

Cross-Informant Assessment of Children's Sympathy: Disentangling Trait and State Agreement

Antonio Zuffiano^{1*}, Stefania Sette², Tyler Colasante³, Marlis Buchmann⁴, Tina Malti³

¹Liverpool Hope University, United Kingdom, ²Faculty of Medicine and Psychology, Sapienza University of Rome, Italy, ³Department of Psychology, University of Toronto, Canada, ⁴Universität Zürich, Switzerland

Submitted to Journal:

Frontiers in Applied Mathematics and Statistics

Specialty Section:

Quantitative Psychology and Measurement

Article type:

Original Research Article

Manuscript ID:

342415

Received on:

08 Jan 2018

Revised on:

05 Mar 2018

Frontiers website link:

www.frontiersin.org

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Author contribution statement

TM corrected and revised the whole manuscript

Keywords

sympathy, social-emotional development, Informant discrepancies, latent state-trait model, longitudinal models.

Abstract

Word count: 227

The use of multiple informants (e.g., caregivers and teachers) is recommended to obtain a comprehensive profile of children's social emotional development. Evidence to date indicates that only a small-to-moderate degree of convergence exists between different informants' assessments of children's social-emotional functioning, especially when the contexts of such informants' observations are also different. However, whether caregivers and teachers primarily disagree about children's dispositional emotional tendencies or situational emotional fluctuations remains unclear. In this study, we investigated the extent to which caregivers and teachers converged in their evaluation of children's dispositional and state sympathy (i.e., a relatively internal and low visibility emotional response of concern for another's wellbeing) in a nationally representative sample of Swiss children (N = 1,273) followed from 6 to 12 years of age. Using analyses based in latent state-trait theory, we found that caregivers and teachers showed moderate-to-large agreement ($r = .510$) at the dispositional, trait level of children's sympathy, but only a small level of agreement in their assessments of children's situational, state-like manifestations of sympathy ($r = .123$). These findings highlight the differential convergence of adults' ratings of one core dimension of children's social-emotional development, i.e., sympathy, at the dispositional and situational levels, and, relatedly the need to investigate the reasons behind discrepancies at both levels of analysis. We elaborate on practical implications for designing social-emotional screening tools across different informants and contexts.

Funding statement

This research was funded by the Swiss National Science Foundation and the Jacobs Foundation.

Ethics statements

(Authors are required to state the ethical considerations of their study in the manuscript, including for cases where the study was exempt from ethical approval procedures)

Does the study presented in the manuscript involve human or animal subjects: Yes

Please provide the complete ethics statement for your manuscript. Note that the statement will be directly added to the manuscript file for peer-review, and should include the following information:

- Full name of the ethics committee that approved the study
- Consent procedure used for human participants or for animal owners
- Any additional considerations of the study in cases where vulnerable populations were involved, for example minors, persons with disabilities or endangered animal species

As per the Frontiers authors guidelines, you are required to use the following format for statements involving human subjects: This study was carried out in accordance with the recommendations of [name of guidelines], [name of committee]. The protocol was approved by the [name of committee]. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

For statements involving animal subjects, please use:

This study was carried out in accordance with the recommendations of 'name of guidelines, name of committee'. The protocol

was approved by the 'name of committee'.

If the study was exempt from one or more of the above requirements, please provide a statement with the reason for the exemption(s).

Ensure that your statement is phrased in a complete way, with clear and concise sentences.

The current study was conducted in Switzerland and consisted of non-invasive and unconstrained parent and teacher questionnaires. According to the regulations in the canton of Zurich in Switzerland (the so-called "Regulations of the Ethics Commission for Psychological Research", 2011), there was no requirement for an ethics committee approval when the study was conducted. According to this regulation (Article 5, paragraph 1), this study was exempted from requiring formal ethical approval. The study fully complies with the ethics guidelines given by this legal regulation (see Article 8, paragraph 2). The regulation is based on the "Ethical Principals of Psychologists and Code of Conduct" (as outlined in the so-called "Ethical Guidelines for Psychologists of the Swiss Society for Psychology, as amended on October 13, 2003) and the ethical standards of the American Psychological Association (APA). Written and informed consent was obtained from all research participants and from the parents / legal guardians of all non-adult participants. The data were analyzed anonymously.

In review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

Cross-Informant Assessment of Children's Sympathy:
Disentangling Trait and State Agreement

Antonio Zuffianò¹, Stefania Sette², Tyler Colasante³, Marlis Buchmann⁴, and Tina Malti^{3,4}

¹ Liverpool Hope University

² Sapienza University of Rome

³ University of Toronto

⁴ Jacobs Center for Productive Youth Development, University of Zurich

Author Note

Antonio Zuffianò, Department of Psychology, Liverpool Hope University; Stefania Sette, Department of Developmental and Social Psychology, Sapienza University of Rome; Tyler Colasante, Tina Malti, Department of Psychology and Department of Psychiatry, University of Toronto; Marlis Buchmann, Tina Malti, Jacobs Center for Productive Youth Development, University of Zurich. This research was funded by the Swiss National Science Foundation and the Jacobs Foundation. The authors would like to express their sincere thanks to the children, parents, and teachers for participating in the study. Moreover, the authors are grateful to all the interviewers and undergraduate students for their help in data collection and coding.

Correspondence concerning this article should be addressed to Antonio Zuffianò, Department of Psychology, Liverpool Hope University, Hope Park Campus, L169JD, Liverpool, UK. Electronic mail may be sent to zuffiaa@hope.ac.uk.

24 Abstract

25 The use of multiple informants (e.g., caregivers and teachers) is recommended to obtain a
26 comprehensive profile of children's social emotional development. Evidence to date indicates
27 that only a small-to-moderate degree of convergence exists between different informants'
28 assessments of children's social-emotional functioning, especially when the contexts of such
29 informants' observations are also different. However, whether caregivers and teachers primarily
30 disagree about children's dispositional emotional tendencies or situational emotional fluctuations
31 remains unclear. In this study, we investigated the extent to which caregivers and teachers
32 converged in their evaluation of children's dispositional and state sympathy (i.e., a relatively
33 internal and low visibility emotional response of concern for another's wellbeing) in a nationally
34 representative sample of Swiss children ($N = 1,273$) followed from 6 to 12 years of age. Using
35 analyses based in latent state-trait theory, we found that caregivers and teachers showed
36 moderate-to-large agreement ($r = .510$) at the dispositional, trait level of children's sympathy,
37 but only a small level of agreement in their assessments of children's situational, state-like
38 manifestations of sympathy ($r = .123$). These findings highlight the differential convergence of
39 adults' ratings of one core dimension of children's social-emotional development, i.e., sympathy,
40 at the dispositional and situational levels, and, relatedly the need to investigate the reasons
41 behind discrepancies at both levels of analysis. We elaborate on practical implications for
42 designing social-emotional screening tools across different informants and contexts.

43 *Keywords:* sympathy, social-emotional development, informant discrepancies, latent
44 state-trait model, longitudinal models.

45

46

47 Cross-Informant Assessment of Children's Sympathy:
48 Disentangling Trait and State Agreement

49 A recommended practice in developmental and clinical research is the use of different
50 informants (e.g., caregivers, teachers, peers, clinicians, etc.) to assess children's social-emotional
51 development, behavioral functioning, and mental health (De Los Reyes, Thomas, Goodman, &
52 Kunder, 2013). From a practical perspective, using data from several sources is important to
53 obtain a comprehensive profile of children's strengths and needs, which can help plan
54 appropriate intervention. Researchers tend to interpret results that are stable across informants as
55 more trustworthy because they do not depend on a specific informant, and the degree of
56 convergence between informants is thus thought to indicate the child's general score for the
57 construct under investigation. However, a large amount of empirical data indicates that only a
58 small-to-moderate amount of agreement exists between different informants of children's social-
59 emotional development and (mal)adaptive behavior (De Los Reyes & Kazdin, 2004, 2005).
60 Although several factors may account for this inconsistency (e.g., different contexts of
61 observation and reference points; De Los Reyes & Kazdin, 2005), the level of analysis at which
62 it occurs remains unclear.

