vienna and were in 14 classes distributed across intervention and waitlist control schools. German students attended public primary schools in rural and urban areas in the western portion of the country. These students were in 32 classes distributed across intervention and waitlist control schools. Available demographic characteristics of students are presented in Table I.

Table I

Demographic Characteristics of Participants

Characteristic	Austrian Sample ($n = 268$)	German Sample ($n = 553$)
Gender		
Boy	53.7	44.1
Girl	46.3	55.9
Grade		
3rd	44.4	47.2
4th	26.9	52.8
5th	13.8	-
6th	14.9	-

Measures

To assess students' social-emotional competencies, we used the SSIS SEL*b*+MHS (Elliott et al., 2020) and an adapted version of the Chedoke-McMaster Attitudes toward Children with Handicap Scale (CATCH) by Rosenbaum, Armstrong, and King (1986; see also de Boer et al. 2014; Schwab, 2015; 2018), consisting

of two subscales, an affective and a behavioral dimension of attitude. Example items/vignettes for these measures and domain definitions can be found in Table II.

Table II

Definitions of SEL Competencies with Example SSIS SELb+MHS-Student Items in English and German

Domain	Definition
Self-Awareness	The ability to accurately recognize one's emotions and thoughts, and their influence on behavior. US English*: I ask for help when I need it.* German: *Ich bitte um Hilfe, wenn ich sie brauche.*
Self-Management	The ability to regulate one's emotions, thoughts, and behaviors effectively in different situations. US English*: I stay calm when dealing with problems.* German: *Ich bleibe ruhig, wenn ich Probleme habe.*
Social- Awareness	The ability to take the perspective of and empathize with others from diverse backgrounds and cultures, to understand social and ethical norms for behavior, and to recognize family, school, and community resources and supports. US English*: I help my friends when they are having a problem.* German: *Ich helfe meinen Freunden und Freundinnen, wenn sie ein Problem haben.*
Relationship Skills	The ability to establish and maintain healthy and rewarding relationships with diverse individuals and groups. US English*: I try to forgive others when they say "sorry."* German: *Ich versuche anderen zu verzeihen, wenn sie "Entschuldigung" sagen.*
Responsible Decision-Making	The ability to make constructive and respectful choices about personal behavior and social interactions based on consideration of ethical standards, safety concerns, social norms, the realistic evaluation of consequences of various actions, and the well-being of self and others. US English*: I am careful when I use things that aren't mine.* German: *Ich bin vorsichtig, wenn ich Dinge benutze, die nicht mir gehören.*
Emotional Behavior Concerns (EBCs)– Externalizing	Negative emotions and behavior mostly directed toward others involving verb or psychological aggression; threatening and bullying of others; poor control of temper; arguing with others; actively excluding others from activities. US English: *I fight with others.* German: *Ich raufe mit anderen.*

Emotional Behavior Concerns (EBCs) - Internalizing	Negative emotions and behaviors mostly directed inwardly involving feeling worried, anxious, sad, or lonely; exhibiting poor self-esteem; lack of interest or limited engagement with others. US English: *I think no one cares about me.* German: *Ich denke, keiner sorgt sich um mich.*
Vignette 1	Alicia s a girl of your age and has just moved to your town. She attends the same class as you. Alicia does not yet understand German well and can hardly speak it.
Vignette 2	Jordan is a boy of your age and has just moved to your town. He attends the same class as you. He does not have any friends. Jordan spends the breaks alone.
Vignette 3	Gerda is a girl of your age and has just moved to your town. She attends the same class as you. Gerda has just started to read and write, but she has difficulties with mathematics. She can play and run like other children, but sometimes forgets the rules of certain games. She needs extra time for solving exercises than the other children and sometimes forgets things. Sometimes it is difficult to understand what Gerda says. For some part of the day, Gerda receives extra learning assistance outside the classroom room.
Vignette 4	Julian is a boy of your age and has just moved to your town. He attends the same class as you. In school, Julian is often restless, fidgety and easily distracted. He often does not follow the teacher's instructions.
Note. All SEL and EBC definitions taken from SSIS SEL*b*+MHS Manual (Elliott et al., 2020). The case vignettes are taken from de Boer et al. (2014) and Schwab (2015; 2018).	

SSIS Brief + Mental Health Scales – Student Form. The SSIS-SEL*b*+MHS is a rating scale of students SEL and mental health that can be completed in less than 5 minutes, is a multi-informant assessment (teacher, parent, and student), and evaluates the social-emotional skills of children and adolescents. We utilized the student version of the SSIS SEL*b*+MHS for this project, which includes 20 SEL-focused items used to generate an overall SEL Composite score, 5 items targeting EBCs in the externalizing behavior domain (EBC-Externalizing), and 5 items targeting EBCs in the internalizing behavior domain (EBC-Internalizing). All items are rated on a Likert scale from 0 *(Never)* to 3 *(Almost Always).* Scores from the SSIS SEL*b*+MHS-Student have evidence for reliability and validity (Elliott et al., 2021). Specifically, reported reliability coefficients from the manual (Elliott et al., 2020) for the SEL Composite, EBC-Externalizing, and EBC-Internalizing scores were .90, .80 and .75 (Cronbach's α); and .87, .60 and .63 (Test-Retest reliability) respectively. Information curves from IRT analyses also supported the reliability of SSIS SEL*b*+MHS scores. Furthermore, extensive validity evidence is available in the form of convergent and discriminant correlations (Elliott et al., 2020); structural validity analyses (Anthony et al., 2021); and known group comparisons (Elliott et al., 2020). Finally, initial work has supported the

reliability and validity of scores from translated versions of the SSIS SELb (the SEL portion of the SSIS SEL*b*+MHS) in European countries (Anthony et al., 2022; 2023), but the current manuscript is the first examination of translated versions of the EBC scales. Cronbach's α for the current sample was .86 for the SSIS SEL Composite, .63 for the EBC-Externalizing Scale, and .70 for the EBC-Internalizing Scale. Single factor models for these scales also yielded ω total coefficients of .87, .67, and .76 for the SEL Composite, EBC-Externalizing Scale and EBC-Internalizing Scale, respectively[2].

