

Assessment of child problem behaviors by multiple informants: a longitudinal study from preschool to school entry

David C.R. Kerr,¹ Erika S. Lunkenheimer,² and Sheryl L. Olson³

¹Oregon Social Learning Center, USA; ²University of Oregon Child and Family Center, USA; ³University of Michigan Psychology Department, USA

Background: Children's early problem behavior that manifests in multiple contexts is often more serious and stable. The concurrent and predictive validity of ratings of externalizing and internalizing by four informants was examined at preschool and early school age in an at-risk sample. **Methods:** Two hundred forty children were assessed by mothers and fathers (Child Behavior Checklist (CBCL)), and teachers and laboratory examiners (Teacher Report Form (TRF)) at ages 3 and 5 years. **Results:** All informants' ratings of externalizing converged on a common factor at ages 3 and 5 that showed strong stability over time ($\beta = .80$). All informants' age 3 externalizing ratings significantly predicted the problem factor at age 5; mothers', fathers', and teachers' ratings were independently predictive. Ratings of internalizing (except by examiners at age 3) also converged at both ages; the problem factor showed medium stability ($\beta = .39$) over time. Only fathers' ratings of age 3 internalizing predicted the age 5 problem factor. **Conclusions:** Findings support the value of multi-informant assessment, uphold calls to include fathers in childhood research, and suggest that examiners provide valid, though non-unique assessment data. Examiner contributions may prove useful in many research contexts. **Key-words:** Externalizing, internalizing, childhood, testing, assessment, longitudinal studies, fathers. **Abbreviations:** CBCL: Child Behavior Checklist; TRF: Teacher Report Form.

For most preschoolers, aggression, defiance, and other externalizing problems represent normal developmental struggles with self-regulation that remit after early childhood (e.g., Campbell, Shaw, & Gilliom, 2000). In some cases, however, such problems follow a troubling trajectory into later childhood and adolescence that is costly to children, their families, and society at large (e.g., NICHD Early Child Care Research Network, 2004; Moffitt & Caspi, 2001). Significant internalizing problems (e.g., anxiety and depressed mood) are less commonly identified in young children, but increase across the childhood years (Gilliom & Shaw, 2004), and are predictive of continued adjustment problems in later childhood (e.g., Gazelle & Ladd, 2003). Understanding processes of stability and desistance in children's behavior problems is aided by assessments that include observations by multiple adults. However, questions remain regarding the validity and utility of ratings of child behavior by specific informant types. Thus, we examined associations among mothers', fathers', teachers' and laboratory examiners' ratings of externalizing and internalizing problems observed in an at-risk sample of children followed from the preschool to early school-age period.

In research and clinical work, behavioral rating scales are most commonly administered to mothers, reflecting their roles as primary caregivers, their stronger participation in behavioral research, and

assumptions that they most accurately observe children's problems (Phares, 1997; Loeber, Green, & Lahey, 1990). Yet, early problems that occur in multiple settings and relationship systems are generally more serious and stable over time (Campbell et al., 2000). Thus, multi-informant assessment of children may contribute to a better understanding of problem severity and future risk (Kagan, Snidman, McManis, Woodward, & Hardway, 2002). Additionally, multi-informant assessment can decrease source and setting error in behavioral research (Merrell, 1999; Offord et al., 1996), and may permit statistical parsing-out of true score variance (Kraemer et al., 2003).

In practice, multi-informant assessment of childhood problems is complicated, since ratings by informants from the same context typically are only moderately related, and ratings from informants from different settings are often only modestly associated (Achenbach et al., 1987; Grietens et al., 2004; Offord et al., 1996). At one time, this disagreement was thought to reflect the unreliability of measures and informant bias. However, discrepancies also can reflect true variation in children's behaviors across diverse settings and relational circumstances (e.g., Stanger & Lewis, 1993; Merrell, 1999). Presently, it is unclear to what extent particular informants' observations are associated with children's problems in other contexts or to their risk for later problems. The unique contributions informants make to prediction of later problems also is of interest. This 'value added' by specific

Conflict of interest statement: No conflicts declared.

© 2007 The Authors

Journal compilation © 2007 Association for Child and Adolescent Mental Health.

Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA

informants may inform study design, and help determine how to integrate discrepant information. Knowledge of the redundancy of informant contributions may guide decisions concerning optimal use of these data when other informants are unavailable.

While mother ratings are most commonly used, other informant types may contribute important information. Apart from studies of fathers' influences on children, little research has addressed the predictive strength of fathers' ratings of behavioral and emotional concerns. Thus, our study aims also addressed the general neglect of fathers in research on developmental psychopathology, an issue that has persisted despite being highlighted 18 years ago (see Phares, Fields, Kamboukos, & Lopez, 2005). Teachers also may provide valuable indices of child problems, because they observe behaviors outside of the home setting, witness children in situations that provoke strong reactions, and develop norms regarding child development due to their exposure to multiple children (Saudino, Ronald, & Plomin, 2005). Teachers' assessments have been used as predictive targets for longitudinal studies of stable disruptive behavior (e.g., Fagot & Leve, 1998; Hayden, Klein, & Durbin, 2005). For these reasons, we included fathers' and teachers' ratings of children's behavior problems in this evaluation of the convergence and validity of multiple informants' ratings.