63 Here, we addressed this gap using the conceptual and methodological framework of
64 latent state-trait (LST) theory (Steyer, Ferring, & Schmitt, 1992). We applied LST to assess the
65 extent to which caregivers and teachers converged in their evaluations of children's sympathy
66 (i.e., affective concern for others' welfare; Eisenberg, Spinrad, & Knafo-Noam, 2015) which is a
67 core dimension of social-emotional development (Malti, Sette, & Dys, 2016; Malti & Song, in
68 press). We investigated this question at two different levels: (1) the dispositional or *trait level*,
69 reflecting children's sympathetic tendencies across time, and (2) the *state level*, reflecting

70 fluctuations in children's sympathetic responses at a given point in time. We focused on
71 children's sympathy because it is regarded as a core social-emotional skill and has been
72 associated with various positive and negative developmental outcomes (for reviews, see
73 Eisenberg et al., 2015; Malti & Song, in press). Its reliable assessment is also highly relevant to
74 clinical contexts ranging in severity (e.g., for the assessment of callous-unemotional traits among
75 high-risk youth [Kimonis, Frick, Muñoz, & Aucoin, 2008] and social-emotional competencies in
76 schools [Malti, Chaparro, Zuffianò, & Colasante, 2016]). We expected caregivers and teachers to
77 agree more at the dispositional versus situational level of children's sympathy because the latter
78 is by definition more ephemeral and sensitive to contextual features, which likely differ
79 significantly for caregivers and teachers at home and school, respectively.

80 **Cross-Informant Convergence in the Assessment of Children's Sympathy**

81 Sympathy is a specific emotional response that includes feelings of concern or sorrow for
82 another's emotional state or welfare (Eisenberg et al., 2015). In comparison to empathy, which
83 generally involves sharing the emotions of another, but not necessarily feeling concern for them,
84 sympathy is more likely to be implicated in prosocial and aggressive behaviors (Eisenberg,
85 Spinrad, & Morris, 2014; Zuffianò, Colasante, Buchmann, & Malti, 2017).

86 Different methods (e.g., questionnaires and observations) and informants (e.g., caregivers
87 and teachers) have been used to assess sympathy across childhood and adolescence (Kienbaum,
88 2014; Malti, Eisenberg, Kim, & Buchmann, 2013). However, the majority of these studies relied
89 on—or at least reported findings from—a single informant using questionnaire items, thus
90 offering only a partial perspective of the development of sympathy across different contexts
91 (e.g., home and school). As a notable exception, Kienbaum (2014) used a multi-method
92 (observations and questionnaires) and multi-informant (caregiver-, teacher-, and self-reports)

93 approach to investigate the development of children's sympathy from 5 to 7 years of age.
94 Correlations between child observations and self-reported sympathy were statistically significant
95 at each of the three time points, whereas the evaluations of teachers and parents were neither
96 associated with each other nor the other methods (correlations ranged from $-.03$ to $.27$).
97 Similarly, Murphy, Shepard, Eisenberg, Fabes, and Guthrie (1999) did not find statistically
98 significant relations between teachers' and parents' evaluations of primary school children's
99 sympathy (the correlation coefficient was $.14$).

100 Several factors might be responsible for this low inter-rater agreement. For instance,
101 caregivers and teachers may perceive children's sympathetic capacities differently based on their
102 shared context with the children, specifically the way in which their respective contexts may
103 differentially set the stage for sympathetic opportunities and ratings. For example, teachers
104 observe children at school amongst a variety of peers (i.e., additional reference points from
105 which to gauge a given child's sympathy), as well as in an environment that generally commands
106 respect for numerous rules. In contrast, caregivers tend to observe their children at home with
107 less reference points (even after considering siblings) and potentially under different sets of rules
108 and expectations. Caregivers may also see their children from a different perspective, given that
109 they are more emotionally involved with the child than the teacher (Funderburk, Eyberg, Rich, &
110 Behar, 2003). Disagreement between informants may also stem from the nature of the construct
111 under investigation and how it is perceived. Sympathy is an internal state that is not easily
112 assessed in children because they may feel concern for another without directly showing it (Stern
113 & Cassidy, 2017). Notably, another important (and less investigated) factor responsible for this
114 disagreement could be the different degree to which the *dispositional characteristics of the child*
115 and *state-like factors* affect the evaluation of each informant. For instance, although caregivers

116 and teachers tend to rate children's behavior and psychological functioning in terms of
117 dispositional (trait) tendencies (e.g., how the child usually behaves or feels; De Los Reyes &
118 Kazdin, 2005), their evaluations can also reflect situational (state) factors. For instance, a teacher
119 may recall a recent event in which a child showed a sympathetic response (e.g., comforting a
120 peer who was teased at school), which may result in an inflated rating of that child's sympathy
121 (compared to his/her general level of sympathy). Therefore, considering that several context- and
122 occasion-specific cues may differently elicit children's sympathy at home (e.g., siblings crying)
123 versus school (e.g., bullying episodes), the disagreement between caregivers and teachers may be
124 further aggravated when the focus of the evaluation (dispositional sympathy versus state
125 sympathy) is not clearly distinguished.

126 In sum, a number of factors may contribute to caregivers and teachers capturing specific
127 aspects of children's sympathy, resulting in difficulties for the interpretation of existing findings,
128 as well as for the integration of information from multiple informants in practical settings (De
129 Los Reyes et al., 2013). Hereafter, we showed how LST theory can shed light on the low cross-
130 informant agreement of children's sympathy by disentangling the level of convergence at both
131 trait level (dispositional sympathy) and state level (momentary manifestation of children's
132 sympathy).

133 **Disentangling Trait and State Agreement using LST**

134 Although a full presentation of LST theory (see Geiser, Bishop, & Lockhart, 2015) is
135 beyond the scope of this paper, we will reference its main assumptions that directly relate to the
136 assessment of trait and state convergence across informants.¹ Developed as an extension of

¹ Throughout this paper, we utilize notations consistent with Geiser et al. (2015).

137 classical test theory, LST theory (Steyer et al., 1992) postulates that an observed, manifest
138 variable (e.g., children's sympathy) can be decomposed into three main components: (1) a trait
139 component ξ that represents the general, stable level of the attribute for that individual, (2) an
140 occasion-specific component ζ that represents state-like deviations from the trait component due
141 to situational and/or interactional (i.e., person x situation) effects, and (3) measurement error.
142 Since, by definition, trait components are stable across time and state components are measured
143 at a specific point in time, only longitudinal data allows for their proper estimation and
144 decomposition (Geiser et al., 2015).

145 For instance, using a structural equation modeling framework, the single-trait-multistate
146 (STMS) model for three observed indicators (e.g., items of a questionnaire) measured at three
147 time points requires the estimation of four latent variables to separate trait and state effects (see
148 Figure 1). First, a common latent trait variable ξ (measured by all nine indicators) is modeled to
149 reflect the general, time-unspecific mean level of the construct under investigation. Importantly,
150 both the factor loading (λ) and intercept (α) of the same item i should be invariant across time to
151 ensure strong (i.e., scalar) longitudinal measurement invariance at the trait-level (i.e., the lack of
152 measurement-related alterations due to different use of the rating scale or interpretations of the
153 items over time; Millsap, 2011; Widaman, Ferrer, & Conger, 2010). Second, three time-specific,
154 latent state residual factors ($\zeta_1, \zeta_2,$ and ζ_3 ; each measured by the three indicators used at each time
155 point) are estimated to capture participants' deviations from the general latent trait. Since latent
156 state residual factors are defined as momentary deviations from the general latent trait, only
157 weak (i.e., metric) longitudinal invariance of factor loadings γ is required (latent state residual
158 factors have a mean of zero by definition).

159 Geiser et al. (2015) extended the STMS model to capture the (in)consistency of trait
160 scores across different fixed situations (e.g., trait anxiety in a neutral versus threatening situation;
161 see Figure 3 on p. 9 of their paper). This revised STMS involves the simultaneous estimation of
162 the same STMS model within each situation (e.g., A and B), thereby allowing the correlation
163 between the resulting latent trait factors ξ_A and ξ_B to be interpreted as an index of the consistency
164 or convergence of the trait scores across the two situations of interest. For our purposes, the
165 revised STMS can also be used to capture (dis)agreement between informants at the trait and
166 state levels. For instance, for caregivers' and teachers' ratings of children's sympathy with a set
167 of items invariant in their content both over time and across informants, the revised STMS
168 allows for the computation of *two relative (rank-order) consistency indexes*: (1) the time-
169 unspecific cross-informant correlation coefficient at the trait level ($\xi_{\text{caregiver}}$ with ξ_{teacher}) with a
170 squared value indicating the degree of cross-informant consistency at the dispositional level of
171 children's sympathy (i.e., both informants rated child A as, *in general*, more sympathetic than
172 child B); (2) the time-specific cross-informant correlation at the state-level ($\zeta_{\text{caregiver}}$ with ζ_{teacher}
173 at time t ; see Figure 2) with a squared value indicating the degree of cross-informant consistency
174 at the momentary, fluctuating level of children's sympathy (i.e., both informants rated child A as
175 more sympathetic than child B *at a specific time point*). Importantly, since latent means are
176 estimated for trait factors, *absolute mean-level differences* across informants in the construct of
177 interest (e.g., trait sympathy) can be also investigated via latent difference score (LDS) models
178 (see de Haan, Prinzie, Sentse, & Jongerling, 2017). The *absolute mean-level differences*
179 represent a further index of (dis)agreement as they indicate to what extent both observers
180 perceive children as having exactly the same mean level of dispositional sympathy (this index is

181 similar to the concept of absolute stability in personality psychology; Santor, Bagby, & Joffe,
182 1997)

183 All these coefficients (dispositional, state, and absolute) reflect distinct indexes of cross-
184 informant (dis)agreement. Failing to distinguish and understand them may lead to misleading
185 interpretations/diagnoses in multi-informant assessment practices (e.g., the ASEBA system;
186 Achenbach & Rescorla, 2001) which, in turn, may affect the selection of appropriate intervention
187 strategies for children.

