Knowing Every Child: Validation of the Holistic Student Assessment (HSA) as a Measure of Social-Emotional Development

Tina Malti[1] · Antonio Zuffianò[2] · Gil G. Noam[3]

Abstract Knowing every child's social-emotional development is important as it can support prevention and intervention approaches to meet the developmental needs and strengths of children. Here, we discuss the role of social-emotional assessment tools in planning, implementing, and evaluating preventative strategies to promote mental health in all children and adolescents. We, first, selectively review existing tools and identify current gaps in the measurement literature. Next, we introduce the Holistic Student Assessment (HSA), a tool that is based in our social-emotional developmental theory, The Clover Model, and designed to measure social-emotional development in children and adolescents. Using a sample of 5946 students (51% boys, $M_{age} = 13.16$ years), we provide evidence for the psychometric validity of the self-report version of the HSA. First, we document the theoretically expected 7-dimension factor structure in a calibration sub-sample ($n = 984$) and cross-validate its structure in a validation sub-sample ($n = 4962$). Next, we show measurement invariance across development, i.e., *late childhood* (9- to 11-year-olds), *early adolescence* (12- to 14-year-olds), and *middle adolescence* (15- to 18-year-olds), and evidence for the HSA's construct validity in each age group. The findings support the robustness of the factor structure and confirm its developmental sensitivity. Structural equation modeling validity analysis in a multiple-group framework indicates that the HSA is associated with mental health in expected directions across ages. Overall, these findings show the psychometric properties of the tool, and we discuss how social-emotional tools such as the HSA can guide future research and inform large-scale dissemination of preventive strategies.

Keywords Social-emotional assessment tool · Child and youth mental health · Social-emotional development theory · Prevention and intervention planning

✉ Tina Malti
tina.malti@utoronto.ca

[1] Department of Psychology, University of Toronto, Deerfield Hall, 3359 Mississauga Road North, Mississauga, ON L5L 1C6, Canada

[2] Department of Psychology, Liverpool Hope University, Liverpool, UK

[3] Harvard Medical School, McLean Hospital, Belmont, MA, USA

Developmental research has provided ample evidence that children's and adolescents' social-emotional skills (e.g., sympathy, age-appropriate emotion understanding, emotion regulation, etc.) are important factors of resilience that can prevent the development of psychopathology and decrease existing behavioral problems. These beneficial effects have been acknowledged in contemporary prevention practices with children and adolescents, and many strengths-based approaches to preventive interventions have been developed, implemented, and evaluated over the past decade. In particular, school-based Social Emotional Learning programs (SEL) have been shown moderate effect sizes in promoting positive social and academic outcomes (Greenberg et al. 2003; Payton et al. 2008) as well as statistically significant decreases in behavioral problems (Durlak et al. 2011).

Given this empirical evidence regarding prevention and intervention programs, it is surprising that the use of systematic screening and assessment tools to guide student selection for the groups and other school-based practices is less developed. The majority of instruments commonly used to assess children's and adolescents' functioning in school contexts have focused on risk and psychopathology, such as bullying, aggression, and emotional symptoms (e.g., Achenbach and Rescorla 2001). What are missing are reliable and valid

instruments that closely align with social-emotional development research and commonly used preventive practices. Therefore, the development and implementation of such tools is warranted to conduct needs assessments, to monitor the implementation and success of SEL practices over time, and to help schools identify tools that could be useful in determining not only pathologies but also assets connected with the success of the preventive practices. Thus, the purpose of this article is threefold: (1) to explain why we need tools to understand and measure child and adolescent social-emotional development in the context of school-based preventive intervention planning, implementation, and evaluation; (2) to introduce the psychometric properties and developmental sensitivity of a new assessment instrument that focuses on children's and adolescents' social-emotional strengths in educational contexts; (3) to discuss the tool's utility in supporting large-scale dissemination of preventive interventions for children and adolescents.

The Need for Developmentally Sensitive, Social-Emotional Assessment Tools

Much research indicates that social-emotional skills, such as normative levels of empathy, play a central role in the development of prosocial orientations and the reduction of problem behaviors across development (Eisenberg et al. 2015). From a practical perspective, understanding each child's social-emotional development helps practitioners identify the child's strengths and needs, as well as how these change over time. In addition to individual empirical portraits of children, assessments can also provide information on strengths of a diverse group of children in a classroom (or an entire school community) that might not be easily detected without a strong data approach. As such, developmentally sensitive assessments can improve the use of intervention strategies that fit the developmental needs of children and adolescents (Malti et al. 2016).

In the past decade, much progress has been made in the design of tailored interventions for children and adolescents. For instance, age-graded differences have been considered, to some extent, in the design and implementation of school-based SEL programs (Durlak et al. 2011). Before systematically translating developmental research into preventive intervention practice, we argue that it is important to understand the normative trajectories of central dimensions of social-emotional development (Malti et al. 2016; Noam et al. 2013), to systematically assess inter-individual differences prior to age-graded intervention delivery, as well as to monitor changes during program implementation. This is because levels of social-emotional capacities vary substantially *across* development and even between children of the *same* chronological age (Malti et al. 2016). Thus, in addition to distinct periods of development (e.g., early versus middle childhood),

it is important for preventive interventions to consider distinct levels of development within periods (e.g., within early childhood) in their theory and logic models.

Social-Emotional Assessment Tools

The systematic use of social-emotional assessment tools is an important step to ensure that developmental resilience factors are utilized in prevention practice. We have argued that early assessments that systematically integrate developmental research and social-emotional resilience factors are likely to facilitate educators' knowledge and ability to promote the delivery of preventive strategies that are sensitive to a child's developmental needs (Malti et al. 2016). In line with this argument, the importance of developing school-based early assessment tools for identifying social-emotional strengths has been underscored, and thus, some assessment tools for use in school and in after-school contexts have been developed (Durlak et al. 2011). Nevertheless, it has been more common to focus on tools that assess psychopathology, such as the Child Behavior Checklist (CBCL; Achenbach and Rescorla 2001). It has also been noted that even well-developed tools used to identify developmental psychopathology only show modest psychometric properties (Angold et al. 2012), which means that the probability that the children in need of care will be correctly identified as such is low (Costello 2016).

