



Published in final edited form as:

J Psychopathol Behav Assess. 2011 March 1; 33(1): 3–10. doi:10.1007/s10862-010-9187-4.

Measuring Context-Specific and Cross-Contextual Components of Hierarchical Constructs

Levent Dumenci,

Department of Social and Behavioral Health, Virginia Commonwealth University, PO Box 980149, Richmond, VA 23298, USA

Thomas M. Achenbach, and

Department of Psychiatry, University of Vermont, 1 South Prospect Street, Burlington, VT 05401, USA

Michael Windle

Department of Behavioral Sciences and Health Education, Emory University, 1518 Clifton Road, Atlanta, GA 30322, USA

Levent Dumenci: LDumenci@vcu.edu; Thomas M. Achenbach: Thomas.Achenbach@uvm.edu; Michael Windle: MWindle@emory.edu

Abstract

Rating scales are often used to measure behavioral constructs. Yet, different informants' ratings may not necessarily agree. The situational specificity (SS) perspective postulates that discrepancies between ratings by different informants are primarily attributable to contextual behavior of the people being rated. The multitrait-multimethod (MTMM) perspective, however, attributes discrepancies between informants to rater bias, i.e., each informant provides a systematically distorted picture of the person being rated. Similarly, the Attribution-Bias-Context (ABC) perspective also attributes informant discrepancies to systematic biases. Within the context of measuring hierarchical constructs, we proposed a hybrid perspective that takes account of variance attributable to the behavior of the person being rated in a particular context from the perspective of a specific informant. We then provided a parametric representation of this perspective and analyses of mother, teacher, and self-ratings of Rule-Breaking and Aggressive Behavior to illustrate features of the model. Strengths and limitations of the SS, MTMM, and hybrid perspectives are discussed.

Keywords

Situational specificity; Method effect; Psychometric modeling; Child psychopathology; Externalizing problems

Introduction

Rating scales are often used to measure behavioral constructs. Examples include mothers' ratings of children's behavioral problems, teachers' ratings of students' achievement, and supervisors' ratings of workers' job performance. An impressive body of empirical evidence documents inconsistencies between different informants' ratings (Achenbach et al. 2005; Achenbach et al. 1987; Duhig et al. 2000; Meyer 2002; Renk and Phares 2004). Two

competing theoretical perspectives, designated as situational specificity (SS) and multitrait-multimethod matrix (MTMM), have provided rationales and methods for explaining discrepancies between informants' ratings. In this article, we propose a perspective that integrates the SS and MTMM perspectives, specify a measurement model consistent with this perspective, and provide data to illustrate features of the proposed model.

Consider a hypothetical study of violence prevention where both the mother and teacher of a youth rate the item "physically attacks people," which is endorsed by the mother but not by the teacher. Rooted in the philosophical view of contextualism (Preyer and Peter 2005), the SS perspective offers a compelling, yet simple explanation for the discrepancy between mother and teacher ratings. That is, people behave differently in different environments. Thus, the youth may display aggressive behavior at home when surrounded by siblings, whereas strict disciplinary practices prevent aggressive behavior at school. There is abundant evidence for contextual specificity of behavior (e.g., Biglan 1995; Morris 1988).

The MTMM perspective (Campbell and Fiske 1959) offers a radically different, yet also compelling, explanation for discrepancies between informants' ratings, i.e., ratings are method dependent. When each type of rater is viewed as a method of measurement, the MTMM perspective holds that discrepancies reflect systematically distorted reports of the behavior that is rated. Therefore, discrepancies between mother and teacher ratings tell us more about the characteristics of the raters than about the behavior being rated. Terms such as "halo effect," "nay saying," and "social desirability" are used to describe various forms of rater bias. De Los Reyes and Kazdin (2005) identified three sources of bias to explain informant discrepancies in the assessment of psychopathology: The actor-observer attribution bias (Jones and Nisbett 1972), memory recall bias (Tversky and Marsh 2000), and the context in which the behavioral ratings are obtained (clinical settings vs. other settings). The evidence for some level of rater bias is compelling (Eid and Diener 2006).