Other relatively untapped sources of information concerning child behavior problems are laboratory examiners who administer research test batteries (e.g., temperament, IQ). Examiners' observations in these settings may be valuable since they occur under standardized conditions and are free of the relational history and context in which problem behavior often occurs (Anderson, Lytton, & Romney, 1986), and of informant bias that is specific to the child (Achenbach et al., 1987). Examiners also can foster a less formal relationship with children than typically occurs with unfamiliar adults (e.g., during medical visits). In our own study, examiners employed an enthusiastic, child-centered style, and engaged in unstructured play and snack time, all of which may ease children's generalized inhibition. The absence of parental support and the demands of testing also may have elicited behavioral and emotional dysregulation that otherwise unfolds more slowly.

Despite potential advantages, it is unclear whether examiners' ratings converge with those of others. The school psychology literature provides some clues, as observations during ability and achievement testing have been found to relate only modestly to children's behaviors in other settings (e.g., Glutting, Youngstrom, Oakland, & Watkins, 1996). Yet, research on one recently released instrument, the Test Observation Form (TOF; McConaughy & Achenbach, 2004) supported significant correspondence with parent and teacher reports on child externalizing problems ($r = .37$ and $.29$, respectively), though not on inter-

nalizing. In the present study, we extended this work by considering the concurrent and predictive validity and utility of problem ratings provided by examiners across the preschool to early school period.

We considered the following research questions. First, we tested whether examiners' ratings would yield internally consistent indices of child externalizing and internalizing problems at ages 3 and 5 years. We expected that problem scales based on examiner reports would be significantly lower than those based on teachers' reports, due to the brief observation period and low base rate of many problem behaviors. Second, we hypothesized that problem ratings contributed by mothers, fathers, teachers, and examiners would be positively correlated within and across time. Within time, we expected informant ratings to converge on externalizing and internalizing problem factors, representing variance in observed problems that is shared across settings. Children were hypothesized to show stability of individual differences on these factors across time. Third, we considered ratings made by mothers, fathers, teachers, and examiners at preschool age first as predictors and then as *unique* predictors of problems at early school age. Findings were expected to provide evidence for or against the concurrent and predictive validity of each informant's ratings.

Method

Participants

Participants were 240 children (118 girls), and their parents and teachers/caregivers, who were part of a longitudinal study. In order to recruit children with a range of behavioral adjustment levels, two different advertisements were periodically placed in regional newspapers and child care centers, one focusing on normally developing toddlers, and the other on hard-to-manage toddlers. Children were assessed at phase 1 (age range = 27 to 45 months, $M = 37.7$, $SD = 2.7$ months), and again at phase 2 (age range = 52 to 71 months, $M = 63.4$, $SD = 2.7$ months). Children initially represented the full range of externalizing symptom severity on the Child Behavior Checklist/2-3 (Achenbach, 1992); children in the medium to high range were oversampled ($T > 60$; 44%). Families were representative of the local population, a predominantly white, middle-class university town and its semi-rural environs in the midwestern United States. The majority of children were European-American (86%); others were African-American (5%) or biracial (8%). Children resided with two parents (88%), a single, never married parent (5%), or a single, divorced parent (7%). The highest education attained for mothers and fathers was high school diploma (19% and 24%, respectively), four-year college degree (46% and 34%, respectively), and graduate or professional training (35% and 42%, respectively). Median annual family income was \$52,000, ranging from under \$20,000 to over \$100,000.

Procedure

Parental reports. Mothers ($N = 235$) and a subsample of fathers ($N = 157$) completed questionnaires. The reasons for father non-participation were not available. Fathers who participated came from families with significantly higher SES [$t(236) = 4.2, p < .001$]. Families were paid for their participation.

Teacher reports. Childcare/preschool teachers (Phase 1 $N = 188$) and teachers (Phase 2 $N = 190$) who agreed to provide child behavior problem ratings were given gift certificates.

Laboratory examiner reports. Most children participated in a three-hour laboratory session at each study phase ($N = 227$ and 203). At each phase, a different examiner administered a series of tasks designed to tap individual differences in self-regulation, cognitive maturity, and social understanding. A study using these data is published elsewhere (Olson, Sameroff, Kerr, Lopez, & Wellman, 2005). Examiners were graduate and advanced undergraduate students in social work or psychology.

Measures

Externalizing and internalizing problems. Parents independently completed the Child Behavior Checklist for ages 2–3 (CBCL/2–3; Achenbach, 1992) at phase 1, and the CBCL for ages 6–18 (Achenbach & Rescorla, 2001) at phase 2; both are measures of children's behavioral and emotional problems. Parents provide ratings of the child's behavior currently or within the previous 2 (CBCL/2–3) or 6 months (CBCL/6–18) using a 3-point scale (very/often true, somewhat/sometimes true, or not true). The CBCL yields factor-analytic derived internalizing and externalizing scales.

Teachers and laboratory examiners completed the Caregiver/Teacher Report Form, Ages 2–5 (CTRF/2–5; Achenbach, 1997) at phase 1, and the Teacher Report Form for ages 6–18 (TRF/6–18; Achenbach & Rescorla, 2001) at phase 2. Teachers reported behavior occurring currently or within the last two months. Examiners rated the child's behavior during the assessment period. Achenbach (1997) and Achenbach and Rescorla (2001) derived broad internalizing and externalizing scales based on these questionnaires.

The CBCL/2–3, CBCL/6–18, C/TRF/2–5 and TRF/6–18 have the same response format and share many of the same items. The content of some items and of the externalizing and internalizing scales varies depending on the form, in order to capture developmental changes and behaviors unique to or more typical of different settings. Since the present study was concerned with individual differences, the scales were used as they were designed and normed. Raw parent report scores were used. Scales based on teachers' and examiners' reports had positively skewed distributions and thus were log-transformed and used unless otherwise specified.