For this study, the SSIS SEL*b*+MHS was translated for the first time into German by two PhD students fluent in both German and English. Both translated the items independently and compared the translation, agreeing on the best translation in case of disagreement. The translated questionnaire was reviewed a final time by a professor working in the field of empirical educational research and sent to a colleague who is a German-speaking teacher with a major in English, who translated the items back into English. These items were again compared to the original item and were very similar to the original items. A pretest was also conducted. The questionnaire was given to two children (elementary school) who completed it. One of the doctoral students who did the translation accompanied this process and took notes of which items the children had difficulty understanding. These items were compared one last time with the original items and slightly adapted to make them more understandable for the target group. The colleague who had done the back translation into English did this one last time for the revised items.

Students' Attitudes Towards Their Peers. Students' attitudes towards peers with different characteristics (see Table I) were assessed by using gender-specific case vignettes. The case vignettes for the study were taken or adapted from de Boer et al. (2014) and Schwab (2015; 2018). After reading each case vignette, students were asked to answer different statements (e.g., "During the breaks I would like to play with [NAME]"; I have often had contact with someone like [NAME]") for each case vignette on a four-point Likert scale from 1 (*Not at all True*) to 4 (*Completely True*). The items used in combination with the vignettes have been validated in past research and showed reliable factor structures (e.g., Hellmich & Loeper, 2019; Schwab, 2015; 2017). The items consisted of two constructs - contact and attitudes – (Bossaert & Petry, 2013) and were adapted from the CATCH. The CATCH was initially developed in Canada to measure the attitudes of children aged 9-13 towards their peers with disabilities (Rosenbaum et al., 1986). Since then, the CATCH has been successfully validated for use with other populations and in several other countries (e.g., Alnahdi et al., 2020; Vignes et al., 2008).

---

[2] All coefficients calculated with the *psych* package in R (Revelle, 2023)

Procedure

FRIEND-SHIP: Improving Students' Social Participation in Primary and Secondary Schools across Europe was funded by Erasmus+ (ref. no. 2019-1AT01-KA201-05 1226). The Ethics Committee of the University of Vienna (ref. no. 00602) approved the evaluation of the intervention. Legal guardians received a declaration of informed consent for their children with the right to withdraw from the study without giving any reason or having any consequences. Schools in Austria and Germany were approached and asked if they would participate as either an intervention or waitlist control group. Project members in Austria and Germany administered the assessment in the schools. Before completing the assessments, students were informed about their right not to participate in the study. The evaluation study used a pretest-posttest waitlist control group design with data collected prior to (October 2021) and after (February to March 2022) the FRIEND-SHIP intervention was completed.

Data Analyses

The assessment data was used of all students in both control and intervention groups data collected in October before any intervention work started. These assessments operationalized the key dependent variables investigated (i.e., SEL composite, EBC-Externalizing, EBC-Internalizing, and Social Inclusion) with the primary independent variables being students' gender and country. The intervention program was not a variable of concern because all students' scores were collected prior to intervention implementation. Several steps were followed to conduct data analyses. First, we evaluated data quality and missingness. Second, we evaluated the measurement invariance of the SSIS SEL*b*+MHS across Austria and Germany student groups. Finally, we examined the validity relations (known group comparisons across boys and girls and relations with scores from the adapted CATCH) across the two countries.

Regarding missing data, 2.8% of cases were missing all SSIS SEL*b*+MHS item-level data and were removed from analyses. Of the remaining cases, missing data was low, ranging from 0.2% to 4.2% across cases (median = 2.2%). Given the low level of missingness, we utilized standard pairwise deletion approaches for our measurement invariance analyses. The SSIS SEL*b*+MHS scales is rated on a 4-point scale and as such, data were treated as ordinal according to recommendations from Rhemtulla et al. (2012) and models were fit using a robust weighted least squares estimator (the WLSMV estimator in lavaan; Rosseel, 2012). When conducting measurement invariance testing with ordinal data, there is often data sparsity that precludes appropriate estimation of thresholds in certain groups. In such cases, one method of addressing the problem is by targeted collapsing of categories where feasible (Liu et al., 2017). In our case, data sparsity arose from 2 items (*I help my friends when they are having a problem; I say "thank you" when someone helps me*) as these items did not have any *A Little* and *Not* ratings in Austria respectively. Collapsing categories resolved all issues for these items.

Measurement Invariance Testing. To avoid the complications of other approaches of assuming invariance of some items in order to identify models to be tested, we followed procedures specified by Wu and Estabrook (2016) for identifying constraints for progressively more constrictive measurement invariance models. All analyses were guided by recommendations laid forth by Svetina et al., (2020). Syntax was generated by the meas.Eq.syntax function of the semTools statistical package (Jorgensen et al., 2022) in R (R Core Team, 2013). All models were fit in lavaan (Rosseel, 2012).

With regard to the specifics of the invariance analyses, we first fit a configural invariance model in which only basic structural constraints were imposed across countries. The fit of this model was evaluated relative to standard fit indices as specified by Hu and Bentler (1999). Specifically, we considered CFI/TLI values greater than or equal to .95, RMSEA values less than or equal to .06 and SRMR values less than or equal to .08 as indications of good overall model fit. As recommended, we used these as guidelines rather than hard-and-fast rules. Once an acceptable fit of the configural model was established, we proceeded to test a model in which thresholds were constrained to equality across groups; a model in which thresholds and loadings were constrained to equality across countries (metric invariance); and a model in which thresholds, loadings, and item intercepts were constrained to equality across countries (scalar invariance). Based on Svetina et al. (2020) procedures we compared progressively more constrained measurement invariance models primarily via ΔRMSEA and ΔCFI values (Δχ² were deemphasized due to sample size sensitivity). With regard to ΔRMSEA, we used a threshold of .015 decrement between models (Chen, 2007) and with regard to ΔCFI we used a threshold of .01 decrement between models as our primary indicators of acceptable fit of progressively more constrained measurement models. Once a final measurement invariance model had been supported, we also constrained latent variable correlations across countries to determine whether factor interrelationships were similar across Austria and Germany.