The MTMM and SS perspectives both enjoy broad empirical support. However, each perspective's strength may reflect the other's weakness. The SS perspective tends to ignore rater bias, whereas the MTMM perspective tends to ignore contextual behavior. Consequently, measurement models originating from the MTMM perspective incorrectly attribute contextual behavior to method bias, resulting in an over-estimation of systematic error variance (i.e., rater effect). Conversely, measurement models originating from the SS perspective incorrectly attribute rater bias to contextual effects, resulting in an over-estimation of systematic variability in behavior (i.e., context specificity). There is a tendency for researchers to use the MTMM perspective to analyze data from multiple informants, select a measurement model rooted in the MTMM perspective, attribute inconsistencies between informants to rater bias, and briefly mention that the estimated method variance may well be contaminated with context specific behavior (e.g., Konold and Pianta 2007).

Purposes of this Study

Evidence for rater bias may argue against the SS perspective, whereas evidence for contextual specificity of behavior may argue against the MTMM perspective. We argue for an alternative view that bridges the SS and MTMM perspectives by proposing that systematic discrepancies between raters are attributable to *contextual behavior as seen from the perspective of a particular rater*. This particular systematic effect will be referred to as "contextual" hereafter. The proposed perspective explicitly attributes discrepancies between informants to both contextual aspects of behavior (i.e., situational specificity) and rater biases (i.e., behavior as seen from the rater's perspective). This alternative view is consistent with the empirical support for both the SS and MTMM perspectives.

The proposed perspective is applicable to hierarchical constructs. Examples include the higher-order constructs of externalizing behavior, which subsumes the rule-breaking and aggressive behavior as lower-order constructs; cognitive ability, which subsumes verbal, quantitative, and spatial abilities; and negative affect, which subsumes anxiety, depression, and withdrawal. Krueger and Markon (2006) argue that comorbid conditions (e.g., rule-breaking and aggressive behavior) should be interpreted as lower-order constructs and that associations between such lower-order constructs are reflected in higher-order constructs (e.g., externalizing problems). Because comorbid conditions are so common, the hierarchical conceptualizations of behavioral constructs may be extremely useful for research and practice.

In the sections that follow, we introduce a model for a three-faceted measurement design for estimating contextual components (i.e., behaviors specific to the context from the perspective of an informant) and cross-contextual components (i.e., behaviors common across contexts and informants) of hierarchical constructs. We then illustrate the model with an example involving mother, teacher, and youth reports of Rule-Breaking Behavior and Aggressive Behavior as lower-order constructs (i.e., syndromal constructs) and Externalizing Problems as a higher-order construct. This illustration involves variance partitioning of ordinal ratings.

Measurement Model

From the measurement perspective, we identify three systematic sources of variability in item responses: (a) a cross-contextual higher-order trait common across all three facets of the measurement design, i.e., item, syndrome, and informant; (b) a contextual higher-order trait common across the item and syndrome facets but specific to each informant; and (c) a contextual lower-order trait specific to all possible combinations of syndrome and informant facets. Returning to the physical aggression example, consider a child exhibiting some externalizing problems (i.e., problems between the child and others). We use the phrase “cross-contextual higher-order trait” to refer to the externalizing problems manifested across multiple contexts (e.g., home and school) as reported by multiple informants. Additionally, the child exhibits externalizing problems in a specific context as reported by a specific informant. We use the phrase “contextual higher-order trait” to refer to externalizing problems when, for example, the child steals from classmates’ backpacks and destroys their belongings in classrooms, as reported by the teacher. Whereas the child’s rule-breaking behavior (e.g., stealing) and aggressive behavior (e.g., destroying others’ belongings) represent comorbid conditions in classrooms as reported by the teacher, each contextual lower-order trait represents a specific type of behavior (e.g., aggression but not rule-breaking) manifested in a specific context (e.g., classroom) as reported by a specific informant (teacher). All other sources of variability in item ratings not represented in these three sources of systematic effects are subsumed under the item residuals (i.e., random error of measurement).