188 Finally, three other advantages of the LST approach are worthy of mention. First, the
189 STMS model disentangles *true* trait and state components using latent variables (ξ and ζ) that are
190 free of measurement error, which is often considered a serious concern in this area of research
191 (De Los Reyes & Kazdin, 2004). Second, it allows us to ascertain the presence of possible
192 differences between caregivers and teachers in their use of the instruments/ratings of items by
193 testing a series of increasingly restrictive measurement invariance models (i.e., configural,
194 metric, and scalar). Establishing strong (scalar) measurement invariance across informants
195 allows us to interpret cross-informant differences as *true* disagreements rather than as biases due
196 to differential use of the rating scales (de Haan et al., 2017; see also Vanderberg & Lance, 2000).
197 Third, the LDS model allows the inclusion of predictors (e.g., children's gender) to explain
198 mean-level inconsistencies across informants (Geiser et al., 2015; for a more technical
199 introduction to LDS models, see McArdle & Hamagami, 2001).

200 **The Present Study**

201 In sum, existing evidence suggests small and not statistically significant cross-informant
202 agreement in the assessment of children's sympathy, especially when informants (i.e., caregivers
203 and teachers) reported children's sympathy from different contexts (i.e., home versus school;

204 Kienbaum, 2014). However, these studies have failed to separate convergence in evaluations of
205 children's dispositional sympathetic tendencies from convergence in evaluations of the
206 fluctuating components of children's sympathy. Moreover, previous works did not clearly focus
207 on distinguishing between agreement in terms of rank-order consistency (e.g., child A is
208 consistently rated as more sympathetic than child B by both informants) and absolute mean-level
209 agreement (e.g., child A has exactly the same mean level of dispositional sympathy according to
210 both informants).

211 In the present study, we aimed to fill this gap using analyses grounded in LST theory
212 (Steyer et al., 1992) and its conceptual extension for fixed situations (Geiser et al., 2015).
213 Specifically, we investigated the convergence of caregivers' and teachers' evaluations of
214 children's sympathy at the trait and state level from age 6 to 12. We expected a higher degree of
215 rank-order convergence between the evaluations of caregivers and teachers at the stable, trait
216 level of children's sympathy (i.e., in terms of how much the child is sympathetic in general)
217 compared to the ephemeral, state level of their sympathy at each time point. We also modeled
218 absolute mean-level (dis)agreement across caregivers and teachers via LDS analysis. Finally,
219 since previous studies reported girls as more sympathetic than boys (Eisenberg et al., 2015), we
220 explored possible differences in mean-level discrepancies of sympathy between genders.

221 Method

222 Participants

223 For illustrative purposes of the STMS model, we analyzed data published in Zuffianò et
224 al. (2017). Data were from a cohort of 6-year-olds (reassessed at ages 9 and 12) from the Swiss
225 Survey of Children and Youth (COCON), a nationally representative study of social-emotional
226 development. At time 1 (T1), 1,273 children (49% girls; $M_{\text{age}} = 6.17$ years, $SD = 0.22$)

227 participated alongside 1,199 primary caregivers (93% biological mothers) and 870 teachers. At
228 time 2 (T2), 1,101 primary caregivers and 853 teachers provided data, and 1,022 caregivers and
229 734 teachers did so at time 3 (T3).

230 **Measures**

231 **Sympathy.** Caregivers and teachers rated children's sympathy (from 1 = *not at all true* to
232 6 = *always true*) using a widely used scale (Eisenberg et al., 1996). For analytical purposes, we
233 only used the three items of the scale (i.e., "feels sorry for others", "feels sorry for other children
234 who are being teased", and "feels sorry for other children who are sad or upset") that were
235 content-invariant across time points and informants. In addition to allowing for our proposed
236 analyses (which are contingent on content invariance), these items captured the prototypical
237 "feeling sorrow" component that is considered the core of sympathy (Zuffianò et al., 2017).
238 Omega reliability coefficients were .663 (95%CI [.610, .716]) at T1, .800 (95%CI [.767, .833])
239 at T2, and .768 (95%CI [.726, .809]) at T3 for caregiver reports, and .908 (95%CI [.893, .923]) at
240 T1, .924 (95%CI [.909, .940]) at T2, and .919 (95%CI [.903, .935]) at T3 for teacher reports.

241 **Results**

242 **Descriptive Statistics**

243 As reported in Table 1, sympathy scores at the manifest level were always positively and
244 statistically significant correlated. Focusing on cross-informant correlations, caregivers and
245 teachers only showed a small degree of convergence, both concurrently (*rs* ranged from .208 to
246 .254) and over time (*rs* ranged from .134 to .207). As expected, boys were consistently rated as
247 less sympathetic than girls.

248 **STMS Results**

249 First, we estimated an STMS model within each informant and ascertained the tenability

250 of time-invariant factor loadings and intercepts by testing a series of increasingly restrictive
251 measurement invariance assumptions (i.e., configural, metric, and scalar; Vandenberg & Lance,
252 2000). We then compared these nested STMS models using the $\Delta\chi^2$ test. However, because the
253 $\Delta\chi^2$ test is sensitive to sample size, we also considered changes in comparative-fit-index (ΔCFI)
254 lower than .010 as indicative of measurement invariance between these nested models (Cheung
255 & Rensvold, 2002). When equality constraints on factor loadings and item intercepts were not
256 tenable, we tested less restrictive models by relaxing some parameter constraints in order to
257 have, at least, partial scalar invariance (i.e., metric and scalar invariance in at least one item
258 beyond the marker item; Byrne, Shavelson, & Muthén, 1989). Second, we estimated a cross-
259 informant STMS model combining the caregiver- and teacher-reported STMS models to evaluate
260 their degree of convergence at the trait and state level. We also tested cross-informant
261 measurement invariance to ensure that differences in children's sympathy scores from caregivers
262 and teachers reflected true informant-based discrepancies. Finally, we explored possible mean-
263 level differences in children's trait-level sympathy using an LDS model (Geiser et al., 2015;
264 McArdle & Hamagami, 2001).

265 To identify our latent variables, we fixed the factor loading of the marker item to 1 and its
266 intercept to 0. We evaluated model fit according to standard criteria (Kline, 2010). Specifically,
267 we considered CFI and Tucker-Lewis-index (TLI) values $> .90$, and root-mean-square-error-of-
268 approximation (RMSEA) values $< .08$ (with a 90% confidence interval; CI) as indicators of
269 acceptable model fit (Kline, 2010). We ran our analyses in *Mplus* 8 (Muthén & Muthén, 1998–
270 2017) and we accounted for missing data with full information maximum-likelihood estimation
271 of the parameters (MLR).²

² With MLR estimation, the formula for $\Delta\chi^2$ also includes the scaling correction factor (*scf*).

272 **Caregiver reports.** As reported in Table 2, we established longitudinal partial scalar
273 invariance for the STMS model according to the Δ CFI criterion. Only the factor loading (at the
274 trait level) and intercept of the item “feels sorry for other children who are sad or upset” were
275 relaxed to be different at T1. Interestingly, squared standardized loadings (see Table 3) indicated
276 that approximately 23% to 38% of the variance of the items stemmed from trait-level variability
277 (average trait consistency coefficient \approx 31%) whereas only 16% to 24% reflected state-level
278 variability (average occasion-specificity coefficient \approx 20%; see Geiser, Keller, & Lockhart,
279 2013; Geiser, Hintz, Burns, & Servera, 2017). Hence, although a large part of the variability of
280 the items was unexplained by the STMS model, caregiver reports mostly captured children’s trait
281 sympathetic tendencies rather than their occasion-specific, sympathetic manifestations.

282 **Teacher reports.** We established full longitudinal scalar invariance for the STMS model
283 involving teacher reports of children’s sympathy, as the Δ CFI was lower than .01 at each step of
284 the measurement invariance analysis (see Table 2). Unlike caregiver reports (see Table 3),
285 squared standardized loadings of the items indicated that teachers mostly captured children’s
286 sympathy at the state level (variance ranging from 50% to 63%, average occasion-specificity
287 coefficient \approx 54%) rather than at the trait level (variance ranging from 23% to 27%, average trait
288 consistency coefficient \approx 25%).