Cross Country Known Group Comparisons and Concurrent Validity Comparisons. We conducted known group comparisons and concurrent validity analyses across countries. Our analytical approach was informed by the general approach for evaluating prediction bias outlined by Lautenschlager and Mendoza (1986). Specifically, we focused on slope differences across groups because the slope best captures concurrent validity differences. Thus, we used hierarchical regression to compare a model in which only the focal predictor (e.g., gender dummy variable and CATCH scores) and a dummy variable indicating country (Austria or Germany) were included with a model in which the focal predictor, dummy variable indicating country, and the interaction between the two variables were included. We conducted separate models for each outcome variable (i.e., score produced by the SSIS SEL*b*+MHS; SEL Composite scores; EBC-Internalizing Scores; EBC-Externalizing Scores) as well as each validity variable (i.e., gender; each of the 4 Case Vignette scores) for a total of 15 models. Moderation was evaluated by ΔR^2 values with

statistically significance indicating differences in slopes and therefore concurrent validity coefficients across countries.

Results

Descriptive statistics for all analytic variables can be found in Table III. These scores on the primary variables investigated provide the foundations for all subsequent analyses.

Table III

Descriptive Statistics across Gender and Country

Characteristic	Austrian Sample				German Sample			
	Girls		Boys		Girls		Boys	
	M	SD	M	SD	M	SD	M	SD
SSIS SEL*b*+MHS								
SEL Composite	2.37	0.37	2.28	0.37	2.45	0.35	2.27	0.40
EBC-Externalizing	0.76	0.60	0.88	0.54	0.58	0.48	0.81	0.65
EBC-Internalizing	1.00	0.71	0.97	0.67	0.81	0.65	0.82	0.68
CATCH								
Vignette 1	3.06	0.55	2.91	0.67	3.06	0.46	2.76	0.60
Vignette 2	3.05	0.59	3.00	0.69	3.04	0.49	2.80	0.64
Vignette 3	2.91	0.64	2.89	0.75	2.89	0.60	2.67	0.71
Vignette 4	2.76	0.74	2.67	0.85	2.71	0.68	2.46	0.82

Note. Mean scores reported. SSIS SEL*b*+MHS = SSIS SEL Brief + Mental Health Scales (Rated on a scale of 0 [*Never*] to 3 [*Almost Always*]); CATCH = Chedoke-McMaster Attitudes towards Children with Handicaps Scale. (Rated on a scale of 1 [*Not at all true*] to 4 [*Completely true*]).

Measurement Invariance Testing

We began measurement invariance testing by evaluating the original five factor correlated factors model of the SSIS SELb. The baseline configural invariance model did not converge with several indications of exceedingly high intercorrelations between SEL factors, as has been observed in similar research on the SSIS SELb-S (Anthony et al., 2023). Thus, we sequentially collapsed factors until a baseline model converged after collapsing Self-Awareness, Self-Management, Relationship Skills, and Responsible Decision-Making factors as had been done in Anthony et al. (2023) resulting in a four-factor correlated model with the merged factor, a Social Awareness factor, and factors for the EBC-Externalizing and EBC-

Internalizing scales[3]. The configural model evidenced adequate fit and subsequent measurement invariance models did not evidence substantial decrements in model fit indicating support for full scalar invariance across Germany and Austria (Table IV). Constraining latent variable correlations to equality across countries also led to improved overall model fit (χ^2 = 1,359.43, *df* = 884, p < .001; RMSEA = .037; SRMR = .071; CFI = .944; TLI = .945) and comparative fit relative to the final measurement invariance model ($\Delta\chi^2$ = 2.81; df = 6; *p* = .83). Unstandardized loadings and interfactor correlations for this final model are presented in Figure 1. As illustrated by this figure, the SEL-Student version items cluster into 2 factors and the EBC-Externalizing and EBC-Internalizing items create 2 more separate factors.

Table IV

Measurement Invariance Fit Statistics for Correlated Factors Models

	χ^2	df	*p*	RMSEA	CFI	TLI	$\Delta\chi^2$	df	*p*	ΔRMSEA	ΔCFI
Configural	1,321.93	798	<.001	.040	.939	.933	-	-	-	-	-
Thresholds	1,355.94	826	<.001	.040	.938	.935	35.52	28	.16	<.001	-.001
Thresholds & Loadings (Metric)	1,362.13	852	<.001	.039	.940	.939	32.27	26	.18	.001	.004
Thresholds, Loadings, & Intercepts (Scalar)	1,406.99	878	<.001	.039	.938	.939	66.00	26	<.001	<.001	-.002
Note. Negative ΔCFI and ΔRMSEA values indicate increases in these indices relative to previous model.											

Known Group and Concurrent Validity Analyses

Regarding known group and concurrent validity comparisons across countries, results can be found in Table V (known gender and country group comparisons) and Tables VI and VII (concurrent validity correlations with social inclusion vignettes). The results featured in these tables provided support for the conclusion that known group comparisons and concurrent validity correlations were statistically similar across Germany and Austria. Specifically, *p*-values for ΔR^2 values were all statistically nonsignificant ranging from .115 to .948. These results support the conclusion that validity evidence across Germany and Austria is similar.