In the following section, we briefly review selected strength-based social-emotional assessment tools. We do not intend to be comprehensive in this brief summary. Rather, the aim is to identify existing psychometrically sound measures that are specifically designed to measure social-emotional strengths in school settings. This includes individual measures from various conceptual backgrounds from childhood to adolescence, population-level indicators (e.g., measures of development or health that are aggregated to a population level), and tools that are not program-specific and that are scientifically sound, which can be used for universal assessments, and that can guide program evaluation and be used to monitor change over time. We identified three commonly used school-based instruments that met these criteria. The first tool includes the Early Development and Middle School Instruments (EDI and MDI, respectively). These instruments are strengths-based but population-level measures. As such, the results are typically not used to evaluate individual developmental trajectories children and needs (Guhn and Goelman 2011). While these tools are of substantial importance for public health and policy planning, they provide little guidance for the planning, implementation, and evaluation of individual-, classroom-, and/or grade-based preventive approaches in school settings. Two other tools that met our criteria are the Social-Emotional Assets and Resiliency Scales (SEARS) and the Devereux Student Strengths Assessment (DESSA). The

SEARS aims at measuring "social resiliency" in 5- to 18-year-olds and includes items related to responsibility, self-regulation, social competence, and sympathy (Cohn et al. 2009).

The items across the three versions (i.e., parent- and teacher-report for 3rd to 6th graders and self-report for 7th to 12th graders) of the system have much in common but differ according to developmental level, setting for the rating, and context of the rating. The tool includes self-, teacher-, and parent-reports. As such, the tool is very useful to generate information about an individual child's, a group's, or a grade-wide group's level of social-emotional functioning, and as such, has much potential to inform prevention practice. However, its psychometric properties have only been reported to some extent (Cohn 2011). The DESSA is a 72-item measure designed for use with children in kindergarten through grade 8. It measures eight key social-emotional competencies, including self-awareness, social awareness, self-management, goal-directed behavior, relationship skills, personal responsibility, decision-making, and optimistic thinking (LeBuffe et al. 2009). It is completed by parents, teachers, or practitioners but does not include a self-report assessment. It generates individual and classroom profiles, and its psychometric properties have been reported (Nickerson and Fishman 2009). While both the SEARS and the DESSA focus on social-emotional strengths and include multiple informants, the DESSA does not include a self-report form for older children and adolescents. Yet, the use of self-reports in youth samples is important as it reveals information on their views on their own strengths. In addition, it is often the most (or only) feasible method to collect information in school and after-school contexts. Importantly, neither the SEARS nor the DESSA explicate how the selection of constructs was guided by theoretical considerations, and also do not present a developmental model of social-emotional skills and expected change. In sum, this brief review yielded relatively few strong assessment choices for schools to choose from when conducting prevention efforts. What are missing are tools based in sound social-emotional developmental theory with an empirically documented sensitivity to age-related change, designed not to be program-specific, have used various informants across ages (e.g., self, teachers, parents), and create individual, classroom, and school-based profiles for use in prevention planning, implementation, and evaluation.

Our social-emotional developmental approach and student self-report tool aimed at filling some of the gaps. The theoretical framework for our tool design, The Clover Model, was developed by Noam at The PEAR Institute at Harvard University and has been described elsewhere (Malti and Noam 2008, 2009; Noam et al. 2012). The Holistic Student Assessment (HSA) builds on the Clover Model. It includes seven dimensions of social-emotional functioning that reflect key aspects of self- and other-oriented skills. Four of those dimensions reflect the central self-oriented social-emotional skills of optimism, emotion control, action orientation, and self-reflection, while three dimensions reflect key other-oriented social-emotional skills, in particular, trust, empathy/sympathy, and assertiveness (Malti and Noam 2008, 2016). An age-appropriate development of self- and other-oriented skills is necessary to establish and maintain resilience and to be socially adapted. For instance, optimism is central to youth well-being and has been shown to be negatively associated with depression and internalizing symptomatology (e.g., Ruini et al. 2009). Vice versa, empathy is related to positive, prosocial outcomes (Eisenberg, 2000; Eisenberg et al. 2015). As such, they are important psychological processes that can be targeted in intervention programming. We have also argued that there is both stability and intra- and inter-individual variation in these skills (i.e., they are subject to change) reflecting developmental processes of growth, decline, and transformation.

HSA student data is collected at the beginning of the school year or program and is shared with teachers and practitioners at the individual and aggregate levels within a week of collection so that practitioners can gain immediate knowledge of their students' strengths and challenges. In addition to this rapid reporting cycle, developmentally appropriate interventions are recommended based on student results. Teachers, student support team (SST) leaders, and other key practitioners receive data interpretation training and developmentally focused coaching to better understand and act on the results. In its original version, the HSA was composed of both self-reported and teacher-reported rating scales designed to assess and guide prevention planning and evaluate outcomes related to the social-emotional strengths and challenges of middle-school students (Malti and Noam 2008; Noam et al. 2012). For this paper, only the student self-report version for grades 4–12 is relevant.

In the following section, we will introduce the HSA tool and provide further evidence for the psychometric properties of its self-report scale. For this purpose, we tested the psychometric properties of the HSA. We, first, ascertained the hypothesized 7-dimension factor structure of the scale in a "calibration sample"; second, we cross-validated the identified factor structure in a large "validation sample" and assessed the tool's measurement invariance across three age groups (late childhood, early adolescence, middle-late adolescence) to evaluate its developmental sensitivity to capture latent mean-level differences; third, we tested the construct validity of the HSA across the three age groups using the Strengths and Difficulties Questionnaire (SDQ; Goodman 1997), a widely used questionnaire for assessing children's and adolescents' externalizing and internalizing symptoms and prosocial behavior. Based on related developmental research, we expected social-emotional skills to be positively related to prosocial behavior (e.g., Flook et al. 2015). We also hypothesized negative links between social-emotional skills with externalizing and internalizing symptoms (Eisenberg et al. 2010; Eisner and Malti 2015).

In line with previous longitudinal research, we also expected developmental differences in the social-emotional skills. Specifically, we assumed an increase in emotion control and assertiveness and a decrease in action control and empathy (see Eisenberg et al. 2015; Eisenberg et al. 2010; Malti and Noam 2009; Malti et al. 2016), whereas we had less firm expectations regarding developmental change in trust, self-reflection, and optimism.

Method

Participants

Data were collected from a total sample of 5946 students (51% boys) attending grades 5–12 (M_{age} = 13.16 years, SD = 1.92) in the USA. The majority of the participating schools were located in Massachusetts (72%), followed by New York State (20%), Michigan (3%), Minnesota (2%), Maine (2%), and Rhode Island (1%). On average across schools, 5% of students opted out of completing the survey. To capitalize on the information included in our dataset, and to test the stability of our factor solution across two different sub-samples (Fabrigar et al. 1999), we extracted two groups of participants from the total sample. The first group of students completed the HSA only (n = 984). This group was used as the "calibration sample" (53% boys; M_{age} = 13.86 years, SD = 2.54). The second group of students completed both the HSA and the SDQ (n = 4962). This group was used as the "validation sample" (49% boys; M_{age} = 13.03 years, SD = 1.73). To test the robustness of the HSA factor across developmental phases, the validation sample was further divided into three age groups: (a) *late childhood* (M_{age} = 10.72, SD = 0.46, age ranged from 9 to 11 years of age; n = 1569), (b) *early adolescence* (M_{age} = 12.79, SD = 0.75, age ranged from 12 to 14 years of age; n = 2777), and (c) *middle-late adolescence* (M_{age} = 15.97, SD = 0.97, age ranged from 15 to 18 years of age; n = 616).

