Assessing Social-Emotional Development:

Reliability and Validity of The Social-Emotional Responding Task (SERT)

Tina Malti [1, 2], Linlin Zhang[3], and H. Melis Yavuz[1, 2]

[1] Department of Psychology, University of Toronto

[2] Centre for Child Development, Mental Health, and Policy, University of Toronto Mississauga

3   School of Psychology, Capital Normal University, Beijing

Author Note

Abstract

The Social-Emotional Responding Task (SERT) assesses children's anticipated emotions in the contexts of transgressions (Malti, 2017). We present a systematic psychometric evaluation of the SERT using data from two different samples of 4- to 8-year-old children from Canada ($N = 291$, $M_{age} = 6.55$ years, SDs = 2.02, 50% boys in sample 1 and $N = 282$, $M_{age} = 6.57$ years, SDs = 1.56, 49% boys in sample 2). Children reported their anticipated emotions in six vignettes describing three domains of transgressions (aggressive acts, prosocial omission, and social exclusion). Caregivers rated children's sympathy and prosocial and aggressive behaviors. Results supported a one-factor ("domain general") model over a three-factor ("domain-specific") model, indicating convergence of anticipated emotions across vignettes and domains. Measurement invariance was established across gender and age groups, indicating robustness of the assessment. Construct validity was supported by associations with sympathy and less robustly, with prosocial and aggressive behavior. We discuss the use of SERT as an assessment tool for children's social-emotional capacities in research and practice settings.

Keywords: social-emotional development, developmental assessment, prosocial development, social emotions, childhood

Public Significance Statement

We present a novel tool to provide information on children's social-emotional capacities and provide evidence for its rigor. The focus on children's social-emotional strengths and developmental needs is timely and can be used to inform population-based interventions and policies to nurture every child's health and development.

Assessing Social-Emotional Development:

Reliability and Validity of The Social-Emotional Responding Task (SERT)

Assessing emotional responses in the context of everyday social and ethical conflict is important because such experiences have been found to serve as an important affective mechanism underlying children's social behaviors and mental health (Arsenio, 2014; Malti & Krettenauer, 2013). Social-Emotional Responding Task (SERT; Malti, 2017; see Malti et al., 2009) is a measure designed to assess children's emotional responses to social and ethical transgression scenarios. While this measure has been used in past research on children's social-emotional development and kind emotions (e.g., Ongley & Malti, 2014), the psychometric properties of the original assessment tool have not been systematically evaluated yet (for an exploration of the internal consistency and factor structure of an earlier version of the tool, see Jansma and colleagues, 2018). Therefore, in this paper we present the first systematic evaluation of SERT's psychometric properties, including measurement invariance assessment across different ages and genders, using two samples of children.

In the developmental literature, the emotional responses to social and ethical transgressions have been prominently studied with the happy-victimizer paradigm. In this tradition, children were presented with hypothetical transgressions (e.g., stealing another child's chocolate) that would involve a positive outcome for the self, and they were interviewed about the emotions anticipated towards the protagonist and the victims. In classical tasks, children were asked about *harmful* acts (e.g., pushing, stealing, hurting another to obtain a positive outcome for the self) exclusively. Over the years, the task to measure the happy victimizer paradigm has been extended to other social domains, such as omission of prosocial duties (not helping, not sharing) and social exclusion (not including peers) (Malti et al., 2009). Research has shown that following

the hypothetical transgressions, the majority of 4-year-olds tend to anticipate positive emotions when imagining themselves as the transgressor (Nunner-Winkler & Sodian, 1988), whereas the majority of 8-year-olds anticipate kind emotions, such as sadness or guilt (for a review, see Malti & Ongley, 2014). The anticipation of negative emotions in these contexts has been interpreted as an indicator of experiencing kind emotions and feelings of ethical guilt because they present an awareness of the ethical norm that has been violated, as well as assuming responsibility for it (Malti, 2016; Malti, 2020).

The kind of emotions an individual anticipates upon transgressions is related to the kind of (un)ethical actions the individual would be inclined to pursue. Anticipating that transgressions will be followed by negative emotions would allow the individual to refrain from the transgression and adjust their behavior to avoid the negative social outcomes (Arsenio et al., 2006). Moreover, feelings of kind emotions (e.g., other-oriented guilt) was suggested to promote positive, reparatory behaviors following a transgression, while the absence of such feelings was suggested to be related to higher antisocial behaviors (Ghorbani et al., 2013; Hoffman, 2000; Malti, 2016; Tangney et al., 2007). Hence, anticipated emotions in the context of social conflict have been hypothesized to serve as an important affective mechanism underlying children's social behaviors (e.g., Malti & Ongley, 2014). Supporting this, the "happy victimizer" pattern has been shown to be associated with more aggression (Arsenio, 2014). Previous research also extensively showed that the anticipated kind emotions following transgression scenarios were associated with higher sympathy and prosocial behaviors (e.g., Jansma et al., 2018; Malti et al., 2007; Malti et al., 2009; also see Krettenauer et al., 2008 for a review). Meta-analytic evidence also suggests that anticipated kind emotions were linked to more prosocial behaviors and less aggression across ages 3 to 20 (Malti & Krettenauer, 2013). Therefore, kind emotions are related

to positive and negative developmental outcomes, and social-emotional development is an

important protective factor for various behavioral and mental health problems (Malti, 2020).

Hence, their assessment is of profound importance for researchers aiming to examine the

protective factors for positive developmental and behavioral outcomes, as well as for

practitioners aiming to prevent and/or intervene upon these outcomes.

Although the significance of these emotions for children's well-being and social

capacities has been acknowledged in the literature, very few, if any, assessment tools have been

developed to date, that examine these emotions across different social contexts and ages reliably.