Regarding the validity coefficients themselves, the direction and magnitude of coefficients was generally in line with expectations. That is, boys rated themselves as having lower SEL behaviors and higher EBC-Externalizing Behaviors (this coefficient became nonsignificant in the interaction model, but

---

[3] This model also provided superior fit to a model in which SEL was included as a unidimensional construct ($\chi2$ = 1,383.63, df = 804, p < .001; RMSEA = .042; SRMR = .072; CFI = .932; TLI = .927; $\Delta\chi^2$ relative to 2 factor model = 62.50, Δdf = 6, *p* < .001)

was statistically significant in the original model), although differences between boys' and girls' ratings of their EBC-Internalizing Behaviors were not statistically significant (Table V). Similarly, scores for the 4 separate case vignettes were positively related to students' SEL Composite ratings but accounted for only a small percentage of variance. The social inclusion scores accounted for a still smaller percentage of variance and were inconsistently related to students' EBC-Externalizing and EBC-Internalizing self-ratings (Tables VI & VII).

Table V

Hierarchical Regression Models for Known Group Gender Comparisons across Country

	Model 1		Model 2	
	B	***p***	**B**	***p***
SEL Composite				
Intercept	2.40	<.001	2.37	<.001
Gender	-0.15	<.001	-0.09	.048
Country	0.03	.241	0.08	.051
Gender* Country	-	-	-0.09	.115
R^2	0.04	<.001	0.04	<.001
ΔR^2	-	-	0.003	.115
EBC-Externalizing				
Intercept	0.72	<.001	0.76	<.001
Gender	0.20	<.001	0.12	.073
Country	-0.13	.003	-0.18	.002
Gender* Country	-	-	0.11	.197
R^2	0.04	<.001	0.04	<.001
ΔR^2	-	-	0.002	.197
EBC-Internalizing				
Intercept	0.98	<.001	1.00	<.001
Gender	-0.004	.933	-0.03	.676
Country	-0.17	<.001	-0.19	.008
Gender* Country	-	-	0.05	.650
R^2	0.01	.004	0.01	.010
ΔR^2	-	-	<0.001	.650
Note. Boy = 1; Germany = 1; Model 1 = Model without Interaction Term; Model 2 = Model with Interaction Term				

Figure 1

Final Scalar Invariance and Constrained Latent Variable Correlation Model for SSIS SELb+MHS Student
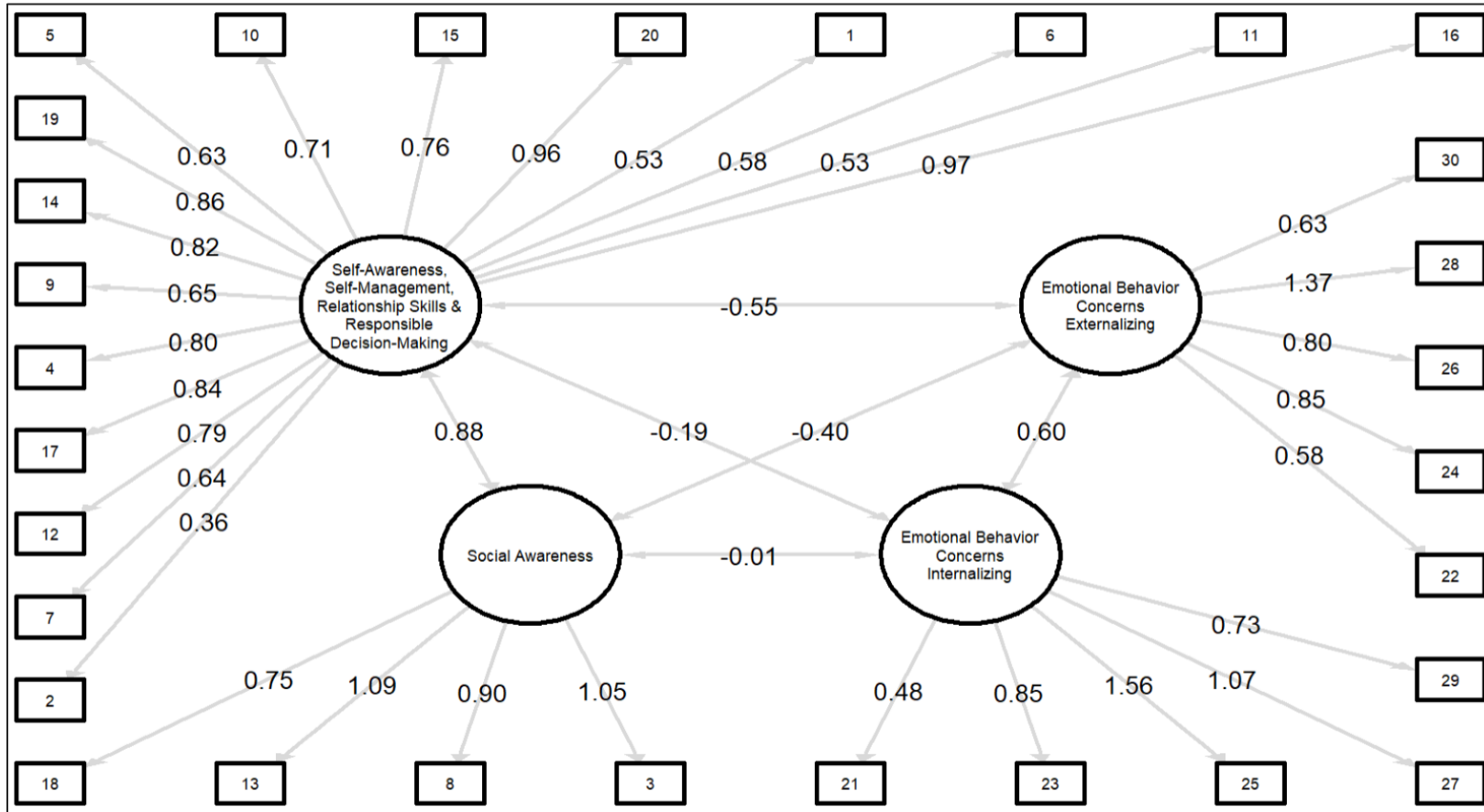
Table VI