Measures

HSA The HSA assesses seven dimensions of social-emotional skills, namely *optimism, emotion control, action orientation, self-reflection, trust, empathy, and assertiveness* (see Table 1). For each of the seven hypothesized dimensions of the HSA, four to six items were generated based on a doctoral dissertation by Song (2003) and the Resilience Inventory developed by Noam and Goldstein (1998) and by subsequent item development at the Harvard based PEAR Institute. For the empathy subscale, some items were adapted from Eisenberg et al. 1996 (see Table 1). The wording of the items was slightly adapted to the different developmental periods to make them age-appropriate. This was done by pilot work, including focus groups with children, adolescents, and caregivers, who gave feedback

Table 1 Holistic student assessment: items of the self-report version

Self-reflection
1. I feel good when I fulfill my responsibilities
2. I like to learn about myself
3. Before I make decisions I think a lot from different sides
4. I try to understand the world I live in

Trust
5. I think most people are fair
6. I trust other people
7. Most people can be trusted

Optimism
8. I am happy with the choices I make in my life
9. More good things than bad things will happen to me
10. I think that I am a lucky person
11. I have more good times than bad times

Empathy
12. When I see another kid who is hurt or upset, I feel sorry for them*
13. I like to help people with their problems
14. When I see someone being bullied I feel sorry for them*
15. I feel bad for other kids who are sad or have problems*

Assertiveness
16. I stand up for things that matter to me
17. I defend myself against unfair rules
18. I speak my mind about my opinions
19. I say what I think even if adults or friends disagree

Action orientation
20. I like being physically active and moving my body
21. I like being active
22. Exercise is important to me

Emotion control (to be reverse coded)
23. I react to things so quickly I get in trouble
24. I get upset easily
25. When things do not go my way I quickly get really mad

* Items adapted from Eisenberg et al. (1996)

on the wording of the items. The initial pool consisted of 32 self-report items tapping into the hypothesized seven dimensions. Each item was formatted with a 4-point rating scale indicating the frequency of the behavior described: 0 (*not at all*), 1 (*sometimes*), 2 (*often*), and 3 (*almost always*). The psychometric properties of the preliminary version of the HSA were examined in the calibration sample (n = 984) via exploratory and confirmatory factor analyses (see "Results" section below).

Insert note below table with the following text:*Items adapted from Eisenberg et al. (1996) SDQ In order to assess the construct validity of the HSA, participants' externalizing and internalizing problems, as well as their prosocial behavior, were measured using the self-report version of the SDQ (Goodman 1997) in the validation group. The SDQ is a widely used screening tool for children and adolescents to evaluate

their prosocial behavior, as well as externalizing and internalizing problems. Although the SDQ was originally developed to assess five dimensions of psychological functioning (i.e., prosocial behavior, emotional symptoms, peer problems, conduct problems, and hyperactivity/inattention), a recent psychometric analysis of the SDQ suggested the usefulness and increased validity of considering two broader dimensions of externalizing (conduct problems and hyperactivity/inattention) and internalizing disorders (emotional symptoms and peer problems), in addition to prosocial behavior (Caci et al. 2015). Omega reliability coefficients for the three dimensions were high in each age group (i.e., late childhood, early adolescence, and middle-late adolescence): prosocial behavior (ωs = .81, .80, and .83, respectively); externalizing symptoms (ωs = .86, .84, and .84, respectively); internalizing symptoms (ωs = .81, .83, and .86, respectively).

Procedure

The self-report forms of the HSA and SDQ were distributed in school or after-school settings. Instruments were filled out by students in group settings with 8 to 12 students and under careful adult supervision. Administration took approximately 20 min.

Results

Factor Structure of the HSA (Calibration Sample)

Exploratory Factor Analysis A series of exploratory factor analyses (EFAs) were conducted to (a) ascertain the goodness of the hypothesized 7-factor solution of the HSA against possible alternative models and (b) select the items with the best psychometric properties. High primary standardized factor loadings (λ) were defined as above .40 (Schaefer et al. 2015), and cross-loadings were defined as having a value $\geq$.32 (Tabachnick and Fidell 2013) or a small gap between the primary and secondary loading (i.e., .20; see Schaefer et al. 2015). Given the ordered categorical nature of our items, we ran our EFAs using the weighted least-squares mean and variance-adjusted (WLSMV) estimation method in M*plus* 7.31 (Muthén and Muthén 2012), which is recommended for categorical items with a limited number of response categories (Beauducel and Herzberg 2006). The WLSMV estimator requires the corrected χ^2 difference test for nested models ($\Delta\chi^2$) as fit-index (Muthén and Muthén 2012). Because the χ^2 statistic is also sensitive to sample size, we also used the comparative-fit-index (CFI), Tucker-Lewis-Index (TLI) > .90, and root-mean-square-error-of-approximation (RMSEA) values <.08 with 90% confidence interval (CI) as indicators of acceptable model fit. Given our large sample size, we set a restricted level of statistical significance at $p \leq .01$.

To evaluate if the hypothesized 7-factor structure was appropriate, we ran three preliminary EFAs, in which we extracted 6, 7, and 8 factors using Geomin as the oblique method of rotation (Muthén and Muthén 2012). We, then, compared their fit and examined the substantive interpretability of their factor solution to determine the number of factors to be retained. The 7-factor solution χ^2 (293) = 598.391, $p < .001$; CFI = .984; TLI = .972, RMSEA = .033 [90% CI: .029, .036] showed a better fit than the 6-factor solution (i.e., $\Delta\chi^2$ (26) = 147.18, $p < .001$). Although the 8-factor solution χ^2 (268) = 493.732, $p < .001$; CFI = .988; TLI = .977, RMSEA = .029 [90% CI: .025, .033] showed a further improvement of the fit compared to the 7-factor solution (i.e., $\Delta\chi^2$ (25) = 107.845, $p < .001$), the solution was difficult to interpret because the seventh factor (only one item had $\lambda > .40$) and eighth factor (only two items showed $\lambda s > .40$) were identified by fewer than three items. Hence, we retained the 7-factor solution as the best fitting model. Next, we deleted those items showing low primary loadings and/or high cross-loadings, and the 7-factor EFA was repeated. After the deletion of seven items, the remaining 25 items loaded strongly only onto their respective intended factor (see Table 1).