Only a few population-based and individual-based measures have been developed and

implemented to measure the general social-emotional functioning (e.g., Early Development

Instrument, Janus & Offord, 2007; and Ages and Stages Questionnaire: Social-Emotional,

Squires et al., 2009; for a review, see Malti et al., 2018). While population-based tools have been

used to inform policy planning by providing information on average social-emotional

functioning across different communities, understanding individual differences is critical for the

implementation of interventions (see Malti et al., 2016). Moreover, even if kind emotions are

significant precursors of ethical actions, hence positive social-emotional outcomes (Arsenio et

al., 2006), these tools are more general and do not specifically target kind emotions. A number of

individual-based social-emotional tools have been developed and implemented as well (see Malti

et al., 2018). However, many of the existing tools did not undergo an in-depth examination of

their psychometric properties. In addition, only some of the existing instruments report a clear

theoretical rationale underlying the selection of the core construct(s) to be measured.

In the current study, we aimed at filling some of these gaps. We introduce the Social-

Emotional Responding Task (Malti et al., 2009), a tool to assess children's anticipated emotions

in hypothetical contexts involving various types of social and ethical conflicts. It relies on the theoretical model of the development of self-conscious emotions, which argues that an understanding of the validity of the ethical rules by itself might not be sufficient for moral motivation and actions, and instead ethical acts require children's acceptance of the rule and the personal commitment to the rule (Malti et al., 2009; Malti & Ongley, 2014; Tangney et al., 2007). Supporting this, Keller and colleagues (2003) showed that children reported kind emotions more when they were asked how they would feel as the transgressor as compared to when they were asked to anticipate emotions of another. These results suggested that anticipating emotions of the self would be different than anticipating emotions of others and emotions of self would be more related to negative feelings upon transgressions. Informed by this theoretical model, the SERT vignettes were designed to capture the development of kind emotions and examine the role of individual differences in social-emotional functioning. In the vignettes, children are presented with hypothetical transgressions and asked to anticipate how they would feel if they had transgressed the respective norm (e.g., stealing chocolate from another child). Responses are coded and collapsed into overt categories (e.g., positive versus kind emotions). The current study aimed to systematically evaluate the psychometric properties of the full instrument, including measurement invariance assessment, using two different samples to ensure the robustness of the results.

Therefore, the purpose of this paper is to present the first systematic psychometric evaluation of SERT in the early and middle childhood years (4-, 6-, and 8-year-olds). We utilized two independent samples to test for the robustness of the psychometric performance of the measure. We aimed to test three main research questions. First, we examined whether children's anticipated kind emotions converge across different social and ethical contexts, or if they are

context-specific. Past research has shown that the domains in which the transgression is situated

(e.g., aggressive acts, omission of prosocial duties, social exclusion), influences the emotions

that children report to anticipate when imagining to be the transgressor (e.g., Miller, 2006;

Smetana, 1981, which might indicate domain specificity of the task (Arsenio, 2014). Yet, in a

study examining an earlier version of the SERT tool, Jansma and colleagues (2018) found that

the task had similar factor structure across different domains, hence it could be considered

domain general. Therefore, we explored if SERT can be used across different domains of

transgression (i.e., the domain general approach), or if a domain-specific approach is more

appropriate, without a specific hypothesis.

Second, we aimed to examine the measurement invariance of the task across age groups

and genders. Examining and establishing measurement invariance among different age and

gender groups would mean that the scale is perceived, understood, and responded to similarly

across the different groups. Since most of the social-emotional and cognitive development takes

place in early and middle childhood, we examined the SERT with 4, 6, and 8-year-old children.

Specifically, we focused on 4-, 6-, and 8-year-old children because these ages are critical in

social-emotional and social-cognitive development like understanding of emotions or theory of

mind understanding, that are related to more advanced moral reasoning (e.g., Malti & Ongley,

2014; Lane et al., 2010). Higher levels of inhibitory control of older children might also help

them suppress possible positive emotions that could be elicited from having a positive outcome

out of the transgression and activate the understanding of another's standpoint, hence activating

more kind emotions. Moreover, the happy victimizer paradigm was also expected to diminish

around ages 6 or 7 (e.g., Arsenio et al., 2006; Krettenauer et al., 2008). Based on the results from

the previous research revealing the happy victimizer paradigm (e.g., Keller et al., 2003; Nunner-

Winkler & Sodian, 1988), we expected 4-year-old children would display less sad emotions when anticipating to be the transgressor, while we expected 6- and 8-year-olds to display more kind emotions.

Third, we aimed to assess the construct validity of SERT. We explored this by testing links between anticipated emotions with both prosocial adaptive outcomes (i.e., sympathy and prosocial behaviors) and aggressive behaviors, since these are the outcomes that have been linked to anticipated kind emotions in the past (e.g., Krettenauer et al., 2008). As suggested by previous research, we hypothesized that the children who report kind emotions upon anticipated transgressions would display higher levels of sympathy and prosocial behaviors and lower levels of aggression (e.g., Malti & Krettenauer, 2013).

Method

Participants

This study used two samples recruited from community events in a major city in Canada. The first sample included 291 children (M_{age} = 6.55 years, SD = 2.02; 144 4-year-olds [min = 4.03, max = 4.99] and 147 8- year-olds [min = 8.01 and max = 9.78]; 50% boys). The second sample included 282 children (M_{age} = 6.57 years, SD = 1.56; 90 4- year-olds [min = 3.46 and max = 5.05] year-olds, 105 6-year-olds [min = 6.05 and max = 7.29], and 87 8- year-olds [min = 8.04 and max = 9.48]; 49% boys). The samples were comparable in ethnical and socioeconomic backgrounds. In both of the samples, the majority of the caregivers who filled in the questionnaires were mothers (83.8% in sample 1 and 80.8% in sample 2) and the majority of children were from intact families (88% in sample 1 and 83.3% in sample 2). In terms of ethnicity, 20% of participants in the first and 21% of participants in the second sample identified their ethnic group as Middle Eastern, 15% of each sample identified as South Asian; 9% of the

first and 12% of the second sample identified as American, 7% of the first and 10% of the

second sample as East Asian; 7% of the first and 11% of the second sample as West European,