Hierarchical Regression Models for Case Vignette 1 and 2 Concurrent Validity Comparisons across Country

	Case Vignette 1				Case Vignette 2			
	Model 1		Model 2		Model 1		Model 2	
	B	*p*	**B**	*p*	**B**	*p*	**B**	*p*
SEL Composite								
Intercept	1.60	<.001	1.62	<.001	1.62	<.001	1.68	<.001
CATCH	0.24	<.001	0.24	<.001	0.23	<.001	0.21	<.001
Country	0.06	.033	0.03	.826	0.07	.013	-0.04	.765
CATCH * Country	-	-	0.01	.837	-	-	0.04	.412
R^2	0.13	<.001	0.13	<.001	0.14	<.001	0.14	<.001
ΔR^2	-	-	<0.001	.837	-	-	<0.001	.412
EBC-Externalizing								
Intercept	1.14	<.001	1.02	<.001	1.11	<.001	1.13	<.001
CATCH	-0.11	.002	-0.06	.236	-0.10	.005	-0.10	.061
Country	-0.16	<.001	0.05	.808	-0.16	<.001	-0.18	.387
CATCH * Country	-	-	-0.07	.327	-	-	0.01	.910
R^2	0.03	<.001	0.03	<.001	0.03	<.001	0.03	<.001
ΔR^2	-	-	0.001	.327	-	-	<0.001	.910
EBC-Internalizing								
Intercept	0.61	<.001	0.67	<.001	0.62	<.001	0.73	<.001
CATCH	0.12	.003	0.11	.107	0.12	.003	0.09	.174
Country	-0.17	.001	-0.26	.307	-0.16	.001	-0.34	.175
CATCH * Country	-	-	0.03	.710	-	-	0.06	.477
R^2	0.03	<.001	0.03	<.001	0.03	<.001	0.03	<.001
ΔR^2	-	-	<0.001	.710	-	-	<0.001	.463
Note. Germany = 1; Model 1 = Model without Interaction Term; Model 2 = Model with Interaction Term								

Table VII

Hierarchical Regression Models for Case Vignette 3 and 4 Concurrent Validity Comparisons across Country

	Case Vignette 3				Case Vignette 4			
	Model 1		Model 2		Model 1		Model 2	
	B	*p*	**B**	*p*	**B**	*p*	**B**	*p*
SEL Composite								
Intercept	1.86	<.001	1.78	<.001	2.11	<.001	2.06	<.001
CATCH	0.16	<.001	0.19	<.001	0.08	<.001	0.10	<.001
Country	0.06	.024	0.19	.122	0.06	.053	0.14	.163
CATCH * Country	-	-	-0.04	.296	-	-	-0.03	.374
R^2	0.08	<.001	0.08	<.001	0.03	<.001	0.03	<.001
ΔR^2	-	-	0.001	.296	-	-	0.001	.374
EBC-Externalizing								
Intercept	0.89	<.001	0.81	<.001	0.79	<.001	0.76	<.001
CATCH	-0.02	.456	0.01	.895	0.01	.601	0.02	.593
Country	-0.15	<.001	-0.02	.917	-0.15	<.001	-0.11	.475
CATCH * Country	-	-	-0.05	.457	-	-	-0.02	.786
R^2	0.02	.002	0.02	.004	0.02	.002	0.02	.006
ΔR^2	-	-	<0.001	.457	-	-	<.001	.786
EBC-Internalizing								
Intercept	0.63	<.001	0.63	<.001	0.89	<.001	0.87	<.001
CATCH	0.12	<.001	0.12	.037	0.03	.290	0.04	.455
Country	-0.16	.001	-0.15	.494	-0.17	<.001	-0.15	.411
CATCH * Country	-	-	-0.005	.948	-	-	-0.01	.899
R^2	0.03	<.001	0.03	<.001	0.02	.002	0.02	.005
ΔR^2	-	-	<.001	.948	-	-	<0.001	.899
Note. Germany = 1; Model 1 = Model without Interaction Term; Model 2 = Model with Interaction Term								

Discussion

This study examined fundamental psychometric features of a German translation of the SSIS SEL*b*+MHS – Student version (Elliott et al., 2020) with elementary school boys and girls in Germany and Austria. This assessment has been demonstrated to yield reliable, valid, and fair scores for English and Spanish-speaking students ages 8 to 18. With the growing global interest in children's social-emotional learning and health, educational researchers in 9 European countries have translated and used the SSIS SEL Brief Scales to evaluate various SEL intervention programs' effects on students' social competence. Until the present study, however, the expanded Mental Health version of these Brief SEL Scales had not been translated, nor had any version of its Emotional Behavior Concern Scales been used with European samples of students. Thus, this investigation

of the SSIS SEL*b*+MHS Student version with German-speaking children contributed new evidence about the validity and utility of this assessment.

Three research questions focused this investigation. Specifically, we collected students' self-reported ratings regarding their SEL skills, externalizing behavior-related, and internalizing behavior-related concerns to address questions about (a) the invariance of the factor structure of the assessment, (b) how scores for boys and girls on these subscales compared, and (c) how the students' SEL/mental health scores correlated with their scores on a social inclusion measure.

Factor structure expectations. We found strong evidence to support our expectation that the German translated Student version of SSIS SEL*b*+MHS would possess a factor structure like that observed in prior research and the structure would be consistent for German and Austrian students sampled. Specifically, based on the elementary students' self-ratings, the SEL*b*+MHS items clustered into 2 SEL factors and 2 EBC factors. The SEL factors were characterized as Social Awareness and a combination of the four other SEL scales (Self-Awareness, Self-Management, Relationship Skills, and Responsible Decision-Making). The EBC items formed 2 neat clusters with one factor representing Internalizing behaviors and the other factor representing Externalizing behaviors.

