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

Antonio Zuffiano[1*], Stefania Sette[2], Tyler Colasante[3], Marlis Buchmann[4], Tina Malti[3]

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

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

Antonio Zuffianò[1], Stefania Sette[2], Tyler Colasante[3], Marlis Buchmann[4], 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.

Abstract

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.

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

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

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

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

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

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

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

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

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

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

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

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

Disentangling Trait and State Agreement using LST

Although a full presentation of LST theory (see Geiser, Bishop, & Lockhart, 2015) is beyond the scope of this paper, we will reference its main assumptions that directly relate to the 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).

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

For instance, using a structural equation modeling framework, the singletrait-multistate (STMS) model for three observed indicators (e.g., items of a questionnaire) measured at three time points requires the estimation of four latent variables to separate trait and state effects (see Figure 1). First, a common latent trait variable ξ (measured by all nine indicators) is modeled to reflect the general, time-unspecific mean level of the construct under investigation. Importantly, both the factor loading (λ) and intercept (α) of the same item *i* should be invariant across time to ensure strong (i.e., scalar) longitudinal measurement invariance at the trait-level (i.e., the lack of measurement-related alterations due to different use of the rating scale or interpretations of the items over time; Millsap, 2011; Widaman, Ferrer, & Conger, 2010). Second, three time-specific, latent state residual factors (ζ_1, ζ_2, and ζ_3; each measured by the three indicators used at each time point) are estimated to capture participants' deviations from the general latent trait. Since latent state residual factors are defined as momentary deviations from the general latent trait, only weak (i.e., metric) longitudinal invariance of factor loadings γ is required (latent state residual factors have a mean of zero by definition).

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

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

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

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

The Present Study

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

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

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

Method

Participants

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

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

Measures

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

Results

Descriptive Statistics

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

STMS Results

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

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

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

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

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

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

Cross-informant STMS. The STMS model with partial scalar invariance across informants[3] (the factor loading and intercept of the caregiver-reported item "feels sorry for other children who are sad or upset" were not constrained to equality) showed a good fit to the data, χ^2 (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.

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

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

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

Discussion

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

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

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

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

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

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

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

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

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

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

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

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

Limitations

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

Conclusions and Future Directions

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

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

ETHICS STATEMENT

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.

References

Achenbach, T.M., McConaughy, S.H., & Howell, C.T. (1987). Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. *Psychology Bulletin, 101*, 213-232. doi:10.1037/0033-2909.101.2.213

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms & Profiles.* Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*, 456–466. doi:10.1037/0033-2909.105.3.456

Chaplin, T. M., & Aldao, A. (2013). Gender differences in emotion expression in children: A meta-analytic review. *Psychological Bulletin, 139*, 735-765. doi:10.1037/a0030737

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indices for testing measurement invariance. *Structural Equation Modeling, 9*, 233–255. doi:10.1207/S15328007SEM0902_5

Church, A. T., Katigbak, M. S., Reyes, J. A. S., Salanga, M. G. C., Miramontes, L. A., & Adams, N. B. (2008). Prediction and cross-situational consistency of daily behavior across cultures: Testing trait and cultural psychology perspectives. *Journal of Research in Personality, 42*, 1199-1215. doi:10.1016/j.jrp.2008.03.007

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

de Haan, A., Prinzie, P., Sentse, M., & Jongerling, J. (2017). Latent Difference Score modeling: A flexible approach for studying informant discrepancies. *Psychological Assessment.* Advance online publication. doi:10.1037/pas0000480

De Los Reyes, A., Augenstein, T. M., Aldao, A., Thomas, S. A., Daruwala, S. E., Kline, K., & Regan, T. (2015). Implementing psychophysiology in clinical assessments of adolescent social anxiety: Use of rater judgments based on graphical representations of psychophysiology. *Journal of Clinical Child & Adolescent Psychology, 44*, 264–279. doi:10.1080/15374416.2013.859080

De Los Reyes, A., & Kazdin, A. E. (2004). Measuring informant discrepancies in clinical child research. *Psychological Assessment,16*, 330–334. doi:10.1037/1040-3590.16.3.330

De Los Reyes, A., & Kazdin, AE. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin, 131*, 483-509. doi:10.1037/0033-2909.131.4.483

De Los Reyes, A., Thomas, S. A., Goodman, K. L., & Kundey, S. M. A. (2013). Principles Underlying the Use of Multiple Informants' Reports. *Annual Review of Clinical Psychology, 9*, 123–149. doi: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*, 405–432. doi:10.1111/j.1467-8624.2010.01564.x.