289 **Cross-informant STMS.** The STMS model with partial scalar invariance across
290 informants³ (the factor loading and intercept of the caregiver-reported item “feels sorry for other
291 children who are sad or upset” were not constrained to equality) showed a good fit to the data, χ^2
292 (155) = 309.825, $scf = 1.112$, $p < .001$, CFI = .973, TLI = .974, RMSEA = .028, 90% CI [.024,

³ In this STMS model, we also constrained the covariances of the residual latent state factors over time to equality ($\zeta_{\text{caregiver}}$ with ζ_{teacher} at T1 = $\zeta_{\text{caregiver}}$ with ζ_{teacher} at T2 = $\zeta_{\text{caregiver}}$ with ζ_{teacher} at T3). The Mplus syntax for this model is reported in the Online Appendix.

293 .033], and was not statistically different ($\Delta\chi^2(4) = 2.993, p = .559; \Delta CFI = .000$) from the partial
294 metric invariance model, $\chi^2(151) = 307.041, scf = 1.111, p < .001, CFI = .973, TLI = .973,$
295 $RMSEA = .029, 90\% CI [.024, .033]$. This latter, in turn, did not worsen the fit of the configural
296 model ($\Delta\chi^2(4) = 4.928, p = .295; \Delta CFI = .000$). Hence, children's sympathy scores could be
297 meaningfully compared across caregivers and teachers. As expected (see Figure 2), caregivers
298 and teachers showed a different degree of rank-order convergence when children's sympathetic
299 scores were disentangled at the trait and state levels. Specifically, caregivers and teachers
300 reported a higher degree of cross-informant consistency at children's trait level of sympathy ($r =$
301 $.510, 95\% CI [.468, .549], p < .001$), compared to their state level ($r = .123, 95\% CI [.069, .177],$
302 $p = .002$ at each time point), with cross-informant agreements of 26% and 2%, respectively.

303 The presence of partial scalar invariance also allowed us to investigate absolute mean-
304 level (dis)agreement across informants. Overall, caregivers (mean $\xi_{\text{parent}} = 5.205, 95\% CI [5.166,$
305 $5.243]$) rated their children as more sympathetic than teachers did (mean $\xi_{\text{teacher}} = 4.906, 95\% CI$
306 $[4.853, 4.959]$). Constraining the two latent trait means to be equal across informants ($\chi^2(156) =$
307 $421.495, scf = 1.114, p < .001, CFI = .954, TLI = .955, RMSEA = .037, 90\% CI [.032, .041]$)
308 worsened the model fit of the partial scalar STMS model ($\Delta\chi^2(1) = 125.445, p < .001; \Delta CFI =$
309 $.020$), thereby revealing statistically significant differences at the mean-level perceptions of
310 children's sympathy across informants. Hence, although parents and teachers showed a
311 moderately high degree of convergence in ranking children relative to their peers based on their
312 dispositional sympathy (e.g., both rated child A as generally more sympathetic than child B),
313 they showed significant differences in capturing the exact mean level of each child's sympathy
314 (e.g., caregiver ratings of children A and B could be 4.3 and 3.8, respectively, whereas teacher
315 ratings of the same children could be 3.9 and 3.2, respectively).

316 To further investigate these absolute mean-level differences at the trait level, we used a
317 LDS analysis (de Haan et al., 2017; Geiser et al., 2015) in which we estimated a second-order
318 latent difference factor (Δf) representing the difference between teachers and caregivers (ξ_{teacher}
319 $- \xi_{\text{caregiver}}$). In the LDS model, $\chi^2(155) = 309.825$, $scf = 1.112$, $p < .001$, CFI = .973, TLI = .974,
320 RMSEA = .028, 90% CI [.024, .033], the mean ($-.299$, $p < .001$) of Δf was statistically
321 significant, indicating, on average, a lower mean value of teacher-reported sympathy compared
322 to caregiver-reported sympathy. In detail, using Cohen's guidelines (1988), the latent mean-level
323 difference between caregivers and teachers could be interpreted as a medium effect (Cohen's $d =$
324 $-.561$, 95% CI $[-.641, -.481]$).⁴ The variance of Δf was also statistically different from zero ($.291$,
325 $p < .001$), highlighting significant inter-individual differences (i.e., caregivers and teachers
326 perceived some children as more different than others). A final conditional LDS model, $\chi^2(171)$
327 $= 347.473$, $scf = 1.112$, $p < .001$, CFI = .971, TLI = .971, RMSEA = .029, 90% CI [.024, .033],
328 revealed that children's gender (girls = 0, boys = 1) predicted the Δf ($\beta = -.490$, $p < .001$, 95% CI
329 $[-.574, -.406]$), suggesting that discrepancies between teachers and caregivers ($\xi_{\text{teacher}} - \xi_{\text{caregiver}}$)
330 were stronger for boys than girls. Specifically, compared to girls, teachers rated boys largely
331 lower than caregivers did (Cohen's $d = -1.125$, 95% CI $[-1.244, -1.006]$).

332 Discussion

333 Understanding the nature of informant discrepancies has attracted the attention of many
334 psychological researchers. This is because this diagnostic information yields potentially
335 important implications when making decisions regarding the selection and implementation of
336 intervention practices aimed at enhancing children's social-emotional development and

⁴ A latent mean score of zero of the Δf would have meant perfect, absolute mean-level agreement between caregivers and teachers in evaluating children's dispositional sympathy.

337 wellbeing. According to meta-analytic findings, only a small-to-moderate degree of convergence
338 ($r = .28$; Achenbach, McConaughy, & Howell, 1987) exists between different types of reporters,
339 such as caregivers and teachers, and this weak agreement tends to be even lower for less
340 observable constructs, such as children's internal affective responses ($r = .21$; De Los Reyes et
341 al., 2015). Hence, prominent developmental and clinical psychologists have emphasized the
342 importance of a multi-informant approach to social-emotional and behavioral assessment
343 because situation-specific effects may reveal meaningful variability in such constructs across
344 contexts (e.g., home versus school; Achenbach et al., 1987; De Los Reyes et al., 2015).

345 In the present study, we highlighted how recent conceptualizations of LST theory (Gesier
346 et al., 2015) can inform children's multi-informant assessment by clearly indicating the level of
347 analysis at which (dis)agreement between informants occurs. We showed that when trait-and
348 state-level variability are distinguished within each informant, two types of relative (rank-order)
349 consistency coefficients can be computed to reflect inter-rater agreement: (1) the trait
350 consistency coefficient (i.e., time-unspecific cross-informant agreement at the trait level of the
351 psychological attribute) and (2) the occasion-specific consistency coefficient (i.e., time-specific
352 cross-informant agreement at the state level of the psychological attribute). To illustrate the
353 advantages of separating these two indexes, we examined the level of (dis)agreement between
354 caregivers and teachers in the evaluation of children's sympathy.

355 At the manifest level, we found that correlations of children's sympathy across
356 informants were low (r s ranging from .13 to .25), reflecting a small amount of agreement
357 between caregivers and teachers. This aligns with previous findings reporting only a small
358 degree of convergence between caregivers and teachers in the assessment of children's sympathy
359 (Kienbaum, 2014). This overall small effect could lead researchers to conclude that only minimal

360 agreement exists between caregivers and teachers and, therefore, that children's sympathetic
361 responses are highly variable across contexts. As a consequence, this high discrepancy may
362 create problems in properly identifying children who may benefit from timely social-emotional
363 interventions to promote their sympathy (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger,
364 2011; Malti, Chaparro, et al., 2016).