Parametrically, the model is given by level 1 and level 2 equations as:

$$y_{ijk}^* = \lambda_{ijk} \eta_{jk} + \varepsilon_{ijk} \quad (1)$$

$$\eta_{jk} = \gamma_{jk} \delta + \gamma_{jk} \xi_k + \zeta_{jk} \quad (2)$$

where y_{ijk}^* is the underlying continuous latent response variable for the i^{th} item ($i=1, \dots, p_j$) that measures the j^{th} syndrome ($j=1, \dots, q$) rated by the k^{th} informant ($k=1, \dots, r$). The symbols

δ , ξ_k , and ζ_{jk} are used to represent (a) the cross-contextual higher-order trait (second-order factor); (b) the contextual higher-order trait (second-order factor); and (c) the contextual lower-order traits (second-order residual), respectively. Also, the 1st-order factor loadings are represented as λ_{ijk} and the 2nd-order factor loadings on the general factor as γ_{jk} and on the group factors as γ_j . The symbol ε_{ijk} is used to represent the random error of measurement, i.e., item residuals as the portion of the item responses not explained by the three types of latent variables. The visual display of the model appears in Fig. 1a and b. In addition to directional relations between observed and latent variables, as well as between latent variables, Fig. 1a and b provide symbols corresponding to in Eqs. 1 and 2 to illustrate the model description.

Variance decompositions of underlying item responses (y_{ijk}^*) follow directly from inserting Eq. 2 into Eq. 1 and taking the variances of both sides of the equation:

$$\text{VAR}(y_{ijk}^*) = \lambda_{ijk}^2 \gamma_{jk}^2 \text{VAR}(\delta) + \lambda_{ijk}^2 \gamma_{jk}^2 \text{VAR}(\xi_k) + \lambda_{ijk}^2 \text{VAR}(\zeta_{jk}) + \text{VAR}(\varepsilon_{ijk}) \quad (3)$$

Therefore, the terms $\lambda_{ijk}^2 \gamma_{jk}^2 \text{VAR}(\delta)$, $\lambda_{ijk}^2 \gamma_{jk}^2 \text{VAR}(\xi_k)$, $\lambda_{ijk}^2 \text{VAR}(\zeta_{jk})$, and $\text{VAR}(\varepsilon_{ijk})$ represent the percentage of variance in item responses accounted for by the cross-contextual higher-order trait, the contextual higher-order trait, the contextual lower-order trait, and the random error of measurement, respectively.

In terms of measurement theory, our model has roots in the latent trait-state model (LTSM; Steyer et al. 1989) and the hierarchical factor model (HFM; Wherry 1959; Yung et al. 1999). The LTSM was originally proposed to account for stable and labile measurement characteristics in longitudinal measurement designs but the underlying characteristic of LTSM is replication, whether cross-sectional, i.e., replications over informants, or longitudinal, i.e., replication over time (Dumenci et al. 2005). Our model utilizes the cross-sectional extension of LTSM, along with the 2nd-order extension of the HFM. The 1st-order factors capture the associations among responses to items by means of contextual lower-order traits, whereas the 2nd-order factors capture the associations among the contextual lower-order traits by means of the cross-contextual higher-order trait and contextual higher-order traits. Thus, the 2nd-order factor residuals represent the contextual lower-order traits after taking account of the cross-contextual and contextual higher-order traits.