5% of the first and 4 % of the second sample as East European, 5% of the first and 3% of the

second sample as Central or South American, 1% of each sample as African. The rest of the

participants either reported another ethnic group (26% of the first and 21% of the second sample)

or did not report ethnic group (5% of the first and 2% of the second sample). In terms of

education, most caregivers had earned a Bachelor's degree both in the first (50%) and second

(41%) samples, followed by a Master's degree (21% of both first and second samples), college

diploma (17% of the first and 18% of the second sample), high school diploma (4% of the first

and 7% of the second sample), doctoral degree (3% of the first and 2% of the second), or trades

diploma or none (2% of the first and 3% of the second sample), 4% of the first and 8% of the

second sample did not report information on education. We measured yearly income with a 9

point Likert scale (1 = less than $9,999; 2 = $10,000-$19,999; 3 = $20,000-$29,999; 4 =

$30,000-$39,999; 5 = $40,000-$49,999; 6 = $50,000-$59,999; 7 = $60,000-$79,999; 8 =

$80,000-$124,999; 9 = 125,000 and more). In both of the samples, the majority of the

participants reported the highest income level (40.9% in sample 1 and 30.9% in sample 2) and

the mean level of income was 7.86 in sample 1 and 7.77 in sample 2 (between $60,000 and

$124,999 band in both samples). In comparison to recent census data, the levels of education and

income and the ethnicity of the participants were representative of the city in which we recruited

participating families (Statistics Canada, 2017).

Procedure

       The study was approved by the institute's Research Ethics Board. Children and their

caregivers were invited to the laboratory and provided oral and written consent, respectively.

Children were interviewed in a private room while caregivers completed the questionnaire in the waiting area. Each assessment took about 20 to 30 minutes. At the end of the study, children and caregivers were debriefed. Children received an age-appropriate book and certificate of participation. The testers were undergraduate psychology students who had been trained in developmental assessment techniques.

Measures

The Social-Emotional Responding Task (SERT)

In the Social-Emotional Responding Task (SERT; Malti, 2017), children were presented six vignettes in total that presented transgressions in different domains (i.e., aggressive acts, prosocial omission, and social exclusion). The children were asked to report how they would feel if they were the transgressor in these vignettes. Two vignettes depicted aggressive acts of transgressions, inducing intentional harm to the other person (i.e., stealing from another child and pushing another child to take their place in line), 2 vignettes depicted transgression via prosocial omission (i.e., not sharing and not helping when that was socially expected), and 2 vignettes depicted transgression via social exclusion (i.e., excluding another child based on non-familiarity and based on low socioeconomic status). The vignettes were presented with non-animated visual drawings that were presented on a computer screen. For each vignette, children were asked to imagine themselves as the transgressor. In each vignette, the picture illustrations of the transgressing story character's gender, age, and skin tone matched the target child. An example vignette from the aggressive acts domain is: "*Imagine that you are at school. Another boy in your class, Mark, shows you a chocolate bar he brought from home. He then puts it back in his backpack. You really like chocolate. When Mark leaves, you wait until nobody is watching, and you take the chocolate bar from his backpack so you can eat it.*" Anticipated emotions were

elicited by the question "How would you feel if you had done this?". If children responded "I

don't know", a probing question was asked, "If you had taken the chocolate and ate it, would you

feel a little good, a little bad, or a little good and bad?"

Coding. Children's responses were coded in two stages using a well-established coding

system (Malti et al., 2009). First, two research assistants independently coded responses into 12

categories, including happy, sad, bad, sorry, guilty, embarrassed, disgusted, scared, anxious,

angry, neutral, and other/undifferentiated. Then, similar categories of negative emotions (i.e.,

sad, bad, sorry, guilty, embarrassed, disgusted; excluding fearful, scared, anxious, and angry)

were collapsed into the kind emotions category (Malti, 2020). We only focused on kind emotions

since this is the appropriate response upon wrongdoing. Each child received a score of 1 if they

reported a kind emotion and received a score of 0 if they reported other emotions. For children

who responded with more than one type of emotion (12% to 24%), the primary (first) response

was used. The coders coded a random subsample (10%) of the data to establish inter-coder

reliability for all 12 categories of emotions (Cohen's κs = 1.00 across the vignettes).

Sympathy, Prosocial Behavior, and Aggressive Behavior

Caregivers rated children's sympathy via Parent Reports of Children's Sympathy scale

(Zhou et al., 2003), using a 7-point Likert scale (0 = *never* to 6 = *always*; Zhou et al., 2003). Five

items (e.g., "feel sorry for others who are less fortunate") were averaged to create a composite of

sympathy (αs = .89 in sample 1 and .92 in sample 2).

Caregivers rated children's prosocial behavior on a 7-point Likert scale (0 = *never* to 6 =

always) using the Strengths and Difficulties Questionnaire (Goodman, 1997). Five items (e.g.,

"helpful if someone is hurt, upset, or feeling ill") were averaged to create a composite of

prosocial behavior (αs = .82 in sample 1 and .92 in sample 2).

Caregivers rated children's overt aggressive behavior on a 7-point Likert scale (0 = *never*

to 6 = *always*) using items from the Child Behavior Checklist (CBCL – 1.5-5 & 6-18 years;

Achenbach & Rescorla, 2000, 2001). Eight items (e.g., "gets in fights") were averaged to create

a composite of aggressive behavior (αs = .85 in sample 1 and .87 in sample 2).

Data Analysis Plan

All analyses were conducted in M*plus* 7 (Muthén & Muthén, 1998-2017). To test our first

two research questions on the factor structure and measurement invariance of the SERT, we ran

confirmatory factor analyses. Structural equation models were conducted to test the construct

validity in relation to the behavioral measures. We used the WLSMV estimator with theta

parameterization in MPlus because the indicators were binary . WLSMV also reduces potential

bias in estimations due to missing data under the assumption of missing at random (Asparouhov

& Muthén, 2010). We adopted the following guidelines to indicate a good fit of CFA models with

categorical data: a non-significant χ^2 ($p > .05$), CFI $\geqslant$ .96, and RMSEA $\leqslant$ .05 (Yu, 2002).