The factor analytic results for the German-translated SEL items replicate almost exactly what Anthony et al. (2023) reported for larger samples of students from 6 other European countries. The results for the two Mental Health Scales, EBC-Internalizing and EBC-Externalizing, were new for translated versions of the SSIS SEL*b*+MHS but functioned entirely consistent with the expectations for 2 factors, both of which negatively correlate with SEL factors.

Similar to prior work by Anthony et al. (2023; 2022), the five-factor proposed structure aligned with CASEL led to estimation problems due to very high interfactor correlations. Such a conclusion is in line with a growing number of similar investigations with other measures in other contexts (e.g., Doromal et al., 2017) in which CASEL domains have been highly correlated or redundant. Given these investigations, strong empirical evidence of the distinguishability of these constructs is much needed. Yet, other, more simple, theoretical models (e.g., a unidimensional model) are also questionable, both from theoretical and empirical grounds. It is likely that the SEL Composite from the SSIS SEL*b*+MHS can be used as a general indicator of SEL skills, but it is clear that much more work remains to establish the scientific foundations for the structure of SEL domains, at least the structure that can be ascertained via behavior rating scales.

Gender difference expectations. We expected that girls would evidence higher levels of positive SEL skills and higher levels of internalizing EBCs, whereas boys would evidence higher levels of externalizing EBCs in both countries and that differences would be of a similar magnitude across countries. This set of expectations was largely supported, although girls and boys sampled did not differ significantly regarding their EBC-Internalizing scores. Given our sample was elementary school-age students, and disproportionately in grades 3 and 4, the finding of no difference on internalizing behavior is not surprising. Thus, it seems elementary girls in both Austria and Germany, and in several other European countries, rate their own positive social-emotional behaviors, on average, consistently higher than same-age boys. And boys, on average, rate

their externalizing behaviors are being more frequently exhibited than do girls in the same grades. This pattern finding is also consistent with research on US girls' and boys' self-ratings of their social-emotional skills and externalizing emotional behavior.

Relationship between SEL and social inclusion scores. Using the adapted CATCH as our measure of social inclusion, we explored the relationship among it and the elementary students' self-ratings of social-emotional functioning. The lack of differential concurrent relationships across countries further supported the conclusion that scores function similarly across countries. It is also interesting to examine the overall magnitude of score relations. Although there has not been extensive prior evaluation in the literature regarding the relationship between SEL and social inclusion, we generally expect that students with higher SEL scores would also have higher social inclusion scores. This expectation was largely borne out, although the magnitude of effects was fairly small and did vary across case vignettes. Relationships between CATCH scores and EBC scores were not statistically significant. Although we did not form hypotheses about these relationships, the current findings are unsurprising because students' emotional and behavioral difficulties do not have as strong of a logical link with social inclusion as SEL. For example, a constituent aspect of SEL includes social awareness, which involves empathic reactions to others that are likely a core feature of social inclusion. Similar logical relationships are not apparent for EBCs. Future research could explore how SEL is related to social inclusion. Indeed, SEL could be an important intervention target for promoting social inclusion in schools in Austria, Germany, and beyond.

Limitations and Future Research

Despite the promising results from this initial study of the validity of the German translation of the SSIS SEL Brief + Mental Health Scales (Student version), there are limitations. First, although score reliability was adequate for the SEL Composite and EBC-Internalizing scores, it was low (Cronbach's $\alpha = .63$; Coefficient ω total = .67) for the EBC-Externalizing score. This low reliability would have less impact on our factor analytic findings as construct relationships with latent variables parse out measurement error but could potentially have been related to the lack of observed relationships between CATCH and EBC-Externalizing scores and gender via diminished statistical power. The reason for this lower reliability is unclear, but there are several possibilities. First, the children in this sample were relatively young and may not have consistent insight into their behavior. Next, it is possible that some item translations resulted in confusing content that was not culturally relevant or clear. For example, children anecdotally shared confusion on two items on the EBC-Externalizing scale (*I fight with others; I do not let others join my group of friends.*) and only one other item on any other scale (*I do my part in a group*). The fact that the EBC-Externalizing scale is short could magnify the impact of small misunderstandings. Future research should corroborate these data with older children and attempt to understand the cultural connotations of all items at a deeper level. Until further data are used, the EBC-Externalizing score could be used without modification, but great caution should be exercised, especially in individual decision-making contexts.

Next, we only used the student version of SSIS SEL*b*+MHS for the current study. Although most students can be reliable raters, it is useful to have concurrent ratings by an adult, such as their teacher or a parent. Future studies should seek to validate German translations of the SSIS SEL*b*+MHS for Teacher and Parent versions. This is particularly important because of the low reliability evidenced by the EBC-Externalizing scale for this sample. Obtaining multi-rater data will also allow for inter-rater comparisons and additional forms of reliability evidence. Finally, as noted, future research should continue to explore the relationships between SEL, mental health, and social inclusion.

Conclusion

This study focused on the validity of the German translation of the SSIS SEL Brief + Mental Health Scales - Student version because it was a key outcome measure for the FRIEND-SHIP Project. The results indicated this assessment yielded meaningful and reliable scores for an elementary sample of girls and boys from Germany and Austria. Importantly, these results were highly consistent with original psychometric studies conducted with large samples of English-speaking students and more recently with samples of students speaking other languages common to Europe (e.g., Greek, Italian, Portuguese, Romanian). Thus, there is now evidence to support the use of the SSIS SEL*b*+MHS in future studies concerning German-speaking students' social emotional well-being.

The exploration of the relationship among the variables of students' social-emotional functioning and their social inclusion of others was less affirming but is worthy of further study with additional measures of inclusion and older samples of students. Gaining insights into students' perceptions of how they are feeling, functioning, and interacting with their classmates remains a challenge essential to address. This study contributed to this knowledge base and provided support for an assessment tool for researchers examining German speaking students' social emotional health.