Eisenberg, N., Fabes, R. A., Murphy, B., Karbon, M., Smith, M., & Maszk, P. (1996). The relations of children's dispositional empathy-related responding to their emotionality,

regulation, and social functioning. *Developmental Psychology, 32*, 195–209. doi:10.1037/0012-1649.32.2.195

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

Eisenberg, N., Spinrad, T., & Morris, A. (2014). Empathy-related responding in children. In M. Killen & J. Smetana (Eds.*), Handbook of moral development* (2nd ed., pp. 184–207). New York, NY: Psychology Press.

Funderburk, B. W., Eyberg, S. M., Rich, B. A., & Behar, L. (2003). Further psychometric evaluation of the Eyberg and Behar rating scales for parents and teachers of preschoolers. *Early Education and Development, 14*, 67–81.

Geiser, C., Bishop, J., & Lockhart, G. (2015). Collapsing factors in multitrait-multimethod models: Examining consequences of a mismatch between measurement design and model. *Frontiers in Psychology: Quantitative Psychology and Measurement, 6*, 946. doi:10.3389/fpsyg.2015.00946

Geiser, C., Hintz, F., Burns, G., Servera, M. (2017). Latent variable modeling of person-situation data. In Rauthmann, J. F., Sherman, R., Funder, D. C. (Eds.), *The Oxford Handbook of Psychological Situations* (pp. 1-57). Oxford University Press.

Geiser, C., Keller, B. T., & Lockhart, G. (2013). First- versus second-order latent growth curve models: Some insights from latent state-trait theory. *Structural Equation Modeling, 20*, 479–503. doi:10.1080/10705511.2013.797832

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

Kienbaum, J. (2014). The development of sympathy from 5 to 7 years: Increase, decline or stability? A longitudinal study. *Frontiers in Psychology, 5*, 468. doi:10.3389/fpsyg.2014.00468

Kimonis E. R., Frick P. J., Muñoz L. C., & Aucoin K. J. (2008). Callous-unemotional traits and the emotional processing of distress cues in detained boys: testing the moderating role of aggression, exposure to community violence, and histories of abuse. *Developmental Psychopathology, 20*, 569–589. doi:10.1017/S095457940800028X

Kline, R. B. (2010). *Principles and practices of structural equation modeling* (3rd ed.). New York, NY: Guilford.

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. doi:10.1080/15374416.2015.1121822

Malti, T., Eisenberg, N., Kim, H., & Buchmann, M. (2013). Developmental trajectories of sympathy, moral emotion attributions, and moral reasoning: The role of parental support. *Social Development, 22*, 773–793. doi:10.1111/sode.12031

Malti, T., Sette, S., & Dys, S.P. (2016). Socioemotional responding: A perspective from Developmental Psychology. In R. Scott, M. Buchmann, & S. Kosslyn (Eds.), *Emerging Trends in the Social and Behavioral Sciences.* Hoboken, NJ: John Wiley and Sons. doi:10.1002/9781118900772.etrds0415

Malti, T., & Song, J. -H. (in press). Social-emotional development and aggression. In T. Malti & K. H. Rubin (Eds.), *Handbook of child and adolescent aggression: Emergence, development, and intervention*. New York, NY: Guilford Press.

McArdle, J.J., & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data. In: Collins, L.M., & Sayer, A.G. (Eds). *New methods for the analysis of change* (pp. 139-175). Washington, DC: APA Press.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York, NY: Taylor and Francis Group.

Murphy, B. C., Shepard, S. A., Eisenberg, N., Fabes, R. A., & Guthrie, I. K. (1999). Contemporaneous and longitudinal relations of young adolescents' dispositional sympathy to their emotionality, regulation, and social functioning. *Journal of Early Adolescence, 19,* 66–97.

Muthén, L.K., & Muthén, B.O. (1998-2017). *Mplus user's guide,8th Edn.* Los Angeles, CA: Muthén & Muthén.

Santor, D. A., Bagby, R. M., & Joffe, R. T. (1997). Evaluating stability and change in personality and depression. *Journal of personality and social psychology, 73*, 1354-1362.