Nested models were compared using $\Delta\chi^2$ from the DIFFTEST function complemented with

ΔCFI, such that a non-significant change in χ^2 ($p > .05$) and a less than .005 decrease in CFI

indicate that the restriction did not significantly worsen the model fit (i.e., evidence for

invariance in measurement invariance tests; Chen, 2007). We run factor analyses to examine the

domain-specificity versus generality of the assessment tool.

Measurement invariance across age groups (i.e., 4-, 6-, and 8-year-olds) and gender was

tested in a multi-group confirmatory factor analysis (MGCFA) framework to ensure that latent

factors of anticipated emotions were comparable across groups. When assessing measurement

invariance, the initial step of testing is to establish configural invariance (Byrne, 2012). Once the

configural invariance is established, further constraints are added to test whether or not the

model is invariant across groups, which would indicate that the measurement tool is suitable to be used with the different groups tested (Vandenberg & Lance, 2000). In line with that, first, we tested a *configural* invariance model, in which all factor loadings and indicator thresholds were freely estimated across groups except for some constraints for identification purpose (i.e., residual variances of indicators were fixed to 1, latent factor mean was fixed to 0, and latent factor variance was fixed to 1). Next, we tested a *metric* invariance model, in which all factor loadings were constrained to be equal across groups above and beyond the configural invariance model. The establishment of metric invariance enables meaningful comparison of latent factor variances and covariances across groups. Lastly, we tested a *scalar* invariance model, in which thresholds of all indicators were constrained to be equal across groups above and beyond the metric invariance model (latent factor mean and variance were now freely estimated in the reference group). The establishment of scalar invariance enables meaningful comparison of latent factor means across groups. When full invariance (i.e., invariance of all indicators) was not supported, we tested partial invariance (i.e., some indicators were constrained to be equal across groups whereas others were not) based on modification indices and theory. For assessing construct validity, the association of the SERT with prosocial and aggressive outcomes were examined via Structural Equal Models in the two samples.

Results

Percentages of kind emotion by vignette and age showed that 8-year-olds displayed more kind emotions across all of the different vignettes in both samples (see Table 1). Most children had complete data across six vignettes (96% in the first and 88% in the second sample). Missing data rates were low on the vignette level (1.7% to 2.7% in the first sample and 4.3% to 7.1% in the second sample).

We examined our research question on the domain-specificity versus generality of the

SERT through model comparison in confirmatory factor analyses. The "*domain specific*" model

was based on the hypothesis that children's anticipated emotions would vary across domains of

aggressive act, prosocial omission, and social exclusion; thus, a three-factor model was expected

to underlie the six vignettes. The "*domain general*" model was based on a competing hypothesis

that children's anticipated emotions would converge across domains of aggressive act, prosocial

omission, and social exclusion; thus, a general factor was expected to underlie all six vignettes.

Models (i.e., the one-factor model versus the three-factor model) were tested and compared using

confirmatory factor analyses (see factor loadings of the one-factor model and the three-factor

model for both samples in Tables S1 and S2, respectively). In sample 1, both "domain specific

(three-factor)" (χ^2 ($df = 6$) = 4.97, p =.55, CFI = 1.00, RMSEA = .00) and "domain general (one-

factor)" (χ^2 ($df = 9$) = 7.30, p =.61, CFI = 1.00, RMSEA = .00) models fit well with the data for

anticipated kind emotion. The two models were not significantly different in terms of model fit

($\Delta\chi^2$ ($df = 3$) = 2.86, p = .41, ΔCFI = .000), lending support to the more parsimonious, "domain

general", model. The pattern of results was fully replicated in sample 2 (domain specific model:

χ^2 ($df = 6$) = 6.93, p =.33, CFI = .99, RMSEA = .02; domain general model: χ^2 ($df = 9$) = 7.39, p

=.60, CFI = 1.00, RMSEA = .00; difference tests: $\Delta\chi^2$ ($df = 3$) = 1.08, p = .78, ΔCFI = .001).

Moreover, standardized factor loadings were also high in both samples (ranging from .70 to .93

in sample 1 and .76 to .83 in sample 2) and the reliability coefficients, assessed with Cronbach's

alpha, were .83 and .85 for anticipated kind emotion, in samples 1 and 2 respectively.

For measurement invariance tests across age groups and gender, model fit of and model

comparison across the configural, metric, and scalar invariance models were evaluated using the

previously mentioned criteria (see Table 2 for the results of the measurement invariance tests).

Across the different age groups, full metric and partial scalar invariance were supported in sample 1. Specifically, the threshold of the "school bus" vignette showed scalar invariance across age groups. Even though in general 8-year-olds reported higher kind emotions in the school bus vignette as compared to 4-year-olds, for children with the same latent score of the anticipated kind emotion in different age groups, 8-year-olds tended to report lower levels of anticipated kind emotion in the "school bus" vignette than 4-year-olds. In sample 2, partial metric and full scalar invariance were supported. Specifically, factor loading of the "school bus" vignette was relatively higher in 6-year-olds than in 4- and 8-year-olds. Across gender, full metric and full scalar invariance were supported in both samples.

Partial or full invariance models also showed a consistent developmental increase in anticipated kind emotions. In sample 1, 8-year-olds had higher latent mean scores on anticipated kind emotions ($M_{diff} = 1.350, p < .001$) than 4-year-olds. In sample 2, 8-year-olds did not differ from 6-year-olds on latent mean scores of anticipated kind emotions, but both groups had higher latent scores ($M_{diff} = 1.264$ and $1.279, p < .001$) than 4-year-olds. No significant gender differences were found on latent mean scores of anticipated kind emotions.