The first factor, labeled *self-reflection*, taps into students' self-awareness and was defined by four items (primary λs ranged from .51 to .65; secondary λs ranged from −.09 to .17). The second factor, labeled *trust*, reflects students' tendency to trust other people and was defined by three items (primary λs ranged from .64 to .69; secondary λs ranged from −.07 to .21). The third factor, labeled *optimism*, captures students' perception of their life under a positive outlook and was defined by four items (primary λs ranged from .50 to .87; secondary λs ranged from −.10 to .21). The fourth factor, labeled *empathy*, reflects students' concern towards others in need and was defined by four items (primary λs ranged from .52 to .77; secondary λs ranged from −.11 to .28). One item (i.e., when I see another kid who is hurt or upset, I feel sorry for them) was taken from Eisenberg et al. (1996). Given the conceptual and measurement overlap between the constructs of "empathy" and "sympathy," we chose to use the term "empathy" because it is commonly used to describe both pure empathy and empathy-related responses in children and adolescents. The fifth factor, labeled *assertiveness*, captures students' tendency to affirm and defend their point of view and was defined by four items (primary λs ranged from .49 to .70; secondary λs ranged from −.09 to .23). The sixth factor, labeled *action orientation*, taps into students' preference for physical activities and was defined by three items (primary λs ranged from .63 to .91; secondary λs ranged from −.09 to .22). The seventh factor, labeled *emotion control*, captures students' ability to manage negative emotions and was defined by three items (primary λs ranged from .62 to .84; secondary λs ranged from −.13 to .20).[1] The factor correlation matrix (Table 2) indicated that self-reflection, trust, optimism, empathy, assertiveness, emotion control, and action

---

[1] Items for this subscale were reversely coded (see Table 1).

Table 2 Latent correlation matrix and latent mean-level differences across age groups

	1	2	3	4	5	6	7	Late childhood[1] $M_{standardized}$	Early adolescence $M_{standardized}$	Middle-late adolescence $M_{standardized}$
1. Self-reflection	—							0.000	0.000	0.000
2. Trust	.33[*]	—						0.000	−0.319[a]	−0.654[b]
3. Optimism	.56[*]	.44[*]	—					0.000	0.000	0.000
4. Empathy	.51[*]	.44[*]	.30[*]	—				0.000	−0.317[c]	−0.283[c]
5. Assertiveness	.41[*]	.33[*]	.45[*]	.19[*]	—			0.000	0.000	0.395[d]
6. Action orientation	.43[*]	.11[*]	.35[*]	.20[*]	.31[*]	—		0.000	−0.254[e]	−0.236[e]
7. Emotion control	.09[*]	.15[*]	.22[*]	.10[*]	.07	−.10[*]	—	0.000	−0.243[f]	−0.225[f]

Note. $N = 984$ for the correlations. [*]$p < .01$

[1] Latent means were fixed to zero in the late childhood group (i.e., the reference group) for identification purposes. Late childhood: $n = 1569$; Early adolescence: $n = 2777$; Middle-late adolescence: $n = 616$.

Differences in latent means across groups are expressed in standard deviation units from the reference group. For each latent dimension (reported per row), values with different letters were significantly different across groups. All the latent mean-level differences were significant at $p < .001$

orientation were overall positively and significantly correlated at $p \leq .01$.

Confirmatory Factor Analysis and Reliability Next, we tested the robustness of the identified 7-factor structure via a confirmatory factor analysis (CFA) with WLSMV estimation and theta parametrization (Muthén and Muthén 2012). The 7-factor structure fits the data well, $\chi^2 (254) = 805.102, p < .001$; CFI = .962; TLI = .955, RMSEA = .047 [90% CI: .043, .051]. All items showed high standardized factor loadings (λs ranged from .52 to .94). Omega (ω) coefficients were computed to assess scale score reliability (McDonald 1970).[2] All scales showed good reliability coefficients: self-reflection ($\omega = .77$), trust ($\omega = .77$), optimism ($\omega = .84$), empathy ($\omega = .87$), assertiveness ($\omega = .76$), action orientation ($\omega = .91$), and emotion control ($\omega = .77$).

Cross-Validation and Measurement Invariance Across Age Groups (Validation Sample)

The 7-factor CFA solution showed an acceptable fit in the validation sample ($n = 4962$), $\chi^2 (254) = 3326.652, p < .001$; CFI = .956; TLI = .948, RMSEA = .049 [90% CI: .048, .051], as well as across age groups: late childhood, $\chi^2 (254) = 1051.453, p < .001$; CFI = .959; TLI = .951, RMSEA = .045 [90% CI: .042, .048]; early adolescence $\chi^2 (254) = 2137.373, p < .001$; CFI = .951; TLI = .942, RMSEA = .052 [90% CI: .050, .054]; and middle-late adolescence $\chi^2 (254) = 714.843, p < .001$; CFI = .955; TLI = .947, RMSEA = .054 [90% CI: .050, .059]. Reliability coefficients for the respective scales were good across the three age groups: self-reflection ($\omega s = .71, .74, .73$); trust ($\omega s = .74, .75, .80$); optimism

($\omega s = .81, .82, .82$); empathy ($\omega s = .88, .87, .91$); assertiveness ($\omega s = .70, .74, .83$); action orientation ($\omega s = .86, .89, .91$); and emotion control ($\omega s = .77, .79, .85$).

In order to test developmental differences at the latent level, we first ascertained that the HSA scores were comparable across the three age groups by establishing its measurement invariance (MI) in a multiple-group analytic framework (Millsap 2011). We tested three nested models of MI adapted for ordered categorical indicators. In each of these models, we imposed increasingly restrictive constraints on the factor loadings (λ) and thresholds (ϑ) of the items composing each scale.[3] The number of ϑs for each item equals the number of k response categories minus one (i.e., three ϑs for our 4-point rating scale; see Muthén and Muthén 2012). Following the recommendations of Millsap and Yun-Tein (2004), we first tested a *configural* invariance model in which both λs and ϑs were freely estimated across the three age groups except for some parameters constrained for identification purposes (i.e., two ϑs for the marker item of each scale and one ϑ for the non-marker items). Next, we tested *metric* invariance in which the λ_s of the items were constrained to be equal across groups. When metric MI holds, the size of the loadings is the same across groups and the instrument can thus be assumed to rank the participants in the same way across each group (Vandenberg and Lance 2000). Lastly, we tested *scalar* (or strong) invariance by imposing both λs and ϑs to be equal across the three age groups. If scalar invariance holds, scores across groups share the same unit of measurement and origin, thereby allowing meaningful comparison of their latent means (Meredith 1993). To test differences among configural, metric, and scalar MI, we calculated $\Delta\chi^2$ tests of these nested models

---

[2] Compared with classical alpha reliability estimates, ω has the advantage of taking into account both the strength of relation between items and constructs (λs) and measurement error, while relaxing the assumption that the items are tau-equivalent.