Stern, J.A., & Cassidy, J. (2017). Empathy from infancy to adolescence: An attachment perspective on the development of individual differences. *Developmental Review*. Advance online publication. doi:10.1016/j.dr.2017.09.002

Steyer, R., Ferring, D. & Schmitt, M. J. (1992). States and traits in psychological assessment. *European Journal of Psychological Assessment, 8*, 79–98.

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*, 4–70. doi:10.1177/109442810031002

Widaman, K. F., & Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives,4*, 10–18. doi:10.1111/j.1750-8606.2009.00110.x

Zuffianò, A., Colasante, T., Buchmann, M., & Malti, T. (2017). The co-development of sympathy and overt aggression from middle childhood to early adolescence. *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	*df*	*scf*	χ^2/df	*p*	CFI	TLI	RMSEA (90%CI)	MC	$\Delta\chi^2$	Δdf	*p*	ΔCFI
Sympathy (Ca)													
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
Sympathy (Te)													
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

		Caregivers			Teachers		
	He/She usually:	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
	Variances						
	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.

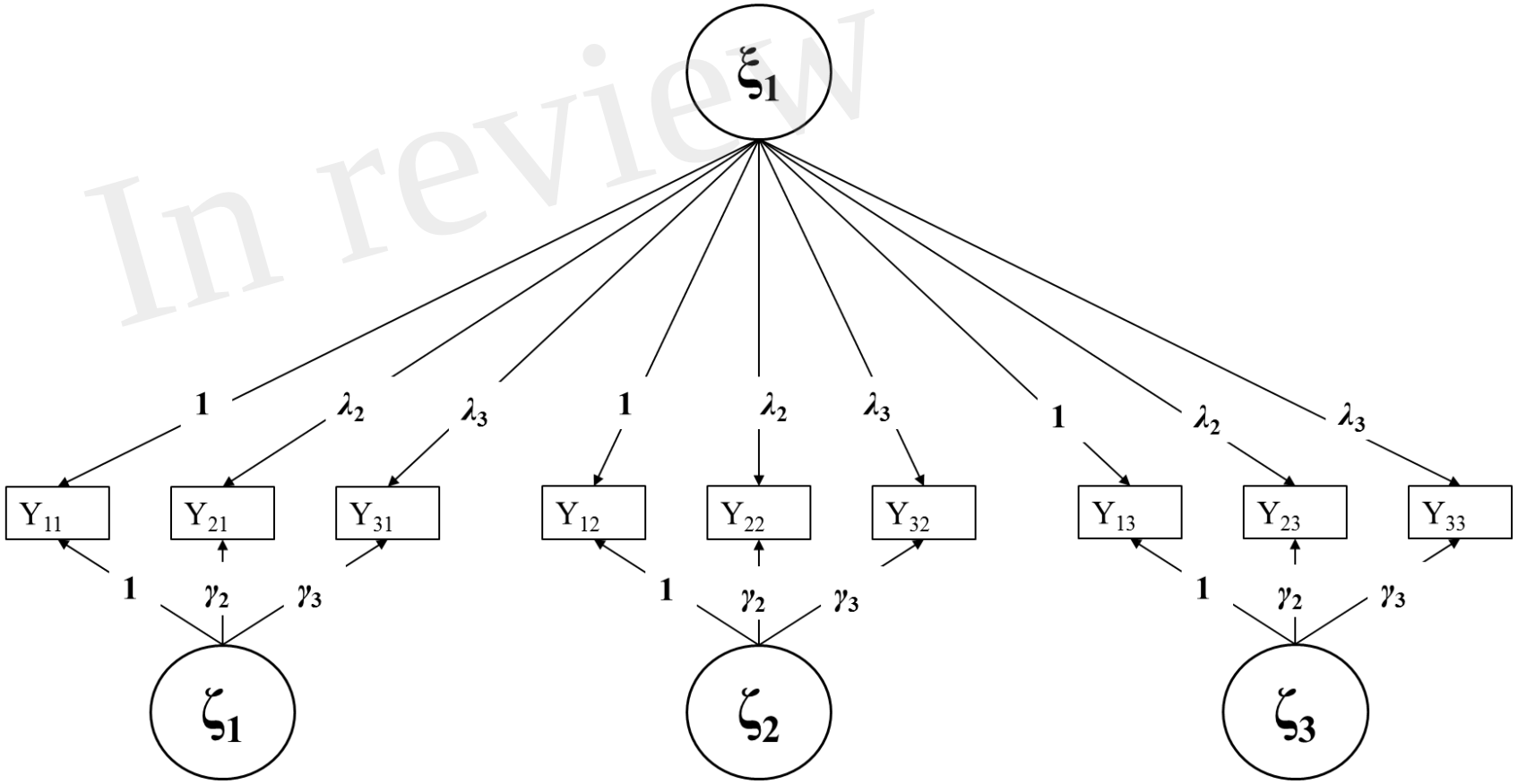

Figure 1. Singletrait-multistate (STMS) Model for Three Waves.

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.

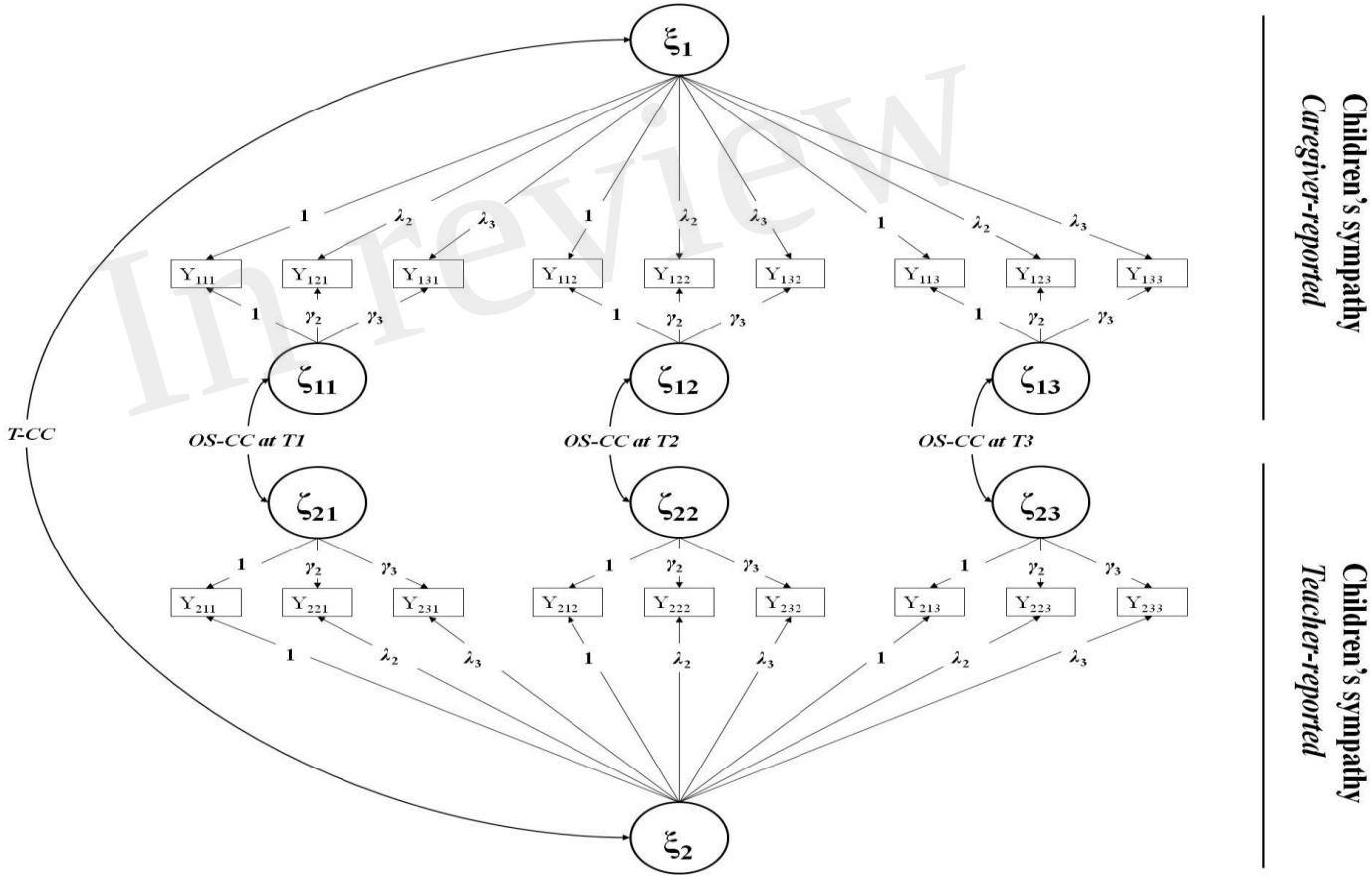

Figure 2. Combined Singletrait-multistate (STMS) Model for Three Waves and Two Informants.