Lastly, we assessed the construct validity by exploring links to prosocial outcomes (i.e., sympathy and prosocial behavior) and aggressive behaviors in Structural Equal Models in the two samples (see Table 3 for construct validity for the total sample and across different age groups and genders). The full or partial invariance model established earlier were used when testing construct validity across age and gender. In sample 1, anticipated kind emotions were positively related to caregiver-reported prosocial behavior ($r = .28, p < .01$) and sympathy ($r = .50, p < .001$), and negatively related to caregiver-reported aggressive behavior ($r = -.20, p < .01$). These relations were not significantly different across gender, Wald $\chi^2(df = 3) = 0.69$ and

2.51, ps = .88 and .47, or age, Wald $\chi^2(df = 3)$ = 1.93 and 2.00, ps = .59 and .57. Yet, when

gender and age were controlled for, only the association between kind emotions and sympathy

remained significant (r = .37, p < .01). In Sample 2, anticipated kind emotions were negatively

related to aggressive behavior (r = -.16, p < .01) and positively related to sympathy (r = .41, p

< .001). Again, the relations were not significantly different across gender, Wald $\chi^2(df = 3)$ = 1.29

and 0.28, ps = .73 and .96, or age, Wald $\chi^2(df = 6)$ = 9.29 and 5.20, ps = .16 and .52. However,

again, sympathy was the only variable related to kind emotions after child gender and age were

controlled for (r = .33, p < .01).

Discussion

Social-Emotional Responding Task (SERT) is a measurement tool aiming to examine the

emotional responses of children to transgression scenarios. This paper examined the context

specificity of the dimensions of SERT, its measurement invariance, and construct validity. SERT

is composed of six vignettes in three different domains (i.e., aggressive acts, prosocial omission,

and social exclusion). The first aim of the study was to examine whether the answers of children

depend on the domain, indicating domain-specificity across contexts, or the answers load under

one factor regardless of the domain, indicating that the task is domain general. The results

provided support for the more parsimonious, one-factor (domain general) structure of SERT,

which was compatible with the factor analysis of the preliminary, Dutch version of the scale

(Jansma et al., 2018). These results supported the reliability of SERT and indicated that the

anticipated emotions upon transgressions would not be dependent on the domain of

transgression. Instead, they suggested that there might be an underlying moral motivation and the

child might be feeling negative emotions when failing to comply with that, regardless of the type

of transgression. These results also suggest that SERT can be used as a full instrument or the

vignettes about different domains of transgression can also be used partially since the answers to different vignettes seem to converge.

The second purpose of this paper was to examine the measurement invariance of SERT across different age groups and genders. The results showed that SERT scores generally have partial scalar or scalar invariance across gender and different age groups indicating that SERT items generally have the same factor structure, same factor loadings, and same intercepts across groups. These findings showed that the mean level differences would indicate differences across different age groups and genders rather than the different interpretations of the instrument by different subgroups. This study was the first to demonstrate measurement invariance of SERT items and to show that SERT is a robust measurement tool that would function uniformly across gender and age groups. One exception was the school bus vignette, which revealed some invariance across age groups. This task violated full scalar invariance showing that the item intercepts were not equal across two age groups in the first sample. This vignette was unique because it specifically targeted social exclusion based on socioeconomic status (i.e., the economically advantaged child being preferred over the economically disadvantaged child). It is not unusual that children with low socioeconomic status are evaluated negatively and are socially excluded (Hjalmarsson & Mood, 2015; Shutts et al., 2016), but the reasons for children to exclude socioeconomically disadvantaged peers has been less widely investigated. In one such study, social exclusion based on economic status versus school membership among 4 and 8-year-olds were compared. The results showed that 8-year-olds (but not 4-year-olds) reported more kind emotions upon social exclusion based on economic status as compared to school membership (Dys et al., 2019). Our results also showed that 8-year-olds reported more kind emotions as compared to 4-year-olds in the school bus vignette, hence were compatible and

suggested that there might be age-related differences in the expression of kind emotions following social exclusion based on socio-economic status. Previous research suggested that starting from preschool years, children become aware of poverty and start to develop stereotypes of low socioeconomic status individuals, and with increased age, they start to attribute poverty to internal factors rather than external ones (e.g., Heberle & Carter, 2020; Sigelman, 2012). However, around 8-9 years of age, they also start to perceive the detrimental effects of poverty on individuals (e.g., Chafel & Neitzel, 2005). It is plausible that the older group is more capable of understanding the psychological effects of poverty on others and feel higher levels of kind emotions based on socially excluding the children with low socioeconomic status. Therefore, these results indicate that social exclusion based on poverty might be a specific case of exclusion as an ethical dilemma in different age groups, and this might explain the partial measurement invariance in different age groups. These results also indicate that the researchers using SERT, and especially the school bus vignette, should be careful about the age group of children and include age as a control variable in their analysis. Moreover, the previous research indicated differences in perceptions of poverty in children from low socioeconomic status (hence poverty) versus children from middle to high socioeconomic status (e.g., Heberle & Carter, 2020). Therefore, it would be beneficial for future studies to target children from different socioeconomic statuses and control for the effect of poverty in the responses of children.

Our results were also in line with the previous literature in revealing that the anticipated kind emotions increased with age. It is likely that young children were focusing on the immediate rewards (e.g., eating the chocolate bar) and had difficulty suppressing the happy emotions upon the anticipation of receiving the reward, while not being able to consider the possible long term negative outcomes and hence activate kind emotions. There was a shift in

anticipated kind emotions between the ages of 4 and 6, while 6- and 8-year-olds anticipated

similar levels of kind emotions upon imagined transgressions. These findings are compatible

with the developmental window of different socio-cognitive skills like the development of self-

awareness, inhibitory control, delay of gratification, and theory of mind understanding, as well as

the development of social-emotions like empathic understanding, and self-evaluative emotions

like guilt upon wrongdoing and understanding standards. These results are also in line with the

age-related findings of previous studies using the happy victimizer paradigm (e.g., Krettenauer et

al., 2008; Nunner-Winkler & Sodian, 1988).