365 Yet, our LST analysis revealed a more complex picture of cross-informant convergence.
366 First, by establishing cross-informant measurement invariance (at the partial scalar level; Byrne
367 et al., 1989), we were able to confidently interpret the relations between caregivers' and
368 teachers' evaluations as reflecting true (dis)agreement rather than methodological biases in their
369 use of the scale. Second, we found that teachers' and caregivers' scores were differentially
370 affected by occasional manifestations of children's sympathy: although both caregivers and
371 teachers attributed a consistent amount of children's sympathetic responses to their dispositional,
372 trait-like characteristics, teachers were more likely than caregivers to capture situational, state-
373 like manifestations of children's sympathy. This difference could be also due to the fact that
374 teachers were different across time (whereas caregivers, mostly mothers, did not change over the
375 duration of the study). Third, cross-informant convergence was different when children's
376 sympathy scores were decomposed into trait and state components. As expected, caregivers and
377 teachers showed moderately high agreement ($r = .510$) in their ratings of children's dispositional
378 tendency to feel sympathetic concern, yet fairly low agreement in their ratings of children's
379 momentary manifestations of sympathy at each time point ($r = .123$). Thus, differently from the
380 correlational results at the manifest level, we found that caregivers and teachers *did* agree in
381 terms of identifying children who were, *in general*, more sympathetic than others. Although this
382 result could be interpreted as further evidence of the relative stability (and visibility) of

383 psychological traits across contexts (e.g., Church et al., 2008), it may also *indirectly* reveal
384 information about the inter-rater agreement concerning the causes of children’s emotional
385 responses. According to the Attribution Bias Context Model (ABC; De Los Reyes & Kazdin,
386 2005), the considerable cross-informant consistency at the trait-level could be related to the fact
387 that informants—such as caregivers and teachers—tend to interpret children’s social-emotional
388 development and behaviors in terms of dispositional tendencies (i.e., child A *is more sympathetic*
389 *in general* than child B; De Los Reyes & Kazdin, 2005). In line with this claim, our LST analysis
390 indicated that both caregivers and teachers captured a considerable portion of the dispositional
391 nature of children’s sympathy (although teacher ratings were more state- than trait-sensitive).
392 Hence, properly isolating agreement at the level at which both informants most attribute the
393 causes of children’s psychological functioning (i.e., the dispositional level) can thus result in
394 relatively high convergence between them, even for a less manifest emotional response like
395 sympathy and for caregivers and teachers who report from different contexts of observation.
396 Interestingly, teachers and caregivers also showed a small, nearly negligible amount of
397 agreement at the state level, reflecting the fluctuating, momentary deviations of children’s
398 sympathy from their general disposition. Hence, situational positive (or negative) spikes in
399 sympathy seemed to have some marginal, time-specific consistency across contexts, which
400 jointly affected caregiver and teacher reports of children’s sympathy at each time point.

401 Although teachers and caregivers generally agreed in terms of identifying children who
402 were more sympathetic than others, we also found that they moderately disagreed regarding the
403 exact, “true” mean level of each child’s dispositional sympathy. Specifically, teacher-reported
404 latent scores were consistently lower than caregiver-reported latent scores. This may be because
405 sympathy is not a highly visible emotional state at school. A child can feel concern for his/her

406 classmates without displaying an obvious emotional response or engaging in immediate
407 prosocial actions that can be clearly seen by the teacher (who is also responsible for numerous
408 other students). From this perspective, parents have the benefit of one-on-one time that increases
409 the chances of gaining insight into their child's sympathetic tendencies. In line with Funderburk
410 et al. (2003), it may also be the case that caregiver ratings are more positive than teacher ratings
411 because of the strong emotional bond underlying the parent-child relationship. Moreover,
412 caregivers and teachers may rely on different cues: they report from different contexts of
413 observation characterized by distinct relationships and opportunities for social interaction which,
414 in the end, provide them with different reference points to calibrate their assessments of
415 children's sympathy (e.g., interactions with siblings versus classmates). Realistically, the
416 abovementioned factors could be jointly responsible for the overall lower dispositional scores of
417 children's sympathy reported by teachers versus caregivers.

418 Finally, we modeled and explained mean-level discrepancies at the trait level using a
419 LDS framework (Geiser et al., 2015) and found systematic, statistically significant variability in
420 how much children were rated lower in sympathy by teachers versus caregivers. Moreover, this
421 variability was predicted by children's gender, such that boys' evaluations were consistently
422 more discrepant (i.e., they were lower in teacher- versus caregiver-reported dispositional
423 sympathy). This finding may stem from gender-typed socialization practices, which could
424 predispose boys to show less sympathy (especially at school where they interact—or at least
425 have the opportunity to interact—more heavily with other peers and adults), thereby reinforcing
426 teachers' stereotypical view of boys as much less sympathetic than girls (Chaplin & Aldao,
427 2013). In addition, boys may express their sympathetic concern in qualitatively different ways
428 from girls (e.g., via nonverbal behaviors such as patting on the shoulder), which might not be

429 easily captured by teachers in the classroom context. Hence, more work is needed to develop
430 social-emotional instruments that include a variety of indicators that tap into both verbal and
431 nonverbal aspects of sympathy-related responding.

432 **Limitations**

433 Despite its strengths, our current approach also has some limitations rooted in LST
434 theory/methodology that may hinder its use for understanding informant discrepancies. First, the
435 STMS requires the use of valid questionnaires that include content-invariant items across raters
436 to establish cross-informant measurement invariance. Although there are some valid multi-
437 informant assessment tools (e.g., The “Child Behavior Checklist”; Achenbach & Rescorla, 2001;
438 the “Strengths and Difficulties Questionnaire”; Goodman, 1997), numerous questionnaires used
439 in the literature have been developed to capture the perspective of a specific informant (e.g., self
440 reports for self-efficacy scales), potentially limiting the use of our current approach for these
441 constructs. Second, because some psychological attributes are more state-like than trait-like by
442 nature (e.g., happiness), researchers should carefully plan appropriate time lags across
443 measurement points to properly model trait and state variability (and to measure associated
444 cross-informant convergence). Third, directly related to the previous point, the STMS assumes
445 the presence of longitudinal data (Geiser et al., 2015), which, very often, is not feasible for
446 several reasons (e.g., time constraints, costs, etc.). Thus, in the absence of longitudinal data, we
447 advise making the level of analysis at which raters should focus their evaluations clear to them
448 (i.e., in the instructions for a particular questionnaire, specify if the rater should focus on how the
449 child generally feels/behaves versus how the child felt/behaved in the last day[s], week[s], or
450 month[s]), thereby increasing the likelihood of convergence between different informants using
451 the scale.

452 Conclusions and Future Directions

453 Although different informants likely capture unique and diverse aspects of children's
454 social-emotional functioning, the extent of their disagreement might be erroneously exacerbated
455 by a mismatch or confusion regarding the level (i.e., dispositional versus situational) at which
456 their assessments are focused. In the present study, we used LST analysis to disentangle these
457 two levels of analysis and we showed how teachers and caregivers had a moderately high degree
458 of convergence in how they evaluated children's dispositional sympathetic tendencies (which is
459 perhaps even more surprising given that sympathy is a relatively difficult internal process to
460 observe). We also highlighted the importance of considering absolute, mean levels of cross-
461 informant (dis)agreement and gender differences thereof.

462 Finally, our findings may also offer some suggestions to help researchers develop better
463 tools to assess essential dimensions of social-emotional functioning in childhood across different
464 informants and contexts. For instance, future multi-informant assessments may benefit from
465 including ad-hoc open questions designed to capture important events (e.g., a specific
466 sympathetic response or related behavior observed) that could account for occasion-specific
467 cross-informant agreement. Moreover, future scales should clearly list the different reference
468 points that can be used to compare children on the basis of psychological functioning (e.g.,
469 siblings, classmates, peers in general, etc.) in order to ease the convergence across informants,
470 especially when they report from different contexts of observation (e.g., home versus school).

471

472

473

474

475

476 **ETHICS STATEMENT**

477 The current study was conducted in Switzerland and consisted of non-invasive and
478 unconstrained parent and teacher questionnaires. According to the regulations in the canton of
479 Zurich in Switzerland (the so-called “Regulations of the Ethics Commission for Psychological
480 Research”, 2011), there was no requirement for an ethics committee approval when the study
481 was conducted. According to this regulation (Article 5, paragraph 1), this study was exempted
482 from requiring formal ethical approval. The study fully complies with the ethics guidelines given
483 by this legal regulation (see Article 8, paragraph 2). The regulation is based on the “Ethical
484 Principals of Psychologists and Code of Conduct” (as outlined in the so-called “Ethical
485 Guidelines for Psychologists of the Swiss Society for Psychology, as amended on October 13,
486 2003) and the ethical standards of the American Psychological Association (APA). Written and
487 informed consent was obtained from all research participants and from the parents / legal
488 guardians of all non-adult participants. The data were analyzed anonymously.

489

490

491 **References**

- 492 Achenbach, T.M., McConaughy, S.H., & Howell, C.T. (1987). Child/adolescent behavioral and
493 emotional problems: implications of cross-informant correlations for situational
494 specificity. *Psychology Bulletin*, *101*, 213-232. doi:10.1037/0033-2909.101.2.213
- 495 Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms &*
496 *Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, &
497 Families.
- 498 Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor
499 covariance and mean structures: The issue of partial measurement invariance.
500 *Psychological Bulletin*, *105*, 456–466. doi:10.1037/0033-2909.105.3.456
- 501 Chaplin, T. M., & Aldao, A. (2013). Gender differences in emotion expression in children: A
502 meta-analytic review. *Psychological Bulletin*, *139*, 735-765. doi:10.1037/a0030737
- 503 Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indices for testing
504 measurement invariance. *Structural Equation Modeling*, *9*, 233–255.
505 doi:10.1207/S15328007SEM0902_5
- 506 Church, A. T., Katigbak, M. S., Reyes, J. A. S., Salanga, M. G. C., Miramontes, L. A., & Adams,
507 N. B. (2008). Prediction and cross-situational consistency of daily behavior across
508 cultures: Testing trait and cultural psychology perspectives. *Journal of Research in*
509 *Personality*, *42*, 1199-1215. doi:10.1016/j.jrp.2008.03.007
- 510 Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ:
511 Lawrence Erlbaum Associates.