Method

Participants

Originally used to develop the 1991 and 2001 scales of the Child Behavior Checklist (CBCL), Teacher's Report Form (TRF), and Youth Self-Report (YSR), the samples included 5,543 youths between the ages of 10 and 18, assessed in U. S. national surveys ($N=2,785$) and in mental health and special education services throughout the U.S. ($N=2,758$; Achenbach 1991, and Achenbach & Rescorla 2001, provide details). For the national surveys, multistage probability methods were used to obtain samples that were representative with respect to gender, socioeconomic status (SES), ethnicity, and rural-suburban-urban residence in the 48 contiguous states. The survey data were obtained in home interviews conducted by Temple University's Institute for Survey Research in 1989 (90% completion rate) and 1999 (93% completion rate). Mean age was 14.2 years ($SD=2.5$) in the survey sample and 13.3 years ($SD=2.2$) in the clinical sample. Mean SES was 5.5 ($SD=2.1$) for the survey sample and 5.7 ($SD=2.2$) for the clinical sample, based on Hollingshead's (1975) 9-point scale for the occupation of the parent holding the higher status job. Parental permission was obtained to collect TRF and YSR data. Of the 5,543

youths, 2,046 had all three forms (i.e., CBCL, TRF, and YSR), 856 had the CBCL and TRF only, 1,964 the CBCL and YSR only, and 677 the TRF and YSR only. All CBCL forms were filled out by mothers because CBCLs completed by fathers, parent surrogates, and other adults (e.g., grandparents) were excluded from the present study.

Instruments and Procedure

The CBCL, TRF, and YSR are standardized questionnaires for obtaining parent, teacher, and self-reports of academic and adaptive functioning and behavioral/emotional problems. Extensive validity and reliability data are provided by Achenbach and Rescorla (2001). This study focused on 22 Externalizing items rated by mothers (CBCL), teachers (TRF), and the youths themselves (YSR): 7 items measuring Rule-Breaking Behavior and 15 items measuring Aggressive Behavior. Each item is rated on the following 3-point scale: 0 = *not true (as far as you know)*; 1 = *somewhat or sometimes true*; and 2 = *very true or often true*. These 22 items are common to the 1991 and 2001 versions of all three instruments. Examples of items on the Rule-Breaking Behavior scale are: *Hangs around with others who get in trouble*; *steals*; and *doesn't seem to feel guilty after misbehaving*. Examples of items on the Aggressive Behavior scale are: *Gets into many fights*; *physically attacks people*; and *threatens people*.

Achenbach and Rescorla (2001) reported high Cronbach's alphas for Rule-Breaking Behavior, Aggressive Behavior, and Externalizing Problems: .85, .94, .94; .95, .95, .95; and .81, .86, .90 for the CBCL, TRF, and YSR, respectively. Test-retest reliabilities estimated at mean intervals of 8 to 16 days were also high for the Rule-Breaking Behavior, Aggressive Behavior, and Externalizing Problems: .91, .90, .92; .83, .88, .89; and .83, .88, .89 for the CBCL, TRF, and YSR, respectively.

Data Analysis

To illustrate our model, we analyzed 5,543 cases that had ratings by at least two informants because parameter estimates based only on complete cases would be biased (Little and Rubin 2002, p. 41). Attrition analyses were conducted to test differences between Externalizing scores reported by two informants for youths who had data from all three informants (i.e., complete data) versus youths who had data from only two informants. In addition, the youths' age, gender, and referral status were used as fixed effects in the analyses. Results from three MANOVAs indicated that, after taking account of youths' age, sex, and referral status, lack of ratings by one informant was not significantly associated with the other two informants' ratings of the youths' Externalizing problems.

The 0-1-2 ratings of each item were treated as ordered categories in the analyses. To estimate the models, we used Weighted Least Squares with robust standard error and mean- and variance-adjusted fit statistic (WLSMV), an asymptotically distribution-free (ADF) estimator (Muthén and Muthén 1998–2004). Thus, the estimation method took account of the ordered categorical item distributions using all available data. To evaluate model fit, we used the Root Mean Square Error of Approximation (RMSEA; Steiger 1990) and the Tucker-Lewis index (TLI; Tucker and Lewis 1973). RMSEA values $\leq .06$ and TLI values $\geq .95$ indicated good model fit for the WLSMV method (Yu and Muthén 2002).