        Therefore, construct validity of SERT was established reliably for sympathy and partially

(in different samples and subsamples) for aggressive and prosocial behaviors. When age and

gender were controlled for, the associations between kind emotions and prosocial and aggressive

behaviors became non-significant. This is partially in contrast with the previous meta-analysis of

Malti and Krettenauer (2013). Yet, gender effects were not controlled for in that meta-analysis. It

is plausible that child gender might have a moderating effect since some previous studies

suggested that female children generally display higher levels of prosocial behaviors while male

children are more prone to aggression (e.g., van der Graaf et al., 2017; Zimmer-Gembeck et al.,

2005). Therefore, controlling gender effects might be nullifying these associations. On the other

hand, the results are in line with previous studies that revealed a significant association between

sympathy and anticipated emotions following hypothetical transgressions (e.g., Jansma et al.,

2018; Malti et al., 2007; Malti et al., 2009), and further provided support to the validity of SERT

scores.

        The results should be interpreted in light of some limitations. In the current study, the

psychometric properties of SERT were measured using children's anticipated kind emotions

upon hypothetical transgressions. Even though the valence of the anticipated emotion is important in predicting behavioral outcomes, the justification of the anticipated emotion is also essential. Moreover, future studies might also use the physiological reactions of children upon hypothetical transgressions since physiological reactions were suggested to be less subjective. Therefore, future studies might benefit from including justifications for kind emotions and including physiological assessment of kind emotions along with children's narrative answers. In the current study, prosocial and aggressive behaviors and sympathy were all reported by mothers and this is another limitation. Future studies might benefit from using multiple informants (Clemans et al., 2014; De Los Reyes et al., 2013). Finally, our data was cross-sectional and hence did not allow for testing associations between anticipated emotions and long-term outcomes, or for examining the changes in social-emotional functioning over time. Future longitudinal studies might address that gap. Despite these limitations, the current study had several very important strengths. First, this study was the first to show the psychometric properties and measurement invariances of SERT. The psychometric properties were established using two samples of children recruited from an ethnically diverse city. The results suggested that SERT scores are reliable and the test score interpretations show the validity of the measurement tool. Measurement invariance tests suggested that the tool can be robustly used with children between the ages of 4 and 8 and across both genders. The addition of a robust measurement tool for anticipated emotions in transgressions is a significant contribution to the literature. Another strength was the age range of the population tested. Previous studies that measured anticipated emotions have generally targeted older, school-aged children, whereas the younger ages are also important to study since the vast majority of socio-cognitive and social-emotional development, which influence kind emotions, take place within that time frame (Killen & Malti, 2015). SERT

was successful in detecting the expected developmental changes, providing further support to the instrument's developmental sensitivity.

Conclusion

In general, the current study contributes to the literature by showing the psychometric validity of the scores in the Social-Emotional Responding Task (SERT) as evidenced by confirmatory factor analyses (i.e., convergence across vignettes) and associations with sympathy. Furthermore, measurement invariance was found across gender and age groups, indicating the robustness of the assessment across different age groups and genders. These results suggest that SERT is a valid assessment tool for anticipated emotions in hypothetical transgressions for children 4 to 8-years-old and for both genders.

Moreover, the assessment of anticipated kind emotions in transgression scenarios has clinical significance. Since social-emotional capacities and moral development are related to positive behavioral outcomes as well as better outcomes in mental health and academic achievement (e.g., Durlak et al., 2011; Jones et al., 2015; Malti et al., 2016), interventions aiming to ameliorate these outcomes might focus on moral development. In such programs and in general, in clinical settings, the tool can be used to predict delays and deviances in kind emotions in early and middle childhood years, which would be informative for practitioners to detect the areas that would benefit from intervention and would be beneficial for tailoring interventions based on the needs of the child. Moreover, such tools have the potential to be used as pre and post assessment strategies for tracking developmental trajectories and for assessing the effectiveness of intervention strategies targeting moral development. The results of this study indicated that SERT would be a theoretically based and psychometrically sound instrument for the assessment of kind emotions.

References

Achenbach, T. M., & Rescorla, L. A. (2000). *Child behavior checklist for ages 1 1/2-5. ASEBA*. University of Vermont.

Achenbach, T. M., & Rescorla, L. (2001). *Manual for the ASEBA school-age forms & profiles: An integrated system of multi-informant assessment.* Burlington, VT: Aseba.

Arsenio, W. (2014). Moral emotion attributions and aggression. In M. Killen & J. Smetana (Eds.), *Handbook of moral development* (2nd ed., pp. 235-255). Psychology Press. https://doi.org/10.4324/9780203581957.ch11

Arsenio, W., Gold, J., & Adams, E. (2006). Children's conceptions and displays of moral emotions. In M. Killen & J. Smetana (Eds.), *Handbook of moral development* (pp. 581–610). Lawrence Erlbaum Associates Publishers.

Asparouhov, T., & Muthén, B. (2010). Weighted least squares estimation with missing data. *Mplus Technical Appendix*, *2010*, 1-10.

Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming.* Routledge.

Chafel, J. A., & Neitzel, C. (2005). Young children's ideas about the nature, causes, justification, and alleviation of poverty. *Early Childhood Research Quarterly, 20*(4), 433-450. https://doi.org/10.1016/j.ecresq.2005.10.004

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464-504. https://doi.org/10.1080/10705510701301834

Clemans, K. H., Musci, R. J., Leoutsakos, J. M. S., & Ialongo, N. S. (2014). Teacher, parent, and peer reports of early aggression as screening measures for long-term maladaptive

outcomes: Who provides the most useful information? *Journal of Consulting and Clinical Psychology, 82*(2), 236-247. https://doi.org/10.1037/a0035918

De Los Reyes, A., Thomas, S. A., Goodman, K. L., & Kundey, S. M. (2013). Principles underlying the use of multiple informants' reports. *Annual Review of Clinical Psychology, 9*, 123-149. https://doi.org/10.1146/annurev-clinpsy-050212-185617

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

Dys, S. P., Peplak, J., Colasante, T., & Malti, T. (2019). Children's sympathy and sensitivity to excluding economically disadvantaged peers. *Developmental Psychology, 55*(3), 482–487. http://dx.doi.org/10.1037/dev0000549