Conflict of interest

We wish to declare a potential conflict of interest as two of the authors, Christopher J. Anthony and Stephen N. Elliott are authors of the SSIS SEL Brief + Mental Health Scales and receive royalties for its sale in the United States.

References

Achenbach, T. M., & Rescorla, L. A. (2001). *Child behavior checklist for ages 6-18* (pp. 6-1). University of Vermont

Alnahdi, G. H. (2020). The construct validity of the Arabic version of the Chedoke-Mcmaster attitudes towards children with handicaps scale, *Cogent Education, 7*(1), 1745540. http://dx.doi.org/10.1080/2331186X.2020.1745540

Anthony, C. J., & DiPerna, J. C. (2017). Identifying sets of maximally efficient items from the Academic Competence Evaluation Scales—Teacher Form. *School Psychology Quarterly, 32*(4), 552-559. https://psycnet.apa.org/doi/10.1037/spq0000205

Anthony, C. J., & DiPerna, J. C. (2018). Piloting a short form of the academic competence evaluation scales. *School Mental Health, 10*, 314-321. https://psycnet.apa.org/doi/10.1007/s12310-018-9254-7

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. https://doi.org/10.1177/0734282920953240

Anthony, C. J., Elliott, S. N., Yost, M., Lei P-W., DiPerna, J. C., Cefai, C., Camilleri, L., Bartolo, P. A., Grazzani, I., Ornaghi, V., Cavioni, V., Conte, E., Vorkapić, S. T., Poulou, M., Martinsone, B., Simões, C., & Colomeischi, A. A. (2022). Multi-informant validity evidence for the SSIS SEL Brief Scales across six European countries. *Front. Psychol.* 13, 928189. https://doi.org/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: An examination of the SSIS SEL Brief Scales with a multi-informant sample from six countries. *European Journal of Psychological Assessment.* https://psycnet.apa.org/doi/10.1027/1015-5759/a000753

Becker, A., Wang, B., Kunze, B., Otto, C., Schlack, R., Hölling, H., Ravens-Sieberer, U., Klasen, F., Rogge, J., Isensee, C., Rothenberger, A., & Bella Study Group, T. (2018). Normative Data of the Self-Report Version of the German Strengths and Difficulties Questionnaire in an Epidemiological Setting. *Zeitschrift fur Kinder- und Jugendpsychiatrie und Psychotherapie, 46*(6), 523–533. https://doi.org./10.1024/1422-4917/a000589

Bossaert, G., & Petry, K. (2013). Factorial validity of the Chedoke-McMaster Attitudes towards Children with Handicaps scale (CATCH). *Research in Developmental Disabilities*, 34(4), 1336–1345. https://doi.org/10.1016./j.ridd.2013.01.007

Branje, S., & Morris, A. S. (2021). The Impact of the COVID-19 Pandemic on Adolescent Emotional, Social, and Academic Adjustment. *Journal of research on adolescence. 31*(3), 486–499. https://doi.org/10.1111/jora.12668

Bücker, S., Nuraydin, S., Simonsmeier, B. A., Schneider, M., & Luhmann, M. (2018). Subjective well-being and academic achievement: A meta-analysis. *Journal of Research in Personality, 74,* 83-94. https://psycnet.apa.org/doi/10.1016/j.jrp.2018.02.007

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.

Cefai C., Simões C., & Caravita, S. (2021). *A Systemic, Whole-School Approach to Mental Health and Well-Being in Schools in the EU*. NESET Report, Luxembourg: Publications Office of the European Union

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. https://doi.org/10.1080/10705510701301834

Cierpka, M. (2001). *FAUSTLOS Ein Curriculum zur Pravention von aggressivem und gewaltbereitem Verhalten bei Kindern der Klassen 1 bis 3*. Hogrefe

Döpfner, M., Plück, J., & Kinnen, C. (2014). *CBCL/6-18R, TRF/6-18R, YSR/11-18R. Deutsche Schulalter-Formen der Child Behavior Checklist von Thomas M. Achenbach*. Hogrefe

de Boer, A., de, Pijl, S. J., Minnaert, A., & Post, W. (2014). Evaluating the effectiveness of an intervention program to influence attitudes of students towards peers with disabilities. *Journal of Autism and Developmental Disorders, 44*(3), 572–583. https://doi.org/10.1007/s10803-013-1908-6

Doromal, J. B., Cottone, E. A., & Kim, H. (2017). Preliminary Validation of the Teacher-Rated DESSA in a Low-Income, Kindergarten Sample. Journal of Psychoeducational Assessment. Advance online publication. https://doi.org/10.1177/0734282917731460

Durlak, J. A., & Weissberg, R. P. (2011). Promoting social and emotional development is an essential part of students' education. *Human Development, 54*(1), 1-3. https://psycnet.apa.org/doi/10.1159/000324337

Eagly, A. H. (2013). *Sex differences in social behavior: A social-role interpretation*. Psychology Press.