[3] With ordered-categorical data, the threshold parameter represents the expected value on the underlying latent variable, which indicates the shift from one response category (e.g., 0 = *not at all*) to another one (e.g., 1 = *sometimes* or higher; see Muthén and Muthén 2012).

using the DIFFTEST function in M*plus* 7.31 (Muthén and Muthén 2012) with a restricted level of statistical significance at $p \leq .01$.[4] However, because the $\Delta\chi^2$ test is overly sensitive to sample size and minor model misspecifications, we also considered the ΔCFI test with a restricted critical level of .005 (Chen 2007). This was done in line with Chen's (2007) guidelines for MI across groups differing in their sample size, as observed decrease in fit can be attributed to sampling error rather than a lack of equivalence when ΔCFI ≤ 0.005.

The configural invariance model fits the data relatively well, χ^2 (762) = 3807.413, $p < .001$; CFI = .954; TLI = .945, RMSEA = .049 [90% CI: .048, .051]. The further constraints of the λs in the metric invariance model χ^2 (799) = 3676.998, $p < .001$; CFI = .956; TLI = .951, RMSEA = .047 [90% CI: .045, .048] did not worsen the model fit (i.e., $\Delta\chi^2$ (37) = 37.628, $p = .44$; ΔCFI = + .002), thereby attesting the plausibility of these constraints. The hypotheses of equal ϑs in the scalar invariance model hold, χ^2 (886) = 3862.678, $p < .001$; CFI = .955; TLI = .954, RMSEA = .045 [90% CI: .044, .047], as suggested by the very small decrease in CFI between the two models (i.e., $\Delta\chi^2$ (87) = 236.439, $p < .001$; ΔCFI = −.001).

To evaluate latent mean-level differences, we assessed fit differences between (1) the *scalar invariance model* in which the latent means were fixed to be zero in one group (i.e., late childhood) and freely estimated in the other two groups and (2) the *latent means invariance model* in which all the means were fixed to equality (i.e., zero) across the three groups. Although the *latent means invariance model* showed a moderate fit to the data, χ^2 (900) = 5124.845, $p < .001$; CFI = .936; TLI = .936, RMSEA = .053 [90% CI: .052, .055], it worsened the *scalar invariance model* fit, thereby indicating the presence of significant mean-level differences across the three age groups (i.e., $\Delta\chi^2$ (14) = 441.903, $p < .001$; ΔCFI = − .019). Next, we constrained the latent means to equality one at a time. The *partial latent means invariance* model, χ^2 (894) = 4160.834, $p < .001$; CFI = .950; TLI = .950, RMSEA = .047 [90% CI: .046, .048], in which only the means of self-reflection and optimism were invariant across the three age groups, did not differ from the scalar invariance model (i.e., $\Delta\chi^2$ (8) = 139.307, $p < .001$; ΔCFI = −.005).

Participants in the early and middle-late adolescence age groups showed lower latent levels of emotion control and lower levels of empathy and action orientation compared to participants in the late childhood age group (Table 2). Trust declined from late childhood to middle-late adolescence, whereas assertiveness showed an increase during middle-late adolescence. Finally, the three age groups did not differ on self-reflection and optimism.

Construct Validity of the HSA Across Age Groups (Validation Sample)

We assessed the construct validity of the HSA by examining the relations between the participants' scores on the seven HSA dimensions with their SDQ scores. We implemented a structural equation model (SEM) within a multiple framework per age group. WLSMV was used as a method of parameter estimation. Specifically, we tested the consistency and strength of our validity model across late childhood (9–11-year-olds, $n = 1569$), early adolescence (12–14-year-olds, $n = 2777$), and middle-late adolescence (15–18-year-olds, $n = 616$) by comparing the fit of two nested models: (1) an *unconstrained SEM* in which the unstandardized effects of the seven HSA dimensions on prosocial behavior, externalizing, and internalizing problems were freely estimated across age groups and (2) a *constrained SEM* in which the same effects were constrained to equality across the three age-groups. Gender (0 = girls; 1 = boys) was included as a control variable. In line with the previous analyses, we set a restricted level of statistical significance at $p \leq .01$ and considered a cutoff of decrease in ΔCFI ≤ 0.005 to accept the plausibility of the constraints imposed (Chen 2007). The SEM validity model is graphically depicted in Fig. 1.

Both the *unconstrained SEM*, χ^2 (1110) = 4808.031, $p < .001$; CFI = .948; TLI = .942, RMSEA = .045 [90% CI: .044, .046] and *constrained SEM*, χ^2 (1173) = 4482.457, $p < .001$; CFI = .953; TLI = .951, RMSEA = .041 [90% CI: .040, .043] showed an acceptable fit to the data. Importantly, the *constrained SEM*, in which the validity parameters were fixed to equality across age groups, was not statistically different from the *unconstrained SEM* (i.e., $\Delta\chi^2$ (63) = 85.917, $p = .029$; ΔCFI = +.005), thereby attesting to the invariance of these effects across late childhood, early adolescence, and middle-late adolescence. The majority of the effects were in the hypothesized directions and therefore confirmed the construct validity of the HSA in each age group. As expected, prosocial behavior was positively associated with self-reflection, trust, empathy, and emotion control (Table 3). Externalizing problems were negatively related to emotion control and reflection. Finally, internalizing problems were negatively associated with optimism, assertiveness, action orientation, and emotion control, whereas positively with self-reflection and empathy.

Discussion

The role of social-emotional development in children's and adolescents' academic functioning and mental health has been widely acknowledged. Yet, there is an urgent need to generate more in-depth understanding of each child's social-emotional development prior to developmentally tailored prevention delivery. Educators also need to understand their classrooms and their schools, as well as their overall school districts.