Results

By simultaneously modeling 7 items measuring Rule-Breaking and 15 items measuring Aggressive Behavior from mother, teacher, and self-ratings, we estimated three construct-relevant components: (a) the cross-contextual higher-order trait of Externalizing, (b) the contextual higher-order trait of Externalizing, and (c) the contextual lower-order traits of

Rule-Breaking and Aggressive Behavior (see Fig. 1a and b). Descriptive fit indices supported the proposed measurement model: RMSEA=.046 and TLI=.976.

Parameter estimates appear in Table 1. Item variances were partitioned by inserting parameter estimates into Eq. 3, e.g., teachers' ratings of item 6. *Swearing or obscene language*:

$$\begin{aligned} \text{VAR}(y_{612}^*) &= (1.012)^2(1.023)^2(.378) + (1.012)^2(1.000)^2(.332) \\ &\quad + (1.012)^2(.055) + .199 \\ &= .405 + .340 + .056 + .199 \end{aligned}$$

reflecting the proportion of variance accounted for by the cross-contextual higher-order trait of Externalizing, contextual higher-order trait (teacher), contextual lower-order trait of Rule-Breaking Behavior (teacher), and measurement error, respectively.

Figure 2 illustrates the summary of variance partitioning. The cross-contextual higher-order trait of Externalizing explained the largest percentage of variance in ratings (range: 21% to 41%), followed by the contextual higher-order trait of Externalizing (range: 8% to 31%). The contextual syndromal traits of Rule-Breaking Behavior and Aggressive Behavior accounted for small but still significant percentages of variance in ratings (range: 2% to 9%). Taken together, the results showed a strong level of agreement among ratings of Rule-Breaking Behavior and Aggressive Behavior by three informants at the latent level. After taking account of the cross-contextual higher-order trait of Externalizing, the results further supported the construct validity of three contextual traits: Rule-Breaking Behavior, Aggressive Behavior, and Externalizing.

In Figure 2, the height of each bar indicates the total amount of construct-relevant variance accounted for by each possible combination of informant-syndromal trait pair. The ratings of Aggressive Behavior had somewhat higher levels of construct-relevant variance than the ratings of Rule-Breaking Behavior across the informants, whereas teachers' ratings had slightly higher levels of construct-relevant variance than mothers' ratings, which were followed by youths' ratings for both Rule-Breaking Behavior and Aggressive Behavior.

Discussion

Context specificity of behavior has long been recognized in prevention and intervention studies. Most prevention programs are, in fact, administered within a specific context, such as peer groups, family, classroom, school, and community. When rating scales are used to assess the effectiveness of a prevention study, it is important to test the degree to which the effectiveness generalizes beyond the specific context where the intervention program is administered. We proposed a general framework for testing context-specific and cross-contextual aspects of behavior and a measurement model originating from this framework. Using data for adolescent psychopathology, we tested the context-specific and cross-contextual components of behavior.

It has long been evident that different informants (e.g., parents and teachers) may observe behaviors that are specific to the environments in which they interact with children and youths (Achenbach et al. 1987). Differential observations and personal interactions may therefore be reflected in informants' responses to assessment instruments. Thus, assessment procedures need to include information from multiple informants in order to obtain comprehensive pictures of problem behavior. In our view, inter-informant discrepancies stem from the combinations of situation-specific behavior and informant characteristics that affect ratings of the behavior. We qualify both lower- and higher-order traits with a prefix

“contextual” to specifically refer to the contextual behavior from the perspective of an informant. For example, the Contextual Aggressive Behavior (Teacher) trait refers to the aggressive behavior in school as seen from teachers’ perspectives. Like the MTMM and SS perspectives, our model does not separate contextual aspects of behavior from rater bias. However, our model also does not assume either the absence of contextual behavior, as the MTMM perspective does, or the absence of rater bias, as the SS perspective does.