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

Elliott, S. N., Lei, P.-W., Anthony, C. J., & DiPerna, J. C. (2021). Screening the whole social emotional child: Integrating emotional behavior concerns to expand the utility of the SSIS SEL brief scales. *School Psychology Review, 52(1), 15–29.* https://doi.org/10.1080/2372966X.2020.1857659

Elliott, S. N., Lei, P. W., Anthony, C. J., & DiPerna, J. C. (2023). Screening the whole social-emotional child: Expanding a brief SEL assessment to include emotional behavior concerns. *School Psychology Review, 52*(1), 15-29. http://dx.doi.org/10.1080/2372966X.2020.1857659

Freudenthaler, H. H., Neubauer, A. C., Gabler, P., Scherl, W. G., & Rindermann, H. (2008). Testing and validating the trait emotional intelligence questionnaire (TEIQue) in a German-speaking sample. *Personality and Individual Differences, 45*(7), 673-678 https://psycnet.apa.org/doi/10.1016/j.paid.2008.07.014

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. Journal of *Child Psychology and Psychiatry, 38*(5), 581–586. https://doi.org/10.1111/j.1469-7610.1997.tb01545.x

Gresham, F. M., & Elliott, S. N. (2017). *Social skills improvement system social emotional learning edition rating forms*. Pearson Assessments

Hellmich, F., & Loeper, M. F. (2019). Children's attitudes towards peers with learning disabilities – the role of perceived parental behaviour, contact experiences and self-efficacy beliefs, *British Journal of Special Education, 46*(2). https://doi-org./10.1111/1467-8578.12259

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

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

Kim, E. K., Anthony, C. J., & Chafouleas, S. M. (2022). Social, emotional, and behavioral assessment within tiered decision-making frameworks: Advancing research through reflections on the past decade. *School Psychology Review, 51*(1), 1–5. https://doi.org/10.1080/2372966X.2021.1907221

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. https://psycnet.apa.org/doi/10.1177/014662168601000202

Lechner, C. M., Knopf, T., Napolitano, C. M., Rammstedt, B., Roberts, B. W., Soto, C. J., & Spengler, M. (2022). The behavioral, emotional, and social skills inventory (BESSI): Psychometric properties of a German-language adaptation, temporal stabilities of the skills, and associations with personality and intelligence. *Journal of Intelligence, 10*(3), 63. http://dx.doi.org/10.3390/jintelligence10030063

Lohbeck, A., Nitkowski, D., Petermann, F., & Petermann, U. (2014). Erfassung von Schülerselbsteinschätzungen zum schulbezogenen Sozial-und Lernverhalten–Validierung der Schülereinschätzliste für Sozial-und Lernverhalten (SSL). *Zeitschrift für Erziehungswissenschaft, 4*(17), 701-722. http://dx.doi.org/10.1007/s11618-014-0582-6

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. https://doi.org/10.1016/j.visinf.2017.01.006

Mahoney, J. L., Weissberg, R. P., Greenberg, M. T., Dusenbury, L., Jagers, R. J., Niemi, K., ... & Yoder, N. (2021). Systemic social and emotional learning: Promoting educational success for all preschool to high school students. *American Psychologist, 76*(7), 1128. http://dx.doi.org/10.1037/amp0000701

Petermann, U., & Petermann, F. (2014). Schülereinschätzliste für Sozial-und Lernverhalten: SSL; Manual. Hogrefe

R Core Team, R. (2013). R*: A language and environment for statistical computing*

Ravens-Sieberer, U., Devine, J., Napp, A. K., Kaman, A., Saftig, L., Gilbert, M., ... & Erhart, M. (2023). Three years into the pandemic: results of the longitudinal German COPSY study on youth mental health and health-related quality of life. *Frontiers in Public Health, 11,* 1129073. https://doi.org/10.3389/fpubh.2023.1129073

Reicher, H., & Matischek-Jauk, M. (2018). Sozial-emotionales Lernen in der Schule Konzepte – Potenziale – Evidenzbasierung. In M. Huber, S. Krause, (Eds.), *Bildung und Emotion*, (pp 249-268). Springer. https://doi.org/10.1007/978-3-658-18589-3_14

Revelle, W. (2023). Package "psych." http:// personality-project.org/r/psych_manual.pdf

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), 354-373 https://psycnet.apa.org/doi/10.1037/a0029315

Rosenbaum, P. L., Armstrong, R. W., & King, S. M. (1986). Children's attitudes toward disabled peers: A self-report measure. *Journal of Pediatric Psychology, 11*(4), 517-530 https://doi.org/10.1093/jpepsy/11.4.517

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of statistical software, 48,* 1-36. https://doi.org/10.18637/jss.v048.i02

Schwab, S. (2015). Einflussfaktoren auf die Einstellung von SchülerInnen gegenüber Peers mit unterschiedlichen Behinderungen [Factors influencing students' attitudes towards peers with different disabilities]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 47*(4), 177–187. https://doi.org/10.1026/0049-8637/a000134

Schwab, S., Hellmich, F., & Görel, G. (2017). Self-efficacy of prospective Austrian and German primary school teachers regarding the implementation of inclusive education. *Journal of Research in Special Educational Needs, 17*(3), 205–217. https://doi.org/10.1111/1471-3802.12379

Schwab, S. (2018). *Attitudes towards Inclusive Schooling. A study on Students', Teachers' and Parents' attitudes.* Waxmann Verlag

Soto, C. J., Napolitano, C. M., Sewell, M. N., Yoon, H. J., & Roberts, B. W. (2022). An integrative framework for conceptualizing and assessing social, emotional, and behavioral skills: The BESSI. *Journal of Personality and Social Psychology, 123*(1), 192. https://doi.org/10.1037/pspp0000401

Spröber, N., Schlottke, P. F., & Hautzinger, M. (2006). ProACT+ E: Ein Programm zur Prävention von "bullying "an Schulen und zur Förderung der positiven Entwicklung von *Schülern. Zeitschrift für Klinische Psychologie und Psychotherapie, 35*(2), 140-150. http://dx.doi.org/10.1026/1616-3443.35.2.140

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 http://dx.doi.org/10.1080/10705511.2019.1602776

Vignes, C., Coley, N., Grandjean, H., Godeau, E., & Arnaud, C. (2008). Measuring children's attitudes towards peers with disabilities: a review of instruments. *Developmental Medicine & Child Neurology, 50*(3), 182-189. https://doi.org/10.1111/j.1469-8749.2008.02032.x

World Health Organization. Mental Health Atlas 2017. (2018). https://www.who.int/publications/i/item/9789241514019 (cited November 14, 2022).

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. https://doi.org/10.1007/s11336-016-9506-0