Missing data considerations. Little's MCAR test (in SPSS 14.0) was used to test systematic differences between cases with and without missing data. This

procedure imputes missing values based on the expectation-maximization method and tests whether means, covariances, and correlations of imputed values of missing cases differ significantly from those based on observed values of non-missing cases. Results did not support differences for externalizing [$\chi^2(204) = 216.6, p = .26$] or internalizing [$\chi^2(204) = 194.6, p = .67$].

Model testing procedures. Structural equation models were conducted in AMOS 6.0 using full information maximum likelihood. Model fit was assessed based on three criteria: 1) model χ^2 with a $p > .05$, 2) root mean square error of approximation (RMSEA) $< .05$, and 3) comparative fit index (CFI) $> .90$. Nested models were used to test changes in model fit when paths were added or constrained; significant changes in fit were indicated by a change in $\chi^2 > 3.84$ and a significant coefficient for any added path ($p < .05$). All models were run separately for externalizing and internalizing problems.

First, confirmatory factor analyses (CFAs) were run based on ratings from all informants at both ages. After base models were specified, significant error covariances were included across time for measures completed by the same informant type so that the estimates of factor stability were not inflated due to the stability of informant specific error. Significant within-time error covariances for informants that used the same questionnaire version also were allowed, in order to increase confidence that ratings loaded on the problem factors due to relations with the underlying constructs rather than shared method variance.

Second, eight models (four informants by two problem syndromes) tested the predictive validity of ratings by each informant at age 3 as the sole predictor of the latent problem behavior factor at age 5. To control for informant-specific variance that was shared across time, ratings by the informant that was considered as the age 3 predictor were covaried with the error term of the ratings by that informant at age 5. Only error covariances among the age 5 indicators that were significant were allowed (to account for shared variance due to method).

Third, two models (one for each problem syndrome) tested the incremental utility of different informants' reports by simultaneously examining each as an independent predictor of the early school age problem behavior factor. Significant covariances among age 3 predictors were allowed (as is convention), as were significant error covariances among age 5 outcomes (to account for shared method), and significant cross-time, informant-specific covariances (to account for error unique to informant).

Results

Means and standard deviations for untransformed problem ratings provided by mothers, fathers, teachers, and examiners are reported in Table 1. Teachers contributed higher externalizing and internalizing problem ratings than did examiners at ages 3 [$t(181) = 7.33$ and 8.95 , respectively, $p < .001$] and 5 years [$t(176) = 4.50$ and 2.89 , respectively, $p < .005$]. Internal consistencies of

Table 1 Means, standard deviations, and ranges of ratings of children's behavior problems provided by multiple informants at ages 3 and 5 years

	Externalizing		Internalizing	
	Mean (SD)	Range	Mean (SD)	Range
Age 3				
Mothers (<i>n</i> = 235)	11.5 (7.3)	0–39	6.5 (4.8)	0–27
Fathers (<i>n</i> = 157)	10.7 (6.7)	0–28	6.0 (4.3)	0–21
Teachers (<i>n</i> = 188)	10.0 (12.4)	0–57	5.8 (6.7)	0–39
Examiners (<i>n</i> = 227)	4.0 (5.3)	0–26	1.4 (2.1)	0–11
Age 5				
Mothers (<i>n</i> = 215)	6.6 (5.9)	0–35	3.7 (4.0)	0–24
Fathers (<i>n</i> = 156)	5.9 (5.4)	0–28	3.4 (3.4)	0–18
Teachers (<i>n</i> = 190)	4.4 (8.1)	0–49	2.5 (3.8)	0–29
Examiners (<i>n</i> = 203)	1.9 (4.9)	0–45	1.8 (3.1)	0–22

Note. Different forms of the CBCL (for ages 2–3 and 6–18) and C/TRF (for ages 2–5 and 6–18) were used at the 3- and 5-year assessments.

untransformed examiner reports at ages 3 and 5 were computed for externalizing (Cronbach's $\alpha = .87$ and $.92$, respectively) and internalizing ($\alpha = .70$ and $.81$, respectively).

Correlations among informants' ratings

Correlations among informants' problem ratings are reported in Table 2. At age 3, externalizing ratings made by fathers, teachers, and examiners were significantly intercorrelated; mothers' ratings of externalizing at age 3 were associated only with fathers'. At age 5, all informants' ratings of externalizing were significantly correlated. Across time, significant stability of individual differences in externalizing was observed for ratings within informant type and across most informant types.

Mothers', fathers', and teachers' internalizing problem ratings were intercorrelated at age 3 years, while examiners' ratings related only to those made by teachers. At age 5, informants' internalizing ratings were significantly correlated, except for fathers' and teachers'. Significant cross-time stability of

individual differences in internalizing was evident primarily within informant type.

Convergence of informant ratings at ages 3 and 5 years

Externalizing behavior problems. Figure 1 depicts an adequately fitting CFA model based on externalizing problem ratings contributed by mothers, fathers, teachers, and examiners at ages 3 and 5 years. All loadings were positive and significant after accounting for the significant error covariance between teachers' and examiners' ratings within time, and cross-time, informant-specific error covariance for mothers', fathers', and teachers' ratings. Stability of individual differences on the externalizing factor was strong across time ($\beta = .80$, $p < .001$). Exploratory analyses did not support the ad hoc observation that mothers' ratings appeared to load more weakly on the problem factor at age 3 than at age 5 (i.e. model fit did not significantly worsen when factor loadings for mothers were constrained to be equal across time). However, an exploratory series of path constraints and nested model comparisons supported the ad hoc observation that factor loadings for fathers' ratings appeared strongest; specifically, fathers' ratings were found to relate more strongly to the externalizing factor than teachers' or examiners' (but not mothers') ratings.