It is essential to compare corresponding models originating from the SS, MTMM, and hybrid perspectives empirically in three-faceted measurement designs involving items, constructs, and informants. Currently available models from the SS and MTMM perspectives do not directly correspond to the model proposed in this study. When available, however, simulation studies will supplement empirical studies in testing the soundness of different conceptualizations. Because De Los Reyes and Kazdin (2005) did not provide an explicit measurement model to represent their Attribution-Bias-Context (ABC) perspective, a parametric comparison between the ABC perspective and other perspectives, including the hybrid perspective, is not yet possible.

Single-informant and multi-informant studies have either treated Rule-Breaking Behavior and Aggressive Behavior as two separate yet correlated constructs, or Externalizing as a unitary construct. Consequently, analyses of Rule-Breaking Behavior, Aggressive Behavior, and Externalizing were problematic in a given statistical model due to linear dependency (i.e., Externalizing = Rule-Breaking + Aggressive). With the measurement model introduced in this study, we provided a broader conceptual orientation, with an emphasis on the measurement of the cross-contextual trait of Externalizing and the contextual traits of Externalizing, as well as the contextual syndromal traits.

The decomposition of syndromal traits showed that the CBCL, TRF, and YSR are all highly reliable measures of Rule-Breaking Behavior and Aggressive Behavior, i.e., the proportion of construct-relevant variance relative to total variance was large for all three forms. The lower-order and higher-order constructs explained variances in observed 0-1-2 ratings as large as 73.5%, as shown in Fig. 2. Overall, mothers’ reports were somewhat more reliable than youths’ self-reports, whereas teachers’ reports were somewhat more reliable than mothers’ reports. Mothers’ reports showed the largest cross-situational consistency, whereas teachers’ reports showed the largest situation-specificity (i.e., contextual syndromal trait). Results further support the contentions that thorough assessment requires information from multiple informants and that informant-specific behavioral patterns may reflect contextually specific characteristics of the people who are being assessed.

Studies reporting modest cross-informant correlations do not tell us how to improve practice. The measurement model introduced in this study effectively shifts the focus from statistical significance and effect size issues to directly accounting for cross-informant correlations. For example, a mother’s report of aggressive behavior contains information above and beyond how the mother alone views her child’s aggressive behavior. It may also reflect the child’s behavior in contexts not directly observed by the mother (e.g., school). There is a need to identify behavioral patterns common to different contexts by obtaining assessment data from multiple informants in order to plan and execute effective prevention and intervention studies. Equally important, informant-specific reports should never be overlooked merely because these reports do not totally agree with one another. Within the constraints of observational research designs, future studies need focus on conceptual, methodological, and statistical approaches that separate the contextual characteristics of the individual being assessed from the characteristics of those who provide the data (e.g., parents, teachers).