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.

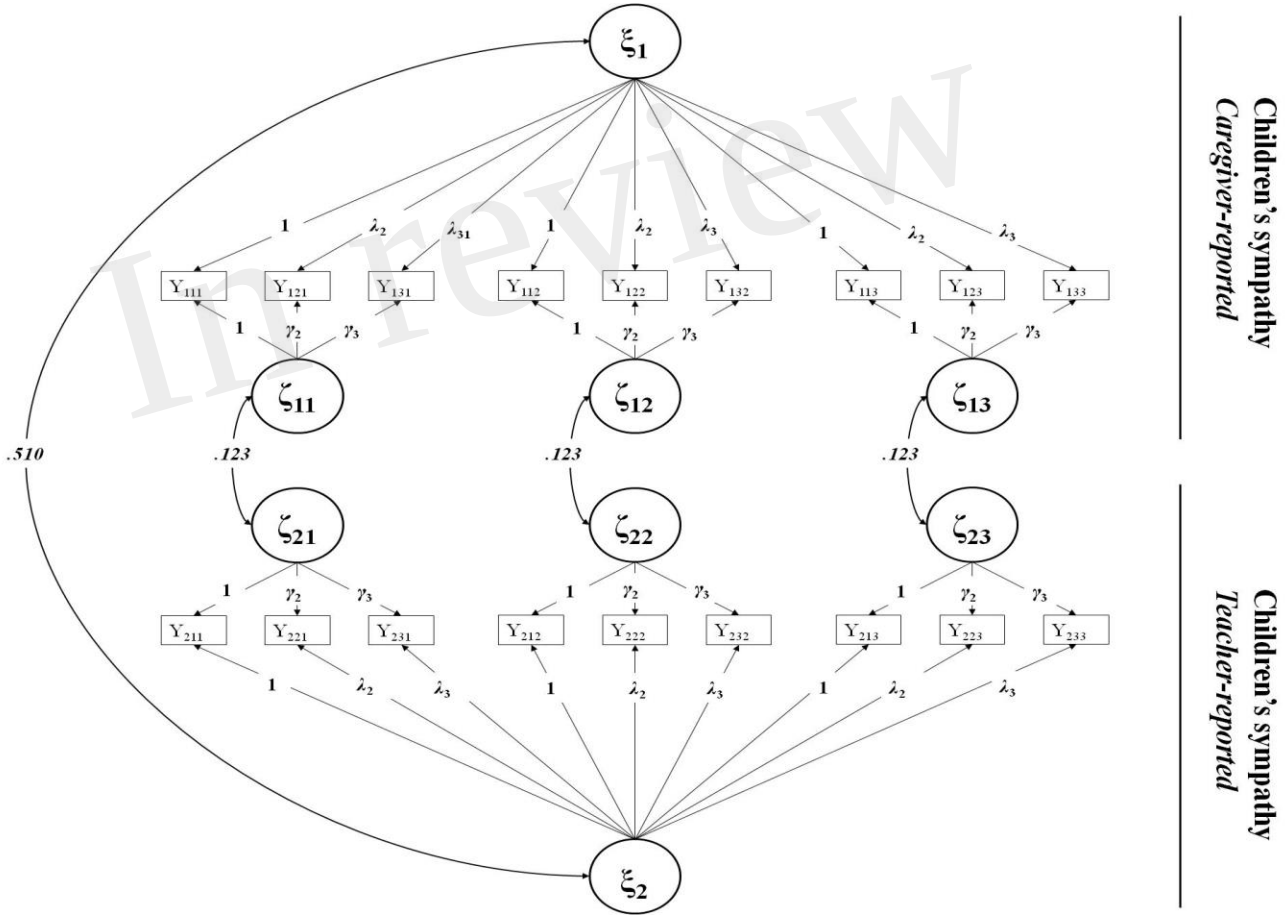

Figure 3. Combined Singletrait-multistate (STMS) Model of Children's Sympathy across Caregivers and Teachers.

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).