Internalizing problems. Figure 2 illustrates a CFA model with satisfactory fit characteristics based on internalizing problem ratings. Except for examiners' ratings at age 3, all informants' ratings loaded significantly on the internalizing factors. Cross-time, informant-specific error covariance was significant for ratings by mothers, fathers, and examiners. Error covariance for teacher and examiner ratings was significant at age 3. Cross-time stability of individual differences on the internalizing factor was significant and of medium strength ($\beta = .39$, $p < .01$). Exploratory analyses supported that apparent differences in factor loadings were significant. Specifically,

Table 2 Correlations among ratings of externalizing (top-right of diagonal) and internalizing (bottom-left) problems by multiple informants

	1	2	3	4	5	6	7	8
Age 3								
1. Mother	–	.44***	.14	.13	.50***	.30***	.23**	.17*
2. Father	.38***	–	.35***	.25**	.40***	.64***	.35***	.18*
3. Teacher	.15*	.18*	–	.43***	.26***	.21*	.44***	.14
4. Examiner	.01	.06	.21**	–	.13	.17*	.27***	.20**
Age 5								
5. Mother	.33***	.16	.00	.04	–	.49***	.46***	.15*
6. Father	.15	.37***	–.02	–.01	.35***	–	.42***	.24**
7. Teacher	.00	.10	.19*	.19**	.28***	.16	–	.43***
8. Examiner	.04	.16	.09	.23**	.20*	.22**	.26***	–

Note. At age 3, *N*s ranged from 129 (fathers and teachers) to 227 (mothers and examiners); at age 5, from 126 (fathers and teachers) to 192 (mothers and examiners); and across time, from 122 [fathers (age 3) and teachers (age 5)] to 215 [mothers (age 3 and age 5)]. * $p < .05$; ** $p < .01$; *** $p < .001$.

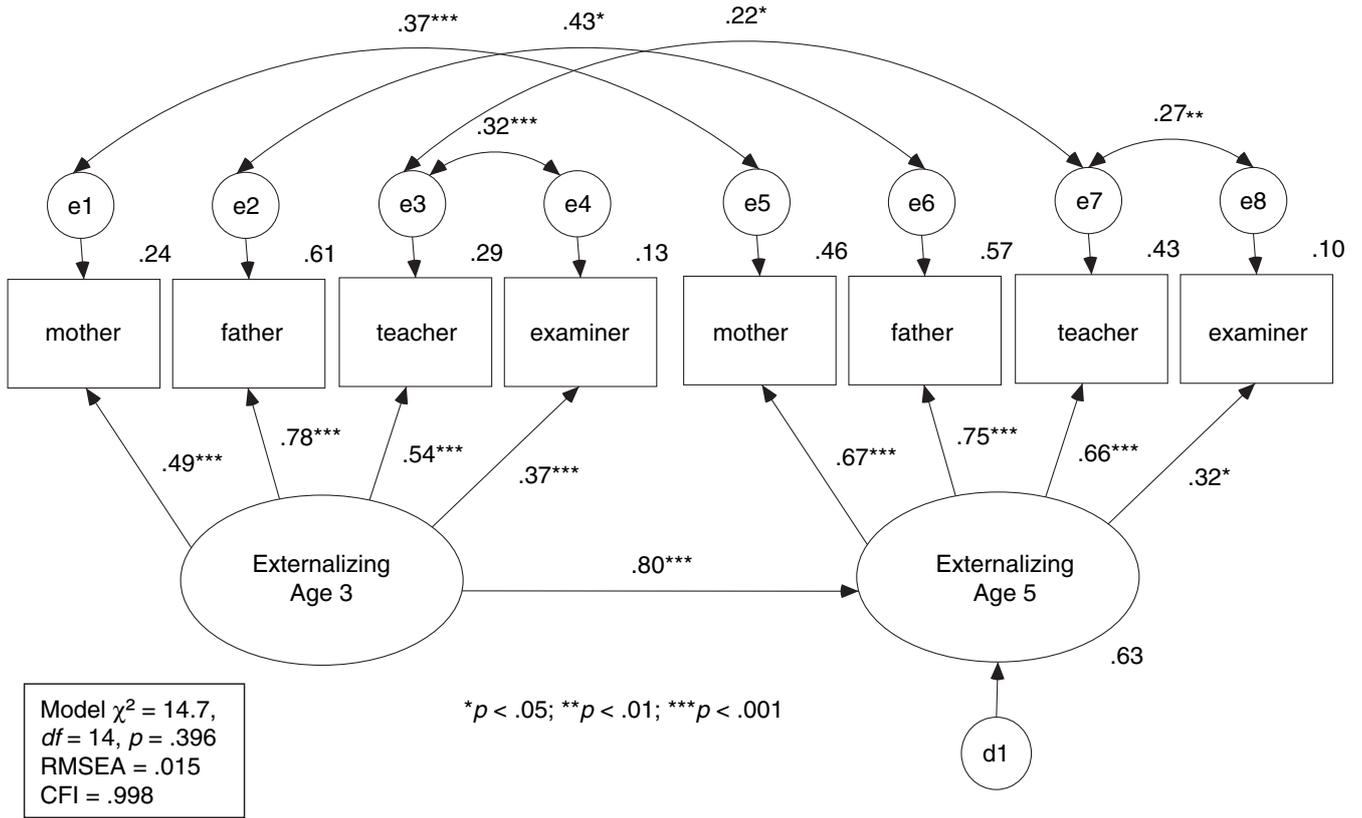


Figure 1 Confirmatory factor analysis of multiple informants' ratings of externalizing behavior problems at ages 3 and 5 years