- 512 de Haan, A., Prinzie, P., Sentse, M., & Jongerling, J. (2017). Latent Difference Score modeling:
513 A flexible approach for studying informant discrepancies. *Psychological Assessment*.
514 Advance online publication. doi:10.1037/pas0000480
- 515 De Los Reyes, A., Augenstein, T. M., Aldao, A., Thomas, S. A., Daruwala, S. E., Kline, K., &
516 Regan, T. (2015). Implementing psychophysiology in clinical assessments of adolescent
517 social anxiety: Use of rater judgments based on graphical representations of
518 psychophysiology. *Journal of Clinical Child & Adolescent Psychology*, *44*, 264–279.
519 doi:10.1080/15374416.2013.859080
- 520 De Los Reyes, A., & Kazdin, A. E. (2004). Measuring informant discrepancies in clinical child
521 research. *Psychological Assessment*, *16*, 330–334. doi:10.1037/1040-3590.16.3.330
- 522 De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of
523 childhood psychopathology: A critical review, theoretical framework, and
524 recommendations for further study. *Psychological Bulletin*, *131*, 483–509.
525 doi:10.1037/0033-2909.131.4.483
- 526 De Los Reyes, A., Thomas, S. A., Goodman, K. L., & Kundey, S. M. A. (2013). Principles
527 Underlying the Use of Multiple Informants' Reports. *Annual Review of Clinical*
528 *Psychology*, *9*, 123–149. doi:10.1146/annurev-clinpsy-050212-185617
- 529 Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The
530 impact of enhancing students' social and emotional learning: A meta-analysis of school-
531 based universal interventions. *Child Development*, *82*, 405–432. doi:10.1111/j.1467-
532 8624.2010.01564.x.
- 533 Eisenberg, N., Fabes, R. A., Murphy, B., Karbon, M., Smith, M., & Maszk, P. (1996). The
534 relations of children's dispositional empathy-related responding to their emotionality,

- 535 regulation, and social functioning. *Developmental Psychology*, 32, 195–209.
536 doi:10.1037/0012-1649.32.2.195
- 537 Eisenberg, N., Spinrad, T. L., & Knafo, A. (2015). Prosocial development. In M. E. Lamb & R.
538 M. Lerner (Eds.), *Handbook of child psychology and developmental science, Vol. 3:*
539 *Social, emotional, and personality development* (7th ed., pp. 610–656). New York, NY:
540 Wiley.
- 541 Eisenberg, N., Spinrad, T., & Morris, A. (2014). Empathy-related responding in children. In M.
542 Killen & J. Smetana (Eds.), *Handbook of moral development* (2nd ed., pp. 184–207).
543 New York, NY: Psychology Press.
- 544 Funderburk, B. W., Eyberg, S. M., Rich, B. A., & Behar, L. (2003). Further psychometric
545 evaluation of the Eyberg and Behar rating scales for parents and teachers of preschoolers.
546 *Early Education and Development*, 14, 67–81.
- 547 Geiser, C., Bishop, J., & Lockhart, G. (2015). Collapsing factors in multitrait-multimethod
548 models: Examining consequences of a mismatch between measurement design and
549 model. *Frontiers in Psychology: Quantitative Psychology and Measurement*, 6, 946.
550 doi:10.3389/fpsyg.2015.00946
- 551 Geiser, C., Hintz, F., Burns, G., Servera, M. (2017). Latent variable modeling of person-situation
552 data. In Rauthmann, J. F., Sherman, R., Funder, D. C. (Eds.), *The Oxford Handbook of*
553 *Psychological Situations* (pp. 1-57). Oxford University Press.
- 554 Geiser, C., Keller, B. T., & Lockhart, G. (2013). First- versus second-order latent growth curve
555 models: Some insights from latent state-trait theory. *Structural Equation Modeling*, 20,
556 479–503. doi:10.1080/10705511.2013.797832

- 557 Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of*
558 *Child Psychology and Psychiatry*, 38(5), 581-586. doi:10.1111/j.1469-
559 7610.1997.tb01545.x
- 560 Kienbaum, J. (2014). The development of sympathy from 5 to 7 years: Increase, decline or
561 stability? A longitudinal study. *Frontiers in Psychology*, 5, 468.
562 doi:10.3389/fpsyg.2014.00468
- 563 Kimonis E. R., Frick P. J., Muñoz L. C., & Aucoin K. J. (2008). Callous-unemotional traits and
564 the emotional processing of distress cues in detained boys: testing the moderating role of
565 aggression, exposure to community violence, and histories of abuse. *Developmental*
566 *Psychopathology*, 20, 569–589. doi:10.1017/S095457940800028X
- 567 Kline, R. B. (2010). *Principles and practices of structural equation modeling* (3rd ed.). New
568 York, NY: Guilford.
- 569 Malti, T., Chaparro, M. P., Zuffianò, A., & Colasante, T. (2016). School-based interventions to
570 promote empathy-related responding in children and adolescents: A developmental
571 analysis. *Journal of Clinical Child and Adolescent Psychology*, 45(6), 718–731.
572 doi:10.1080/15374416.2015.1121822
- 573 Malti, T., Eisenberg, N., Kim, H., & Buchmann, M. (2013). Developmental trajectories of
574 sympathy, moral emotion attributions, and moral reasoning: The role of parental support.
575 *Social Development*, 22, 773–793. doi:10.1111/sode.12031
- 576 Malti, T., Sette, S., & Dys, S.P. (2016). Socioemotional responding: A perspective from
577 Developmental Psychology. In R. Scott, M. Buchmann, & S. Kosslyn (Eds.), *Emerging*
578 *Trends in the Social and Behavioral Sciences*. Hoboken, NJ: John Wiley and Sons.
579 doi:10.1002/9781118900772.etrds0415

- 580 Malti, T., & Song, J. -H. (in press). Social-emotional development and aggression. In T. Malti &
581 K. H. Rubin (Eds.), *Handbook of child and adolescent aggression: Emergence,*
582 *development, and intervention.* New York, NY: Guilford Press.
- 583 McArdle, J.J., & Hamagami, F. (2001). Latent difference score structural models for linear
584 dynamic analyses with incomplete longitudinal data. In: Collins, L.M., & Sayer, A.G.
585 (Eds). *New methods for the analysis of change* (pp. 139-175). Washington, DC: APA
586 Press.
- 587 Millsap, R. E. (2011). *Statistical approaches to measurement invariance.* New York, NY: Taylor
588 and Francis Group.
- 589 Murphy, B. C., Shepard, S. A., Eisenberg, N., Fabes, R. A., & Guthrie, I. K. (1999).
590 Contemporaneous and longitudinal relations of young adolescents' dispositional
591 sympathy to their emotionality, regulation, and social functioning. *Journal of Early*
592 *Adolescence, 19,* 66–97.
- 593 Muthén, L.K., & Muthén, B.O. (1998-2017). *Mplus user's guide, 8th Edn.* Los Angeles, CA:
594 Muthén & Muthén.
- 595 Santor, D. A., Bagby, R. M., & Joffe, R. T. (1997). Evaluating stability and change in personality
596 and depression. *Journal of personality and social psychology, 73,* 1354-1362.
- 597 Stern, J.A., & Cassidy, J. (2017). Empathy from infancy to adolescence: An attachment
598 perspective on the development of individual differences. *Developmental Review.*
599 Advance online publication. doi:10.1016/j.dr.2017.09.002
- 600 Steyer, R., Ferring, D. & Schmitt, M. J. (1992). States and traits in psychological
601 assessment. *European Journal of Psychological Assessment, 8,* 79–98.