Ghorbani, M., Liao, Y., Çayköylü, S., & Chand, M. (2013). Guilt, shame, and reparative behavior: The effect of psychological proximity. *Journal of Business Ethics*, *114*(2), 311-323. https://doi.org/10.1007/s10551-012-1350-2

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Child Psychology & Psychiatry & Allied Disciplines, 38*(5), 581-586. https://doi.org/10.1111/j.1469-7610.1997.tb01545.x

Heberle, A. E., & Carter, A. S. (2020). Young children's stereotype endorsement about people in poverty: Age and economic status effects. *Children and Youth Services Review, 108*. https://doi.org/10.1016/j.childyouth.2019.104605

Hjalmarsson, S., & Mood, C. (2015). Do poorer youth have fewer friends? The role of household

and child economic resources in adolescent school-class friendships. *Children and Youth*

Services Review, 57, 201-211. https://doi.org/10.1016/j.childyouth.2015.08.013

Hoffman, M. L. (2000). *Empathy and Moral Development: Implications for caring and justice.*

Cambridge University Press. https://doi.org/10.1017/CBO9780511805851

Jansma, D., Malti, T., Opdenakker, M. C., & van der Werf, G. (2018). Assessment of anticipated

emotions in moral transgressions. *European Journal of Psychological Assessment, 34*(2),

111-126. https://doi.org/10.1027/1015-5759/a000467

Janus, M., & Offord, D. R. (2007). Development and psychometric properties of the Early

Development Instrument (EDI): A measure of children's school readiness. *Canadian*

Journal of Behavioural Science/Revue Canadienne des Sciences du Comportement,

39(1), 1-22. https://doi.org/10.1037/cjbs2007001

Jones, D. E., Greenberg, M., & Crowley, M. (2015). Early social-emotional functioning and

public health: The relationship between kindergarten social competence and future

wellness. *American Journal of Public Health, 105*(11), 2283-2290.

https://doi.org/10.2105/AJPH.2015.302630

Keller, M., Lourenço, O., Malti, T., & Saalbach, H. (2003). The multifaceted phenomenon of

'happy victimizers': A cross-cultural comparison of moral emotions. *British Journal of*

Developmental Psychology, 21(1), 1-18. https://doi.org/10.1348/026151003321164582

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.

https://doi.org/10.1016/bs.acdb.2014.11.007

Krettenauer, T., Malti, T., & Sokol, B. (2008). The development of moral emotions expectancies

and the happy victimizer phenomenon: A critical review of theory and applications.

European Journal of Developmental Science, 2(3), 221–235.

https://doi.org/10.3233/DEV-2008-2303

Lane, J. D., Wellman, H. M., Olson, S. L., LaBounty, J., & Kerr, D. C. (2010). Theory of mind

and emotion understanding predict moral development in early childhood. *British*

Journal of Developmental Psychology, 28(4), 871-889.

https://doi.org/10.1348/026151009X483056

Malti, T. (2020). Kindness: A perspective from developmental psychology. *European Journal of*

Developmental Psychology. Advance online publication.

https://doi.org/10.1080/17405629.2020.1837617

Malti, T. (2017). The Social-Emotional Responding Task. Unpublished tool by T. Malti, 2017.

Copyright by T. Malti, 2017.

Malti, T. (2016). Toward an integrated clinical-developmental model of guilt. *Developmental*

Review, 39, 16-36. https://doi.org/10.1016/j.dr.2015.11.001

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, 45*(6)*,* 718-731.

https://doi.org/10.1080/15374416.2015.1121822

Malti, T., Gummerum, M., & Buchmann, M. (2007). Contemporaneous and 1-year longitudinal

prediction of children's prosocial behavior from sympathy and moral motivation*. The*

Journal of Genetic Psychology, 168(3), 277–299.

https://doi.org/10.3200/GNTP.168.3.277-300

Malti, T., Gummerum, M., Keller, M., & Buchmann, M. (2009). Children's moral motivation, sympathy, and prosocial behavior. *Child Development*, *80*(2), 442-460. https://doi.org/10.1111/j.1467-8624.2009.01271.x

Malti, T., & Krettenauer, T. (2013). The relation of moral emotion attributions to prosocial and antisocial behavior: A meta-analysis. *Child Development, 84*(2)*, 397-412. https://doi.org/10.1111/j.1467-8624.2012.01851.x

Malti, T., & Ongley, S.F. (2014). The development of moral emotions and moral reasoning. In M. Killen & J. Smetana (Eds.), *Handbook of moral development* (2nd ed., pp. 163-183). Psychology Press. https://doi.org/10.4324/9780203581957.ch8

Malti, T., Zuffianò, A., & Cheung, C. (2018). Measuring social-emotional correlates of aggression in children and adolescents. In T. Malti & K. H. Rubin (Eds.), *Handbook of child and adolescent aggression* (pp. 297-317). The Guilford Press.

Miller, J. G. (2006). Insights into moral development from cultural psychology. In M. Killen & J. Smetana (Eds.), *Handbook of moral development* (pp. 375-398). Lawrence Erlbaum Associates Publishers.

Muthén, L.K. & Muthén, B.O. (1998-2017). *Mplus User's Guide. Eighth Edition*. Muthén & Muthén.

Nunner-Winkler, G., & Sodian, B. (1988). Children's understanding of moral emotions. *Child Development,* 59(5), 1323-1338. https://doi.org/10.2307/1130495

Ongley, S.F., & Malti, T. (2014). The role of moral emotions in the development of children's sharing behavior. *Developmental Psychology, 50*(4), 1148–1159. doi:10.1037/a0035191

Shutts, K., Brey, E. L., Dornbusch, L. A., Slywotzky, N., & Olson, K. R. (2016). Children use

wealth cues to evaluate others. *PLoS ONE, 11*(3), e0149360.

https://doi.org/10.1371/journal.pone.0149360

Sigelman, C. K. (2012). Rich man, poor man: Developmental differences in attributions and

perceptions. *Journal of Experimental Child Psychology, 113*(3), 415–429.

https://doi.org/10.1016/j.jecp.2012.06.011

Smetana, J. G. (1981). Preschool children's conceptions of moral and social rules. *Child

Development, 52*(4), 1333-1336. https://doi.org/10.2307/1129527

Squires, J., Twombly, E., Bricker, D., & Potter, L. (2009). *ASQ-3 Ages and stages questionnaires

user's guide*. Lane County.