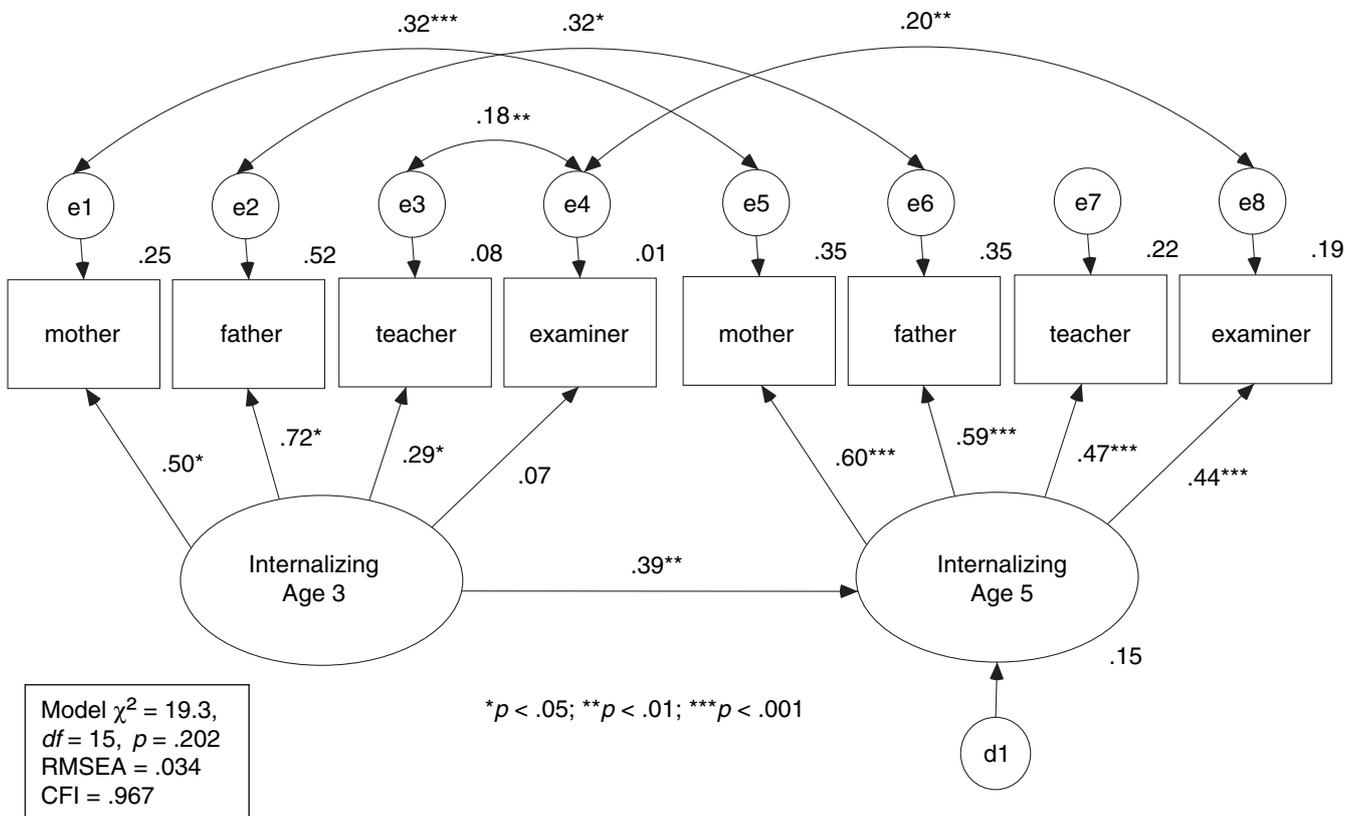


Figure 2 Confirmatory factor analysis of multiple informants' ratings of internalizing problems at ages 3 and 5 years

the model fit worsened significantly ($\Delta\chi^2 = 8.0$) when factor loadings for examiner ratings were constrained to be equal at both ages. Additionally, at both ages, fathers' ratings were found to relate more strongly to the internalizing factor than did teachers' or examiners' (but not mothers') ratings.

Predictive validity of ratings by multiple informants

Externalizing behavior problems. In separate predictive validity models by informant, age 3 ratings by mothers ($\beta = .42$), fathers ($\beta = .57$), teachers ($\beta = .44$), and examiners ($\beta = .31$) each were significantly associated with the age 5 externalizing factor defined by the four informants' ratings (all $p < .001$); models fit adequately.

To ascertain incremental validity, informants' ratings of externalizing problems at age 3 years were considered simultaneously as competing predictors of the age 5 externalizing problems factor defined by ratings from all four informants. In the final model [χ^2 ($df = 12$) = 16.3, $p = .18$; $RMSEA = .038$; $CFI = .987$], ratings by mothers ($\beta = .20$, $p < .05$), fathers ($\beta = .37$, $p < .001$), and teachers ($\beta = .37$, $p < .001$) each predicted the externalizing factor score at age 5, and together explained 52% of the variance in this outcome. Other details of the model were that the age 3 predictors all covaried significantly ($p < .05$); cross-time, within-informant error covariances were significant for mothers' ($r = .32$, $p < .001$) and fathers' ($r = .36$, $p < .001$) ratings; and the error covariance of teachers' and lab examiners' ratings at age 5 years was significant ($r = .28$, $p < .01$).