References

- Achenbach, TM. Manual for the child behavior checklist/4–18 and 1991 profile. Burlington: University of Vermont, Department of Psychiatry; 1991.
- Achenbach TM, Krukowski RA, Dumenci L, Ivanova MY. Assessment of adult psychopathology: meta-analysis and implications of cross-informant correlations. *Psychological Bulletin* 2005;131:361–382. [PubMed: 15869333]
- Achenbach TM, McConaughy SH, Howell CT. Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. *Psychological Bulletin* 1987;101:213–232. [PubMed: 3562706]
- Achenbach, TM.; Rescorla, LA. Manual for the ASEBA school-age forms & profiles. Burlington: University of Vermont, Research Center for Children, Youth, and Families; 2001.
- Biglan, A. Changing culture practices: A contextualistic framework for intervention research. Reno: Context Press; 1995.
- Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin* 1959;56:81–105. [PubMed: 13634291]
- De Los Reyes A, Kazdin AE. Informant discrepancies in the assessment of child psychopathology: a critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin* 2005;131:483–509. [PubMed: 16060799]
- Duhig AM, Renk K, Epstein MK, Phares V. Interparental agreement on internalizing, externalizing, and total behavior problems: a meta-analysis. *Clinical Psychology: Science and Practice* 2000;7:435–453.
- Dumenci, L.; Windle, M.; Achenbach, TM. Latent Trait-State- Model for cross-sectional research designs. Paper presented at the European Congress of Psychology at Granada; Spain. Jul. 2005
- Eid, M.; Diener, E., editors. Handbook of multimethod measurement in psychology. Washington: American Psychological Association; 2006.
- Hollingshead, AB. Unpublished paper. New Haven, CT: Yale University, Department of Sociology; 1975. Four factor index of social status.
- Jones, EE.; Nisbett, RE. The actor and the observer: Divergent perceptions of the causes of behavior. In: Jones, EE.; Kanouse, DE.; Kelly, HH.; Nisbett, RE.; Valins, S.; Weiner, B., editors. Attribution: Perceiving the causes of behavior. Morristown: General Learning Press; 1972. p. 79-94.
- Konold TR, Pianta RC. The influence of informants on ratings of children's behavioral functioning. *Journal of Psycho-educational Assessment* 2007;25:222–236.
- Krueger RF, Markon KE. Reinterpreting comorbidity: a model-based approach to understanding and classifying psycho-pathology. *Annual Review of Clinical Psychology* 2006;2:111–133.
- Little, RJA.; Rubin, DB. Statistical analysis with missing data. 2. Hoboken: Wiley; 2002.
- Meyer, GJ. Implications of information gathering methods for a refined taxonomy of psychopathology. In: Beutler, LE.; Malik, ML., editors. Rethinking the DSM: A psychological perspective. Washington: American Psychological Association; 2002. p. 69-105.
- Morris EK. Contextualism: the world view of behavior analysis. *Journal of Experimental Child Psychology* 1988;46:289–323.
- Muthén, LK.; Muthén, BO. Mplus: User's guide. Los Angeles: Muthén & Muthén; 1998–2004.
- Preyer, G.; Peter, G., editors. Contextualism in philosophy: Knowledge, meaning, and truth. Oxford: Oxford University Press; 2005.
- Renk K, Phares V. Cross-informant ratings of social competence in children and adolescents. *Clinical Psychology Review* 2004;24:239–254. [PubMed: 15081518]
- Steiger JH. Structural model evaluation and modification: an interval estimation method. *Multivariate Behavioral Research* 1990;25:173–180.
- Steyer R, Majcen AM, Schwenkmezger P, Buchner A. A latent state-trait anxiety model and its application to determine consistency and specificity coefficients. *Anxiety Research* 1989;1:281–299.

- Tucker LR, Lewis C. The reliability coefficient for maximum likelihood factor analysis. *Psychometrika* 1973;38:1–10.
- Tversky B, Marsh EJ. Biased retellings of events yield biased memories. *Cognitive Psychology* 2000;40:1–38. [PubMed: 10692232]
- Wherry RJ. Hierarchical factor solutions without rotation. *Psychometrika* 1959;24:45–51.
- Yu, CY.; Muthén, BO. Evaluation of model fit indices for latent variable models with categorical and continuous outcomes (Technical Report). Los Angeles: University of California at Los Angeles, Graduate School of Education & Information Studies; 2002.
- Yung YF, Thissen D, McLeod LD. On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika* 1999;64:113–128.