---

[4] Given that χ^2 values are not exact using WLSMV as method of parameter estimation, χ^2 and resulting CFI values can be non-monotonic with model complexity (Muthén and Muthén 2012).
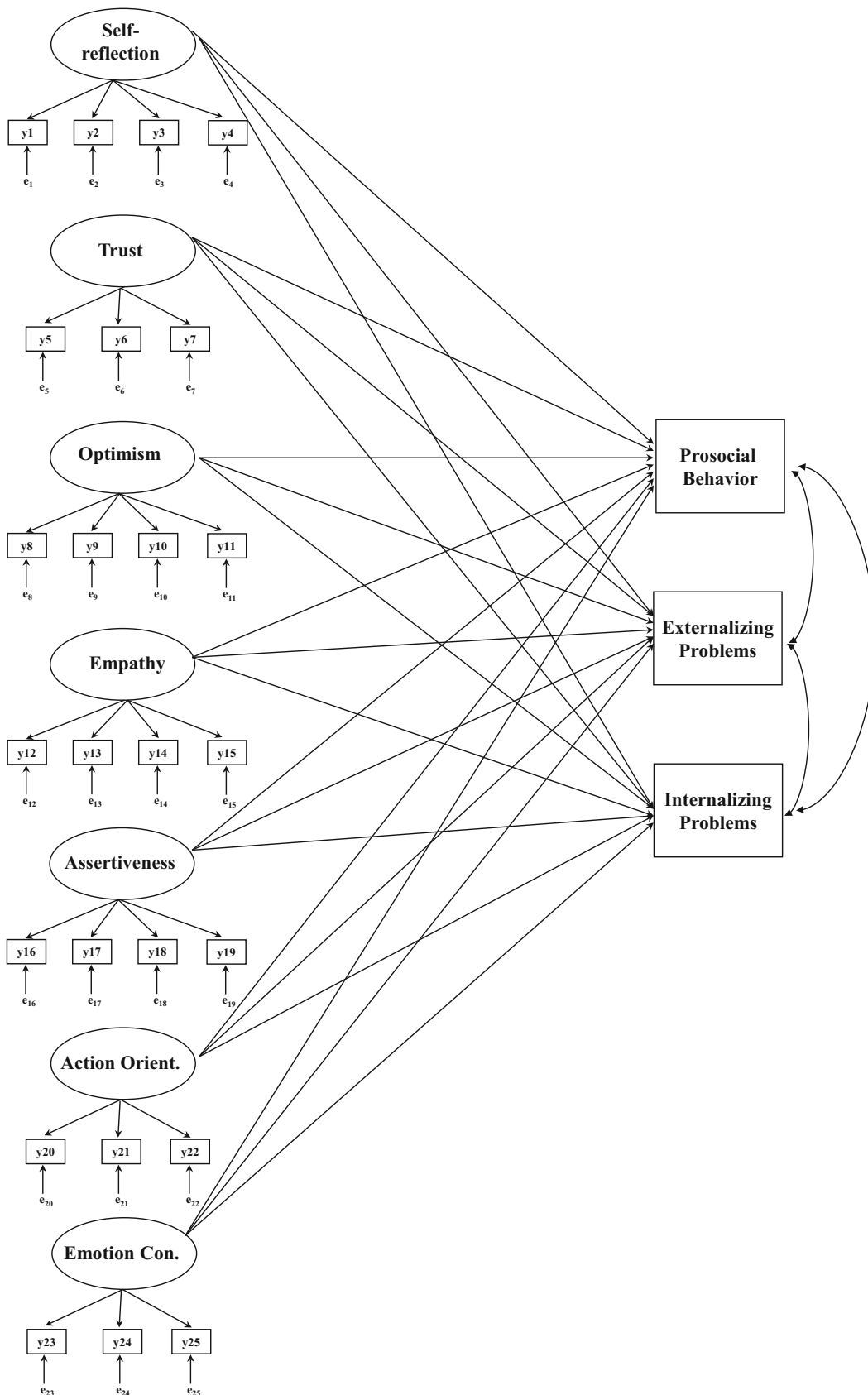
Fig. 1 SEM validity model. *Note*. The SEM validity model was implemented simultaneously across the three age groups (in a multiple-group framework). The effects of sex and correlations among HSA latent factors were estimated but not depicted for simplicity

Table 3 Standardized regression coefficients of the validity analysis

Gender → HSA and outcome dimensions	Late childhood	Early adolescence	Middle-late adolescence
Self-reflection	−.05[*]	−.05[*]	−.05[*]
Trust	.08[**]	.08[**]	.07[**]
Optimism	.06[**]	.06[**]	.06[**]
Empathy	−.23[**]	−.23[**]	−.22[**]
Assertiveness	−.03	−.03	−.03
Action orientation	.21[**]	.23[**]	.22[**]
Emotion control	.06[*]	.07[**]	.07[**]
Prosocial behavior	−.03	−.03	−.03
Externalizing problems	.07[**]	.08[**]	.08[**]
Internalizing problems	−.02	−.02	−.02
HSA dimensions → prosocial behavior	($R^2 = .47$)	($R^2 = .47$)	($R^2 = .43$)
Self-reflection	.30[**]	.32[**]	.28[**]
Trust	.07[*]	.07[*]	.08[*]
Optimism	−.08[†]	−.09[†]	−.08[†]
Empathy	.42[**]	.41[**]	.43[**]
Assertiveness	.01	.01	.01
Action orientation	.01	.01	.01
Emotion control	.14[**]	.10[**]	.10[**]
HSA dimensions → externalizing problems	($R^2 = .62$)	($R^2 = .60$)	($R^2 = .58$)
Self-reflection	−.25[**]	−.28[**]	−.25[**]
Trust	.01	.02	.02
Optimism	−.05	−.05	−.05
Empathy	−.01	−.01	−.01
Assertiveness	.03	.03	.03
Action orientation	.03	.03	.03
Emotion control	−.69[**]	−.67[**]	−.69[**]
HSA dimensions → internalizing problems	($R^2 = .34$)	($R^2 = .34$)	($R^2 = .38$)
Self-reflection	.33[**]	.36[**]	.32[**]
Trust	−.05	−.05	−.06
Optimism	−.33[**]	−.36[**]	−.36[**]
Empathy	.15[**]	.14[**]	.15[**]
Assertiveness	−.27[**]	−.27[**]	−.29[**]
Action orientation	−.12[**]	−.11[**]	−.11[**]
Emotion control	−.47[**]	−.44[**]	−.45[**]

Note. Standardized regression coefficients were reported for late childhood ($n = 1569$), early adolescence ($n = 2777$), and middle-late adolescence ($n = 616$). Standardized regression coefficients can slightly differ across age groups since unstandardized parameters were constrained to equality within the multiple-group analysis. Gender (0 = girls; 1 = boys). †$p < .05$; *$p < .01$; **$p < .001$

Embracing a strength-based developmental focus is important because it can provide support in a developmentally appropriate way and thus contribute to improved intervention effectiveness. Here, we have argued that one powerful way of identifying children's and adolescents' social-emotional strengths is to use tools to screen and assess central dimensions of self-oriented and other-oriented social-emotional functioning. Measuring social-emotional development is important because those capacities not only affect academic functioning and mental health but also tend to vary

substantially both across development and between children of the *same* chronological age (Malti et al. 2016). Based on this argument, the purpose of this article was to discuss the role of theory-guided tools in the planning, implementation, and evaluation of school-based preventive practices aimed at promoting social-emotional development and mental health in all children and adolescents. Social-emotional learning programs, albeit having proven to be effective in practice, still incorporate developmental considerations only in part, including the absence of the systematic use of tools to generate

knowledge on children's social-emotional devleopment that can inform the intervention planning process. Guided by this logic, we first reviewed selected existing tools and identified the current gaps in the literature. We found that only few validated social-emotional assessment instruments exist that have been developed and validated for use in school contexts thus far. Two assessment tools are the SEARS and the DESSA which include age-graded items. However, these tools are not based in a strong conceptual model, and the selection of the core constructs remains to some extent arbitrary. It also remains unclear how age-related change is expected to happen, and SEARS' psychometric properties have only been validated to some extent. In contrast, the DESSA has documented its validity but does not include self-report scales.