Internalizing problems. In separate predictive validity models by informant, only fathers' internalizing ratings at age 3 were associated with the internalizing factor at age 5 ($\beta = .31$, $p < .01$); age 3 ratings by mothers ($\beta = .15$), teachers ($\beta = .05$), and examiners ($\beta = .09$) were not significant predictors. All models fit adequately, though the model evaluating examiner ratings required a direct path between this predictor and teachers' ratings of internalizing at age 5 ($\beta = .18$, $p < .05$), which seemed reasonable since both informants made observations in non-home settings. Given these results, it was not surprising that the final incremental validity model for internalizing problems [χ^2 ($df = 17$) = 19.1, $p = .32$; $RMSEA = .023$; $CFI = .983$] indicated that only fathers' ($\beta = .33$, $p < .01$) ratings at age 3 years uniquely predicted the age 5 internalizing factor (11% explained variance). Also in the model was the direct path between age 3 examiner ratings and later teacher ratings ($\beta = .18$, $p < .01$); significant age 3 predictor covariances between mothers' and fathers' ratings ($r = .38$, $p < .001$), and between teachers' and examiners' ratings ($r = .19$, $p < .05$); and cross-time, within-informant covariances for mothers', fathers', and examiners' ratings ($r = .18$ to $.33$).

Discussion

In the current study, we examined child problem behaviors in an at-risk sample of children from preschool to early school age. Our main goal was to examine the stability of individual differences in externalizing and internalizing problem ratings of children across different settings and relationships (i.e. by mothers, fathers, teachers, and lab examiners), and between two developmental periods. Such associations would support the validity of ratings by different informants. Another aim of the study was to determine whether different informants' ratings of preschoolers' behaviors contribute predictive information regarding children's behavioral and emotional problems at early school age, and, if so, to what extent that information is unique. Findings offered new information on fathers and lab examiners, who have been underutilized as behavioral informants in past research on child emotional and behavioral concerns. Notably, study models accounted for the often significant informant- and method-specific variance shared among informants' ratings.

Cross-informant agreement

The pattern of cross-sectional correlations among informants' externalizing behavior ratings was generally consistent with the literature (Duhig, Renk, Epstein, & Phares, 2000). Confirmatory models suggested that significant variance in informants' ratings was explained by common externalizing factors at ages 3 and 5. Fathers' ratings of externalizing problems appeared to be more strongly associated with multi-informant problem factors than were ratings by teachers and examiners, a finding that strongly upholds calls for greater inclusion of fathers in research on development and psychopathology (e.g., Phares et al., 2005).

Correlations between mothers' and fathers' ratings of internalizing problems were medium in strength; correlations between ratings by other informant combinations were generally small to negligible at age 3, and small at age 5. Though not directly tested here, others have remarked on the lower levels of cross-informant agreement on internalizing than on externalizing problems (e.g., Glaser, Kronsoble, & Forkner, 1997; Grietens et al., 2004). CFA models supported the presence of internalizing factors that explained significant variance across situations and relationships, and across time; the only exception was that variations in examiners' observations at age 3 did not overlap significantly with what was common to observations by mothers, fathers, and teachers. At both ages, fathers' ratings of internalizing were again more strongly reflected in the underlying problem constructs than were ratings by teachers and examiners.

Importantly, the present study highlights the significant situational and temporal stability of

individual differences. On balance, however, in the majority of cases, the variance in informants' ratings that went *unexplained* by others' ratings was high. Though beyond the scope of the present study, research on the measurement of informant discrepancy, its correlates, and efforts to minimize it are reviewed by De Los Reyes and Kazdin (2005).

Temporal stability of informant ratings and problem factors

The temporal stability of individual differences in externalizing and internalizing, respectively, was large to medium for parents' ratings, medium to small for teachers' ratings, and small for examiners' ratings. These findings parallel prior research on rater agreement within and across time and context (Barkley, 1998; Duhig, et al., 2000; Grietens et al., 2004; Offord et al., 1996). The modest stability of individual differences in internalizing problems from ages 3 and 5 years may reflect changes in children's initially limited abilities to express internal distress verbally (Christensen, Margolin, & Sullaway, 1992; Kazdin, 1988) and/or changes in the degree to which such problems are situationally specific.

Temporal stability of individual differences in the externalizing and internalizing latent problem factors was large and medium, respectively. These findings on the presence of problem factors and stability in children's relative ordering on these factors over time are consistent with trait-like, or partially temperament-based problem behaviors (e.g., Olson et al., 2005; Gilliom & Shaw, 2004). The stability of internalizing problems must be taken in perspective, however, as 85% of the variance in the school-age problem factor was not explained by prior problems. Still, the successful derivation of factor scores reflecting situational and temporal stability in problem behavior is highly promising for developmental psychopathology researchers. Methodologically, this approach may permit better approximation of underlying constructs, by accounting for informant and method-specific error (e.g., Kraemer et al., 2003; Merrell, 1999; Offord et al., 1996). Conceptually, derivation of a multi-informant problem estimate is valuable for understanding problem antecedents and sequelae, since children who show disruptive behavior problems in multiple contexts are more likely to have serious and stable problems (Campbell et al., 2000).