Statistics Canada. (2017). *Census profile, 2016 Census* (Statistics Canada Catalogue no. 98-316-

X2016001). https://www12.statcan.gc.ca/census-recensement/2016/dp-

pd/prof/index.cfm?Lang=E

Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral emotions and moral behavior. *Annual

Review of Psychology*, *58*, 345-372.

https://doi.org/10.1146/annurev.psych.56.091103.070145

Van der Graaff, J., Carlo, G., Crocetti, E., Koot, H. M., & Branje, S. (2017). Prosocial behavior

in adolescence: gender differences in development and links with empathy. *Journal of

Youth and Adolescence*, *47*(5), 1086-1099. https://doi.org/10.1007/s10964-017-0786-1

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance

literature: Suggestions, practices, and recommendations for organizational research.

Organizational Research Methods, 3(1), 4-70. https://doi.org/10.1177/109442810031002

Yu, C. Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes* [Unpublished doctoral dissertation]. University of California, Los Angeles.

Zhou, Q., Valiente, C., & Eisenberg, N. (2003). Empathy and its measurement. In S. J. Lopez & C. R. Snyder (Eds.), *Positive psychological assessment: A handbook of models and measures* (p. 269–284). American Psychological Association. https://doi.org/10.1037/10612-017

Zimmer-Gembeck, M. J., Geiger, T. C., & Crick, N. R. (2005). Relational and physical aggression, prosocial behavior, and peer relations: Gender moderation and bidirectional associations. *The Journal of Early Adolescence*, *25*(4), 421-452. https://doi.org/10.1177/0272431605279841

Table 1

Percentage of Kind Emotions by Vignettes and Age Groups in Sample 1 and Sample 2

	Sample 1		Sample 2		
	4-year-old $n = 144$	8-year-old $n = 147$	4-year-old $n = 90$	6-year-old $n = 105$	8-year-old $n = 87$
1. AA: Chocolate	47.1	83.0	42.9	67.0	79.3
2. AA: Lollipop	51.8	85.7	50.6	77.5	82.1
3. PO: Ice-cream	45.3	76.9	39.2	68.9	68.3
4. PO: Music	57.4	87.1	45.0	80.6	86.6
5. SE: Painting	38.7	75.5	40.2	77.2	80.8
6. SE: School bus	26.5	82.3	32.1	64.4	78.3

Note. AA = Aggressive Acts; PO = Prosocial Omission; SE = Social Exclusion.

Table 2

Measurement Invariance Tests Across Development and Gender for Anticipated Kind Emotions

Model	χ^2	df	p	CFI	RMSEA	Model Comparison	$\Delta\chi^2$	Δdf	p	ΔCFI	Δ RMSEA
Age											
Sample 1											
1. Configural	22.21	18	.13	0.986	.040						
2. Metric	20.11	24	.69	1.000	.000	2 vs. 1	1.85	6	.93	.014	-.040
3. Scalar	30.23	28	.35	0.993	.023	3 vs. 2	8.90	4	06	-.007	.023
3a. Partial Scalar	21.41	27	.77	1.000	.000	3a vs. 1[1]	1.52	9	.99	.014	-.040
Sample 2											
1. Configural	19.00	27	.87	1.000	.000						
2. Metric	51.19	39	.09	0.981	.058	2 vs. 1	23.01	12	.03	-.019	.058
2a. Partial Metric	39.86	38	.39	0.997	.023	2a vs. 1	15.33	11	.17	-.003	.023
3. Scalar	42.36	46	.63	1.000	.000	3 vs. 1[1]	21.25	19	.32	.000	.000
Gender											
Sample 1											
1. Configural	12.06	18	.84	1.000	.000						
2. Metric	19.00	24	.75	1.000	.000	2 vs. 1	5.13	6	.53	.000	.000
3. Scalar	27.48	28	.49	1.000	.000	3 vs. 2	7.71	4	.10	.000	.000
Sample 2											
1. Configural	14.75	18	.68	1.000	.000						
2. Metric	19.57	24	.72	1.000	.000	2 vs. 1	4.72	6	.58	.000	.000
3. Scalar	22.68	28	.75	1.000	.000	3 vs. 1[1]	8.04	10	.63	.000	.000

Note. Incremental fit indices that indicate non-invariance were bolded. [1]Estimation problem occurred in chi-square difference tests when comparing the scalar invariance models to the metric invariance models; thus, the scalar invariance models were compared to the configural invariance model instead.

Table 3

Construct Validity in Relation to Prosocial Behavior, Aggressive Behavior, and Sympathy for the Total Sample and by Age and Gender

	Sample 1			Sample 2		
	Prosocial Behavior	Aggressive Behavior	Sympathy	Prosocial Behavior	Aggressive Behavior	Sympathy
Total Sample						
No covariates	0.28**	-0.20**	0.50***	0.12	-0.16*	0.41***
Age and gender as covariates	0.21+	0.04	0.37**	0.08	-0.06	0.33*
By Gender						
Girl	0.29**	-0.13	0.43***	0.12	-0.13	0.32**
Boy	0.12	-0.17*	0.29*	0.06	-0.11	0.29*
By Age						
4-yrs	0.17+	0.05	0.32**	0.08	0.02	0.29+
6-yrs	-	-	-	-0.06	-0.04	0.10
8-yrs	0.09	-0.01	0.12	0.26	-0.17	0.32

Note. ***$p<.001$; **$p<.01$; *$p < .05$; +$p < .10$.