- 602 Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance
603 literature: suggestions, practices, and recommendations for organizational
604 research. *Organizational Research Methods*,*3*, 4–70. doi:10.1177/109442810031002
- 605 Widaman, K. F., & Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal
606 structural equation models: Measuring the same construct across time. *Child*
607 *Development Perspectives*,*4*, 10–18. doi:10.1111/j.1750-8606.2009.00110.x
- 608 Zuffianò, A., Colasante, T., Buchmann, M., & Malti, T. (2017). The co-development of
609 sympathy and overt aggression from middle childhood to early adolescence.
610 *Developmental Psychology*. Advance online publication. doi:10.1037/dev0000417

Table 1

Correlations, Means, and Standard Deviations (SD) of Sympathy

	Mean (SD)	1	2	3	4	5	6	7
1. Sex	– (–)	–						
2. Sympathy_T1 (Ca)	5.117 (0.772)	-.165	–					
3. Sympathy_T2 (Ca)	5.076 (0.906)	-.158	.420	–				
4. Sympathy_T3 (Ca)	5.067 (0.870)	-.175	.384	.505	–			
5. Sympathy_T1 (Te)	4.914 (1.047)	-.262	.208	.201	.134	–		
6. Sympathy_T2 (Te)	4.737 (1.167)	-.337	.187	.254	.176	.288	–	
7. Sympathy_T3 (Te)	4.620 (1.113)	-.324	.177	.207	.225	.174	.383	–

Note. Sex (boys = 1, girls = 0). Ca = caregiver report. Te = teacher report. Teachers and caregivers rated sympathy on a 6-point scale from 1 to 6. All correlation coefficients were statistically significant at $p < .001$.

Table 2

Measurement Invariance

	χ^2	<i>df</i>	<i>scf</i>	χ^2/df	<i>p</i>	CFI	TLI	RMSEA (90%CI)	MC	$\Delta\chi^2$	Δdf	<i>p</i>	ΔCFI
<i>Sympathy (Ca)</i>													
1. Configural	83.156	22	1.191	3.780	<.001	.960	.934	.047 (.037, .058)					
2. Metric partial	99.426	29	1.217	3.428	<.001	.954	.942	.044 (.035, .054)	2vs.1	16.913	7	.018	.006
3. Scalar partial	112.765	34	1.228	3.317	<.001	.948	.981	.043 (.035, .052)	3vs.2	13.520	5	.019	.006
<i>Sympathy (Te)</i>													
4. Configural	79.090	22	0.984	3.595	<.001	.982	.971	.048 (.036, .042)					
5. Metric	94.331	30	1.026	3.144	<.001	.980	.976	.044 (.034, .054)	4 vs.5	16.606	8	.034	.002
6. Scalar	111.668	36	1.024	3.102	<.001	.977	.977	.043 (.034, .052)	5 vs.6	17.320	6	.008	.003

Note. In addition to the χ^2 , the following fit indexes are reported: Comparative-fit-index (CFI); Tucker-Lewis-index (TLI), Root-mean-square-error-of-approximation (RMSEA) with 90% confidence intervals (CI). Ca = Caregiver; Te = Teacher; *df* = degrees of freedom; *scf* = scaling correction factor; MC = model comparison

Table 3

Factor Loadings, Intercepts, and Variances from Final STMS Models

<i>He/She usually:</i>		Caregivers			Teachers		
		λ	γ	α	λ	γ	α
T1	feels sorry for others	1.000 (0.617)	1.000 (0.462)	0.000	1.000 (0.507)	1.000 (0.734)	0.000
	feels sorry for other children who are being teased	1.131 (0.474)	1.290 (0.405)	-0.902	1.120 (0.518)	1.070 (0.716)	-0.873
	feels sorry for other children who are sad or upset	0.926 (0.506)	1.194 (0.488)	0.353	1.037 (0.475)	1.144 (0.759)	-0.360
T2	feels sorry for others	1.000 (0.545)	1.000 (0.408)	0.000	1.000 (0.486)	1.000 (0.704)	0.000
	feels sorry for other children who are being teased	1.131 (0.572)	1.290 (0.488)	-0.902	1.120 (0.512)	1.070 (0.708)	-0.873
	feels sorry for other children who are sad or upset	1.170 (0.596)	1.194 (0.455)	-1.069	1.037 (0.495)	1.144 (0.791)	-0.360
T3	feels sorry for others	1.000 (0.552)	1.000 (0.413)	0.000	1.000 (0.498)	1.000 (0.721)	0.000
	feels sorry for other children who are being teased	1.131 (0.561)	1.290 (0.479)	-0.902	1.120 (0.512)	1.070 (0.708)	-0.873
	feels sorry for other children who are sad or upset	1.170 (0.581)	1.194 (0.443)	-1.069	1.037 (0.488)	1.144 (0.779)	-0.360
<i>Variances</i>							
	Trait variability (ξ)	.284	$p < .001$.325	$p < .001$	
	State variability (ζ_1)	.159	$p < .001$.681	$p < .001$	
	State variability (ζ_2)	.159	$p < .001$.681	$p < .001$	
	State variability (ζ_3)	.159	$p < .001$.681	$p < .001$	

Note. Item intercepts (α), unstandardized factor loadings, and standardized factor loadings (in parentheses) for sympathy at both trait level (λ) and state level (γ) are reported. All factor loadings (λ and γ) were statistically significant at $p < .001$. Time 1 = T1; Time 2 = T2; Time 3 = T3.

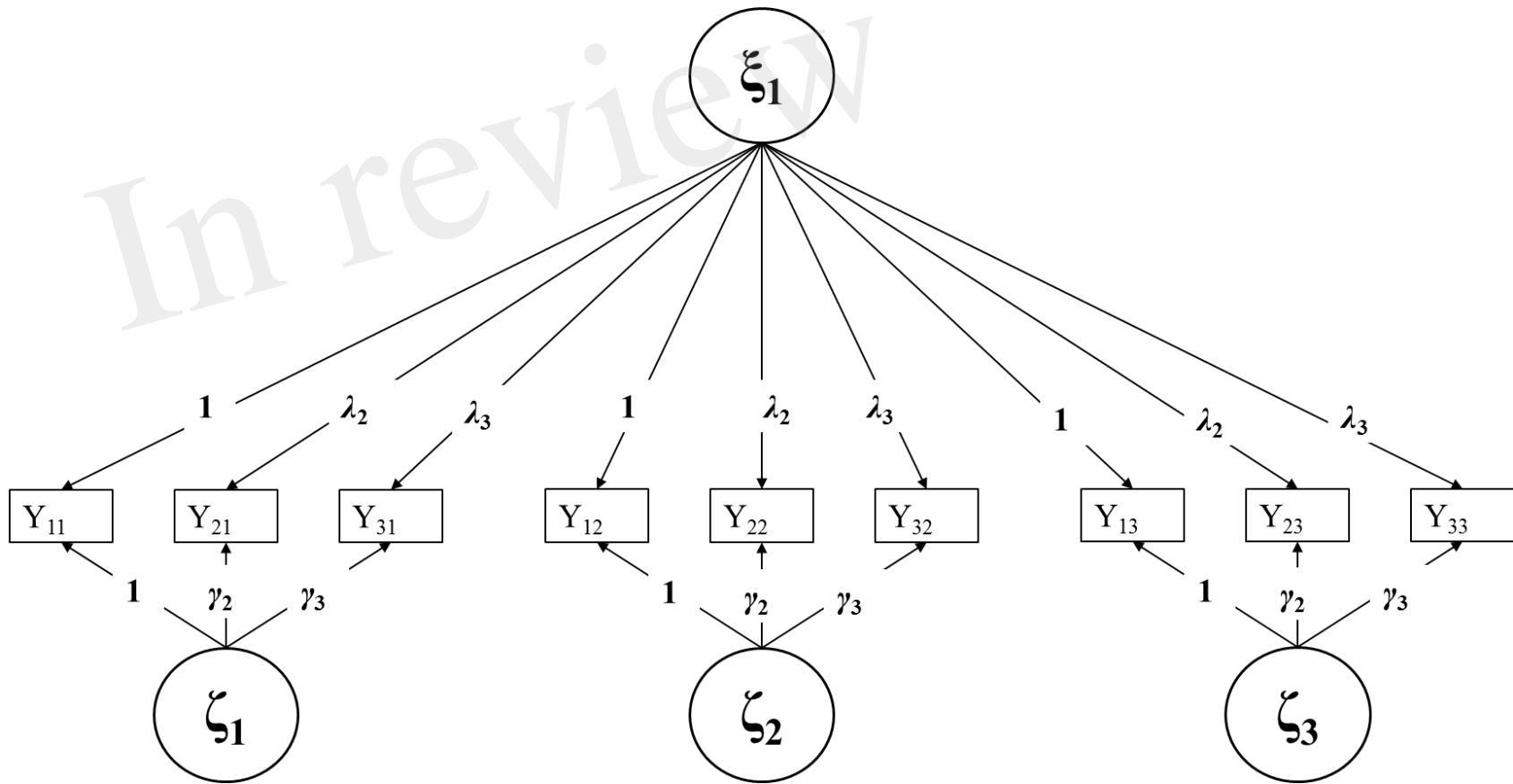


Figure 1. Singletrait-multistate (STMS) Model for Three Waves.