Predictive validity of reports from specific informants

Ratings of preschoolers' externalizing problems by mothers, fathers, teachers, and examiners each predicted 9% to 33% of the variance in the latent problem factor at early school age, even after accounting for cross-time, informant-specific variance. Though modest in some cases, findings

support the predictive validity of all informant types. Furthermore, the incremental predictive utility of mothers', fathers', and teachers' ratings of age 3 externalizing was supported in relation to age 5 problems. Examiners offered valuable but redundant information.

Surprisingly, only fathers' ratings of internalizing at age 3 were associated with the internalizing factor at age 5, further supporting the value of father participation in developmental studies. Mothers' and teachers' observations tapped the internalizing problem factor at preschool age, but did not predict later internalizing problems that were observed across settings. Notably, a modest but statistically significant portion of the unique variation in teachers' internalizing ratings at age 5 overlapped with variance in prior ratings by examiners; perhaps features of the age 3 lab setting (e.g., testing demands) elicited the expression of difficulties that would manifest later in school.

The validity and utility of laboratory examiner ratings

A unique feature of this study was the extensive investigation of examiners' ratings of children's problems. These ratings showed adequate internal consistency at both ages and for both externalizing and internalizing problems. As predicted, examiners provided lower estimates of problem severity compared to teachers. Limited variance in internalizing ratings at age 3 may explain why examiner ratings did not relate to this problem factor; low sensitivity of examiner ratings to preschoolers' internalizing problems is consistent with findings using the Testing Observation Form (TOF; McConaughy & Achenbach, 2004). Remarkably, however, the associations between examiners' observations and those of other informants suggested that within a three-hour period with an unfamiliar adult, children showed patterns of misbehavior, inhibition, and negative affect that showed some consistency with their functioning in other settings and relational contexts. Moreover, externalizing behavior observed during the brief preschool laboratory setting explained variation in early school-age problems, though not independently from variation accounted for by parent observations. Yet, even overlapping information can be useful, especially when other informant data are missing. Results also support that using a multi-informant approach, examiner ratings can be used to partial out shared variance better attributed to error.

There is limited research on whether informants outside of the home and school settings provide meaningful information on children's behavior problems. For example, disruptive behavior problems are difficult to assess in a brief clinic visit (e.g., Shaffer, Lucas, & Richters, 1999), though ratings made by psychiatrists in young, non-clinical

samples have yielded promising findings (Clarke-Stewart, Allhusen, McDowell, Thelen, & Call, 2003). Externalizing ratings on the TOF (McConaughy & Achenbach, 2004) also have been found to correlate significantly with ratings by parents and teachers. Instruments that are normed in large samples and include items and scales most relevant to the testing context (e.g., TOF) may be more sensitive indicators for future research of this kind.

Limitations and considerations for future research

Limitations of the present study mainly relate to sample characteristics; children were predominantly European-American and from middle-class, two-parent families. Thus, our findings may not generalize to racially and ethnically diverse populations of children or to families experiencing significant contextual disadvantage. Additionally, although participants were recruited to oversample those with higher externalizing problems, the sample was not representative of clinic-referred youth, and may not have been sensitive to variation in internalizing problems.

Multi-informant assessment in early childhood research is challenging for several reasons. As reviewed in Phares (1997), study recruitment may be less likely to target fathers. Additionally, some young children have limited contact with adults. Many American children live in one-parent households (e.g., 28% lived in single-parent homes in 2004; Child Trends, 2004), and/or do not attend school or childcare (e.g., 39% of children under six years did not receive nonparental child care in 2001; Federal Interagency Forum on Child and Family Statistics, 2002). Excluding or less comprehensively assessing children without multiple informants may bias findings. Thus, the present results regarding the utility of fathers' and examiners' reports suggests that alternative, viable assessment options may be open to childhood researchers.

Acknowledgements

This research was supported by a grant from the National Institute of Mental Health (RO1MH57489) to Sheryl L. Olson and Arnold J. Sameroff.

The authors thank the study participants, laboratory examiners, and study administrators, especially Gail Benninghoff and Meribeth Pezda, as well as Lee Owen and Dave DeGarmo of Oregon Social Learning Center for statistical consultation.

Correspondence to

David Kerr, Oregon Social Learning Center, 10 Shelton McMurfhey Blvd, Eugene OR 97401, USA; Email: davidk@oslc.org