Next, we evaluated the psychometric properties of the HSA, a tool that was meant to fill some of the previous gaps in the assessment literature. Using a calibration sample, we first confirmed the 7-factor structure of the tool and cross-validated the structure in a large validation sample. The findings confirmed that the HSA's factor structure is robust. Our empirical analysis also provided evidence for measurement invariance across age groups using a multiple-group analytic framework. The findings confirmed that latent mean-level differences can be attributed to actual developmental differences rather than alterations and/or differential use of the instruments across the three age groups (Millsap 2011). Lastly, we documented the construct validity of the HSA by exploring expected relations between its dimensions with externalizing and internalizing psychopathology, as well as with adaptive, prosocial behavior. While prosocial behavior was positively associated with self-reflection, trust, empathy, and emotion control, externalizing problems were negatively related to emotion control and reflection. In addition, internalizing problems were negatively associated with optimism, assertiveness, action orientation, and emotion control, and they were positively related to self-reflection and empathy. In sum, the majority of the effects were in the expected directions, confirming the construct validity of the HSA across age.

Collectively, the findings provide evidence for the factor structure, measurement invariance, developmental sensitivity, and construct validity of the HSA. The results of the HSA can be used to generate individual, classroom, and school-wide profiles of self- and other-oriented social-emotional skills. Such profiles are not only important to generate the knowledge on a child's, classroom's, and/or school's level of social-emotional functioning that is needed to help practitioners and policy-makers identify similarities and differences in self- and other-oriented social-emotional skills across ages, cultural groups, and communities (for a sample profile, see Malti and Noam 2016). Rather, it is also useful for the selection of developmentally tailored preventive intervention strategies as they can increase fit between a child's needs and developmental level with type of intervention and pedagogical strategy

that is being implemented in a particular context (for a more detailed discussion, see Malti and Noam 2016).

The findings also documented expected change in HSA dimensions of social-emotional development across ages. More specifically, we found that empathy, trust, and action orientation decreased with age, whereas assertiveness increased, and no age-related differences in optimism and self-reflection were found. Taken together, these findings support previous reseach and are overall in line with current theorizing regarding the development of social-emotional skills (Malti and Noam 2016; Noam et al. 2013). The findings that trust and emotion control decreased with age were somewhat unexpected. Possibly, the neural changes that characterize adolescence may explain the decrease in emotion control, as it has been shown that the various neural changes in the adolescent brain may lead to heightened emotional responsivity compared to childhood and adulthood (Casey et al. 2008). In addition, adolescents might become more aware of the multifaceted nature of peer interactions and realize that they and others often have to balance issues of fairness with self-interest, conventions, and a multitude of other concerns (Killen and Malti 2015). This increasing awareness may be associated with a decrease in trust in others.

Our empirical approach had several strengths: first, we tested the psychometric properties of the self-report scale of the HSA in a large sample of students. Second, we were able to establish the measurement invariance of the HSA across a broad age range (i.e., from late childhood to late adolescence) which allowed us to meaningfully interpret HSA mean scores as actual developmental differences. Third, we documented the HSA's construct validity within a multiple group SEM framework which permitted us to test simultaneously the strength of the relations between HSA dimensions and SDQ scores across the three age groups. Despite these strengths, we acknowledge several limitations. First, we only evaluated the self-report version of the HSA, and the findings between the HSA and the SDQ might be inflated due to shared informant variance. Future investigations with multiple informants (e.g., self-report, teacher- report, parent-report) might further strengthen the validity of the tool. In addition, despite the large-scale nature of our sample, it was cross-sectional, and future longitudinal designs are warranted to test the tool's developmental sensitivity further, both in terms of intraindividual change sensitivity and interindividual differences at different developmental periods. Given the character of the data collection across various school and after-school sites in contexts of research and practice, we were unable to collect complete informaiton on participation rates which limits the generalizibility of findings. Finally, we also recognize that the three age groups were unbalanced in terms of subjects. Although our partitioning of the sample was guided by theoretical arguments (i.e.,

developmental stages) and we were able to establish the measurement invariance of the HSA across late childhood, early adolescence, and middle-late adolescence, future studies should replicate our factor structure using more balanced age groups.

In sum, this article presented evidence for the psychometric properties and developmental sensitivity of the self-report version of the HSA in a large sample of US middle school students, an assessment tool grounded in social-emotional developmental theory and research. It is clear that theory-driven developmental measures that focus on socioemotioanl strengths and are sensitive to change can be more effective in engaging students in high-quality in-school and out-of-school-time activities that fit their developmental strengths and needs (Malti et al. 2016). In addition, we have shown elsewhere that the scale scores of the HSA generate individual-, classroom-, and school-wide profiles of students' social-emotional development (Malti and Noam 2008, 2016; Noam et al. 2012). This can guide the planning, selection, and implementation of developmentally sensitive preventive interventions. For example, information on baseline differences in social-emotional skills in a classroom can result in intervention strategies being more effective with certain children and classrooms than others (Malti and Noam 2009). Some children, regardless of age, may show relatively high levels of baseline social-emotional functioning, while others may show low levels. This could have an immediate effect on program outcomes if less differentiated children lack the social-emotional capacity to comprehend and implement the skills being taught in the classroom. Therefore, simply adjusting existing programs for lower or higher age groups may not suffice. Instead, program content and delivery should be further modified based on an understanding of social-emotional developmental theory and information collected by assessment tools, which includes intricate developmental differences within and between age groups. Future work will need to assess these targeted interventions and their effect on social-emotional development. In sum, the HSA profiles can help practitioners to understand the status of social-emotional development, inform on inter-individual differences prior to age-graded intervention delivery, and monitor changes during program implementation and outcome evaluation of preventive strategies, policies, and practices to promote social-emotional development and reduce mental health risks.

References

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms and profiles*. Burlington: University of Vermont, Research Center for Children, Youth, and Families.

Angold, A. E. J., Erkanli, A., Copeland, W., Goodman, R., Fisher, P. W., & Costello, E. J. (2012). Psychiatric diagnostic interviews for children and adolescents: A comparative study. *Journal of the American Academy of Child & Adolescent Psychiatry, 51*, 506–517. doi:10.1016/j.jaac.2012.02.020.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling, 13*, 186–203. doi:10.1207/s15328007sem1302_2.

Caci, H., Morin, A. J. S., & Tran, A. (2015). Investigation of a bifactor model of the strengths and Difficulties questionnaire. *European Child & Adolescent Psychiatry, 24*, 1291–1301. doi:10.1007/s00787-015-0679-3.

Casey, B. J., Jones, R. M., & Hare, T. A. (2008). The adolescent brain. *Annals of the New York Academy of Sciences, 1124*, 111–126. doi:10.1196/annals.1440.010.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*, 464–504. doi:10.1080/10705510701301834.