Note. Latent variables indicate both trait (ξ) and state (ζ) components. For the sake of simplicity, the mean-structure (i.e., intercepts) of the model is not depicted.

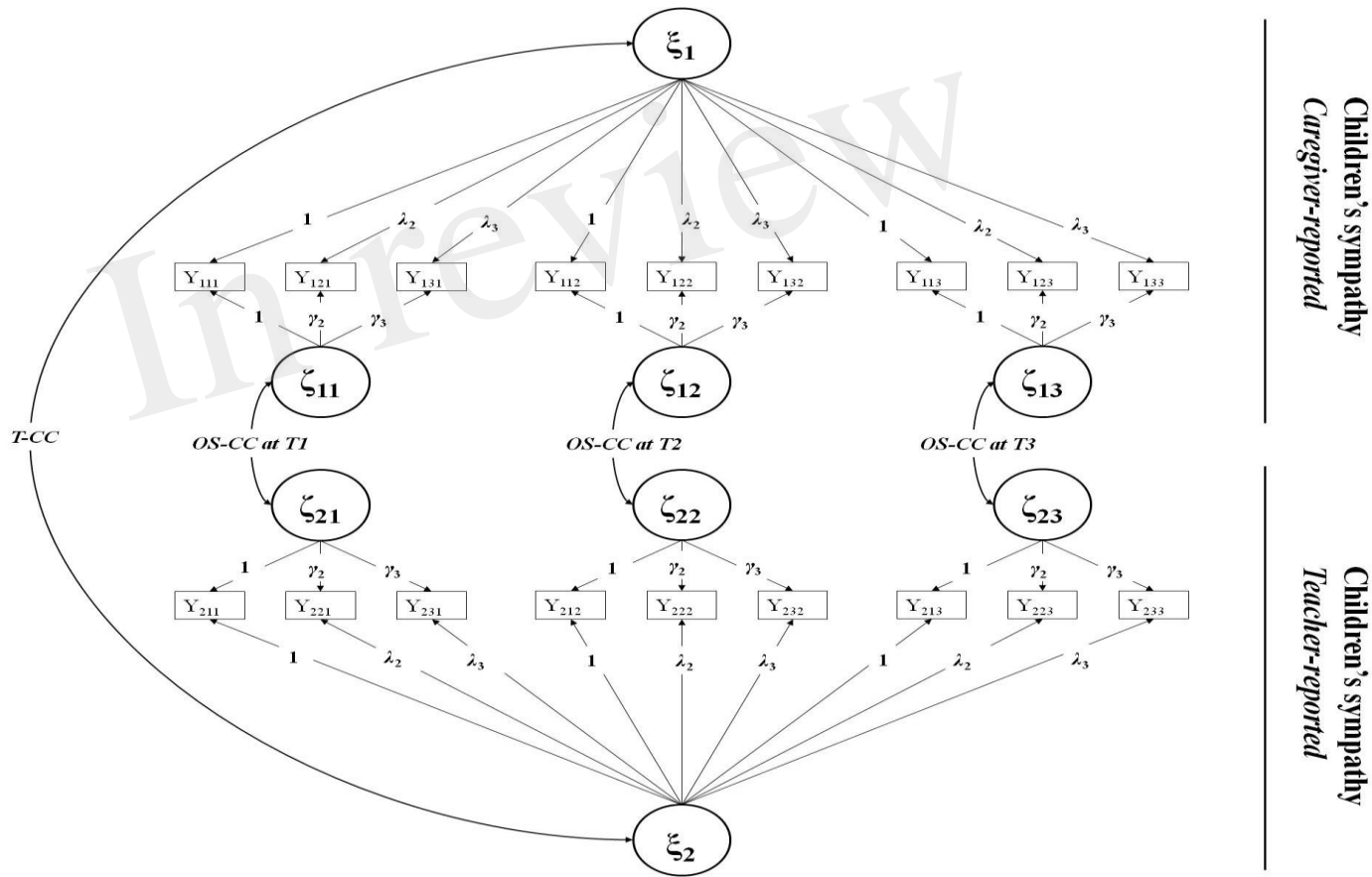


Figure 2. Combined Singletrait-multistate (STMS) Model for Three Waves and Two Informants.

Note. Latent variables indicate both trait (ξ) and state (ζ) components for each informant. Cross-informant trait consistency coefficient ($T-CC$) and cross-informant occasion-specific consistency coefficients ($OS-CC$) are reported. For the sake of simplicity, the mean-structure (i.e., intercepts) of the model is not depicted.

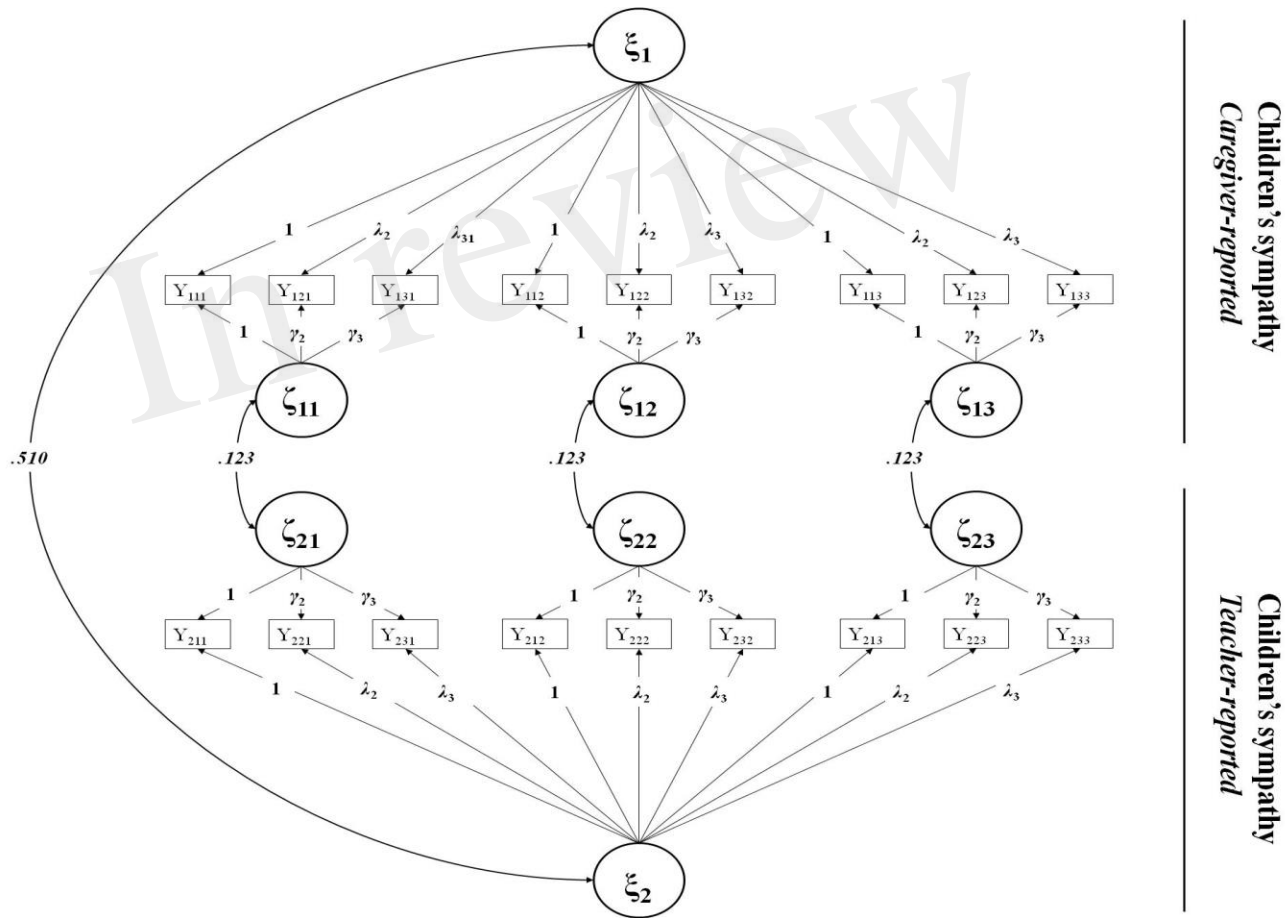


Figure 3. Combined Singletrait-multistate (STMS) Model of Children's Sympathy across Caregivers and Teachers.

Note. Latent variables indicate both trait (ξ) and state (ζ) components for each informant. Cross-informant trait consistency coefficient and cross-informant occasion-specific consistency coefficients were statistically significant ($p < .01$).