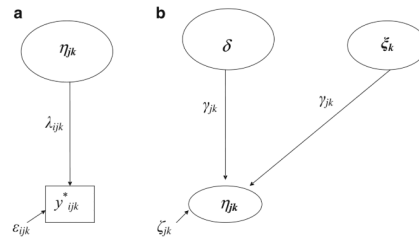


Fig. 1.
a. Measurement model for contextual behavior: Level 1 **b.** Measurement model for contextual behavior: Level 2

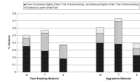


Fig. 2. Measurement model for Externalizing Behavior: Partitioning of variance estimates. M = mother; T = teacher; S = self

Table 1

Parameter estimates

	η_{11}	η_{12}	η_{13}	η_{21}	η_{22}	η_{23}	ϵ_{ijk}
γ^*_{111}	1						.439
γ^*_{211}	1.025						.411
γ^*_{311}	1.087						.337
γ^*_{411}	0.570						.818
γ^*_{511}	1.000						.439
γ^*_{611}	1.089						.336
γ^*_{711}	0.729						.702
γ^*_{112}	1						.218
γ^*_{212}	0.848						.437
γ^*_{312}	0.920						.338
γ^*_{412}	0.489						.806
γ^*_{512}	0.853						.430
γ^*_{612}	1.012						.199
γ^*_{712}	0.736						.576
γ^*_{113}	1						.816
γ^*_{213}	1.656						.495
γ^*_{313}	1.623						.515
γ^*_{413}	0.967						.828
γ^*_{513}	1.703						.466
γ^*_{613}	1.731						.448
γ^*_{713}	1.023						.807
γ^*_{121}				1			.425
γ^*_{221}				1.122			.277
γ^*_{321}				0.927			.506
γ^*_{421}				1.067			.346

	η_{11}	η_{12}	η_{13}	η_{21}	η_{22}	η_{23}	ξ_{ijk}
γ^*_{521}				1.138			.255
γ^*_{621}				1.105			.299
γ^*_{721}				1.092			.315
γ^*_{821}				0.948			.484
γ^*_{921}				1.006			.419
γ^*_{1021}				0.967			.452
γ^*_{1121}				0.901			.533
γ^*_{1221}				0.825			.609
γ^*_{1321}				1.121			.278
γ^*_{1421}				1.175			.206
γ^*_{1521}				0.950			.482
γ^*_{122}					1		.220
γ^*_{222}					1.013		.200
γ^*_{322}					0.933		.321
γ^*_{422}					0.925		.333
γ^*_{522}					1.062		.121
γ^*_{622}					1.007		.210
γ^*_{722}					1.003		.216
γ^*_{822}					0.904		.363
γ^*_{922}					0.956		.288
γ^*_{1022}					0.937		.316
γ^*_{1122}					0.870		.410
γ^*_{1222}					0.910		.355
γ^*_{1322}					1.024		.183
γ^*_{1422}					1.048		.143
γ^*_{1522}					0.921		.339
γ^*_{123}						1	.632

	η_{11}	η_{12}	η_{13}	η_{21}	η_{22}	η_{23}	ϵ_{ijk}
$y^*_{.223}$						1.088	.565
$y^*_{.323}$						0.713	.813
$y^*_{.423}$						1.214	.458
$y^*_{.523}$						1.311	.368
$y^*_{.623}$						1.280	.398
$y^*_{.723}$						1.262	.415
$y^*_{.823}$						0.893	.707
$y^*_{.923}$						0.698	.821
y^*_{1023}						0.783	.775
y^*_{1123}						0.662	.839
y^*_{1223}						0.899	.703
y^*_{1323}						1.112	.545
y^*_{1423}						1.352	.327
y^*_{1523}						0.785	.773
	δ	ζ_1	ζ_2	ζ_3	ζ_{jk}		
η_{11}	1.020	<u>1</u>			0.044		
η_{12}	1.023	<u>1</u>	<u>1</u>		0.055		
η_{13}	0.488			<u>1</u>	0.007*		
η_{21}	<u>1</u>	<u>1</u>			0.073		
η_{22}	1.037		<u>1</u>		0.041		
η_{23}	0.755			<u>1</u>	0.066		
Variance:	0.378		0.123	0.332	0.087		

All parameters are significant ($p < .01$), except for marked with an "*" ($p > .10$). Fixed parameters are underlined.