Cohn, B. (2011). *Innovations in strength-based social-emotional assessment: Factor analysis, psychometric analysis, and cross-informant comparisons with the SEARS-T* (Unpublished doctoral dissertation). University of Oregon, Eugene.

Cohn, B., Merrell, K. W., Felver-Grant, J., Tom, K., & Endrulat, N. (2009, February). *Strength-based assessment of social and emotional functioning: SEARS-C and SEARS-A*. Boston: Presented at the Meeting of the National Association of School Psychologists.

Costello, J. (2016). Early detection and prevention of mental health problems: Developmental epidemiology and systems of support. *Journal of Clinical Child & Adolescent Psychology, 45*, 710–717. doi:10.1080/15374416.2016.1236728.

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*, 405–432. doi:10.1111/j.1467-8624.2010.01564.x.

Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annual Review of Psychology, 51*, 665–697. doi:10.1146/annurev.psych.51.1.665.

Eisenberg, N., Fabes, R. A., Murphy, B., Karbon, M., Smith, M., & Maszk, P. (1996). The relations of children's dispositional empathy-related responding to their emotionality, regulation, and social functioning. *Developmental Psychology, 32*, 195–209.

Eisenberg, N., Spinrad, T. L., & Eggum, N. D. (2010). Emotion-related self-regulation and its relation to children's maladjustment. *Annual Review of Clinical Psychology, 6*, 495–525. doi:10.1146/annurev.clinpsy.121208.131208.

Eisenberg, N., Spinrad, T. L., & Knafo-Noam, A. (2015). Prosocial development. In M. E. Lamb & R. M. Lerner (Eds.), *Handbook of hild psychology and developmental science, Vol. 3: Social, emotional and personality development* (7th ed., pp. 610–656). New York: Wiley.

Eisner, M. P., & Malti, T. (2015). Aggressive and violent behavior. In M. E. Lamb & R. M. Lerner (Eds.), *Handbook of child psychology and developmental science, Vol. 3: Social, emotional and personality development* (7th ed., pp. 795–884). New York: Wiley.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272–299. doi:10.1037/1082-989X.4.3.272.

Flook, L., Goldberg, S. B., Pinger, L., & Davidson, R. J. (2015). Promoting prosocial behavior and self-regulatory skills in preschool children through a mindfulness-based kindness curriculum. *Developmental Psychology, 51*, 44–51. doi:10.1037/a0038256.

Goodman, R. (1997). The strengths and Difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38*, 581–586. doi:10.1111/j.1469-7610.1997.tb01545.x.

Greenberg, M. T., Weissberg, R. P., O'Brien, M. U., Zins, J. E., Fredericks, L., Resnik, H., & Elias, M. J. (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *American Psychologist, 58*, 466–474. doi:10.1037/0003-066X.58.6-7.466.

Guhn, M., & Goelman, H. (2011). Bioecological theory, early child development and the validation of the population-level early development instrument. *Social Indicators Research, 103*, 193–217. doi:10.1007/s11205-011-9842-5.

Killen, M., & Malti, T. (2015). Moral judgments and emotions in contexts of peer exclusion and victimization. *Advances in Child Development and Behavior, 48*, 249–276. doi:10.1016/bs.acdb.2014.11.007.

LeBuffe, P. A., Shapiro, V. B., & Naglieri, J. A. (2009). *The Devereux student strengths assessment (DESSA)*. Lewisville: Kaplan Press.

Malti, T., & Noam, G. G. (2008). Where youth development meets mental health and education: The RALLY approach. *New Directions for Youth Development*, *No. 120*.

Malti, T., & Noam, G. G. (2009). A developmental approach to the prevention of adolescents' aggressive behavior and the promotion of resilience. *European Journal of Developmental Science, 3*, 235–246. doi:10.3233/DEV-2009-3303.

Malti, T., & Noam, G. G. (2016). Social-emotional development: From theory to practice. *European Journal of Developmental Psychology, 13*, 652–665. doi:10.1080/17405629.2016.1196178.

Malti, T., Chaparro, M. P., Zuffianò, A., & Colasante, T. (2016). School-based interventions to promote empathy-related responding in children and adolescents: A developmental analysis. *Journal of Clinical Child and Adolescent Psychology*, 1–14. doi:10.1080/15374416.2015.1121822.

McDonald, R. P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology, 23*, 1–21. doi:10.1111/j.2044-8317.1970.tb00432.x.

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika, 58*, 525–543. doi:10.1007/BF02294825.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: Routledge.

Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research, 39*, 479–515. doi:10.1207/S15327906MBR3903_4.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus User's Guide* (Seventh ed.). Los Angeles: Muthén & Muthén.

Nickerson, A. B., & Fishman, C. (2009). Convergent, divergent, and predictive validity of the Devereux elementary student strengths assessment. *School Psychology Quarterly, 24*, 48–59. doi:10.1037/a0015147.

Noam, G. G., & Goldstein, L. S. (1998). *The resilience inventory*. Unpublished research report, Harvard Medical School and McLean Hospital.

Noam, G. G., Malti, T., & Guhn, M. (2012). From clinical-developmental theory to assessment: The holistic student assessment tool. *International Journal of Conflict and Violence, 6*, 201–213.

Noam, G. G., Malti, T., & Karcher, M. J. (2013). Mentoring relationships in developmental perspective. In D. DuBois & M. J. Karcher (Eds.), *Handbook of youth mentoring* (pp. 99–116). Thousand Oaks: Sage.

Payton, J. W., Weissberg, R. P., Durlak, J. A., Dymnicki, A. B., Taylor, R. D., Schellinger, K. B., & Pachan, M. (2008). *Positive impact of social and emotional learning for kindergarten to eighth-grade students: Findings from three scientific reviews (report)*. Chicago: Collaborative for Academic, Social, and Emotional Learning.

Ruini, C., Ottolini, F., Tomba, E., Belaise, C., Albieri, E., Visani, D., Offidani, E., Caffo, E., & Fava, G. A. (2009). School intervention for promoting psychological well-being in adolescence. *Journal of Behavior Therapy and Experimental Psychiatry, 40*, 522–532.

Schaefer, L. M., Burke, N. L., Thompson, J. K., Dedrick, R. F., Heinberg, L. J., Calogero, R. M.,, .. Svami, V. (2015). Development and validation of the Sociocultural Attitudes Towards Appearance Questionnaire-4 (SATAQ-4). *Psychological Assessment*, 27(1), 54–67. doi:10.1037/a0037917.

Song, M. (2003). *Two studies on the Resilience Inventory (RI): Toward the goal of creating a culturally sensitive measure of adolescence resilience*. Unpublished doctoral dissertation, Harvard University, Cambridge.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston: Pearson.

Vandenberg, R., & Lance, C. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4–70. doi: 10.1177/109442810031002 .