References

- Achenbach, T.M. (1992). *Manual for the Child Behavior Checklist/2-3 and 1992 Profile*. Burlington, VT: University of Vermont Department of Psychiatry.
- Achenbach, T.M. (1997). *Teacher/caregiver REPORT FORM for ages 2-5*. Burlington, VT: University of Vermont, Department of Psychiatry.
- Achenbach, T.M., Edelbrock, C., & Howell, C.T. (1987). Empirically based assessment of the behavioral/emotional problems of 2- and 3-year-old children. *Journal of Abnormal Child Psychology*, 15, 629-650.
- Achenbach, T.M., & Rescorla, L.A. (2001). *Manual for the Child Behavior Checklist/6-18*. Burlington, VT: University of Vermont Department of Psychiatry.
- Anderson, K.E., Lytton, H., & Romney, D.M. (1986). Mothers' interactions with normal and conduct-disordered boys: Who affects whom? *Developmental Psychology*, 22, 604-609.
- Barkley, R.A. (1998). *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment* (2nd edn). New York: Guilford Press.
- Campbell, S.B., Shaw, D.S., & Gilliom, M. (2000). Early externalizing behavior problems: Toddlers and preschoolers at risk for later maladjustment. *Development and Psychopathology*, 12, 467-488.
- Child Trends. (2004). *America's families and living arrangements: 2004. Calculations of U.S. Census Bureau, Current Population Survey, 2004: Annual Social and Economic Supplement. Table C-2*. Washington, DC: US Census Bureau.
- Christensen, A., Margolin, G., & Sullaway, M. (1992). Interparental agreement on child behavior problems. *Psychological Assessment*, 4, 419-425.
- Clarke-Stewart, K., Allhusen, V., McDowell, D., Thelen, L., & Call, J. (2003). Identifying psychological problems in young children: How do mothers compare with child psychiatrists? *Journal of Applied Developmental Psychology*, 23, 589-624.
- De Los Reyes, A., & Kazdin, A.E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*, 131, 483-509.
- Duhig, A.M., Renk, K., Epstein, M.K., & Phares, V. (2000). Interparental agreement on internalizing, externalizing, and total behavior problems: A meta-analysis. *Clinical Psychology: Science and Practice*, 7, 435-453.
- Fagot, B.I., & Leve, L.D. (1998). Teacher ratings of externalizing behavior at school entry for boys and girls: Similar early predictors and different correlates. *Journal of Child Psychology and Psychiatry*, 39, 555-566.
- Federal Interagency Forum on Child and Family Statistics. (2002). *America's children: Key national indicators of well-being*. Washington, DC: US Government Printing Office.
- Gazelle, H., & Ladd, G.W. (2003). Anxious solitude and peer exclusion: A diathesis-stress model of internalizing trajectories in childhood. *Child Development*, 74, 257-278.
- Gilliom, M., & Shaw, D.S. (2004). Codevelopment of externalizing and internalizing problems in early

- childhood. *Development and Psychopathology*, 16, 313–333.
- Glaser, B.A., Kronsoble, K.M., & Forkner, C.B.W. (1997). Parents and teachers as raters of children's problem behaviors. *Child and Family Behavior Therapy*, 19, 1–13.
- Glutting, J.J., Youngstrom, E.A., Oakland, T., & Watkins, M.W. (1996). Situational specificity and generality of test behaviors for samples of normal and referred children. *School Psychology Review*, 25, 94–107.
- Grietens, H., Onghena, P., Prinzie, P., Gadeyne, E., Van Assche, V., Ghesquiere, P., & Hellinckx, W. (2004). Comparison of mothers', fathers', and teachers' reports on problem behavior in 5- to 6-year-old children. *Journal of Psychopathology and Behavioral Assessment*, 26, 137–146.
- Hayden, E.P., Klein, D.N., & Durbin, C.E. (2005). Parent reports and laboratory assessments of child temperament: A comparison of their associations with risk for depression and externalizing disorders. *Journal of Psychopathology and Behavioral Assessment*, 27, 89–100.
- Kagan, J., Snidman, N., McManis, M., Woodward, S., & Hayward, C. (2002). One measure, one meaning: Multiple measures, clearer meaning. *Development and Psychopathology*, 14, 463–475.
- Kazdin, A.E. (1988). The diagnosis of childhood disorders: Assessment issues and strategies. *Behavioral Assessment*, 10, 67–94.
- Kraemer, H.C., Measelle, J.R., Ablow, J.C., Essex, M.J., Boyce, W.T., & Kupfer, D.J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *The American Journal of Psychiatry*, 160, 1566–1577.
- Loeber, R., Green, S.M., & Lahey, B.B. (1990). Mental health professionals' perception of the utility of children, mothers, and teachers as informants on childhood psychopathology. *Journal of Clinical Child Psychology*, 19, 136–143.
- McConaughy, S.H., & Achenbach, T.M. (2004). *Manual for the Test Observation Form for Ages 2–18*. Burlington, VT: University of Vermont, Center for Children, Youth, & Families.
- Merrell, K.W. (1999). *Behavioral, social, and emotional assessment of children and adolescents*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Moffitt, T.E., & Caspi, A. (2001). Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. *Development and Psychopathology*, 13, 355–375.
- NICHD Early Child Care Research Network. (2004). Trajectories of physical aggression from toddlerhood to middle childhood. *Monographs of the Society for Research in Child Development*, 69, vii–129.
- Offord, D.R., Boyle, M.H., Racine, Y., Szatmari, P., Fleming, J.E., Sanford, M., & Lipman, E.L. (1996). Integrating assessment data from multiple informants. *Journal of American Academy of Child and Adolescent Psychiatry*, 35, 1078–1085.
- Olson, S.L., Sameroff, A.J., Kerr, D.C.R., Lopez, N.L., & Wellman, H.M. (2005). Developmental foundations of externalizing problems in young children: The role of effortful control. *Development and Psychopathology*, 17, 25–45.
- Phares, V. (1997). Accuracy of informants: Do parents think that mother knows best? *Journal of Abnormal Child Psychology*, 25, 165–171.
- Phares, V., Fields, S., Kamboukos, D., & Lopez, E. (2005). Still looking for Poppa. *American Psychologist*, 60, 735–736.
- Saudino, K.J., Ronald, A., & Plomin, R. (2005). The etiology of behavior problems in 7-year-old twins: Substantial genetic influence and negligible shared environmental influence for parent ratings and ratings by same and different teachers. *Journal of Abnormal Child Psychology*, 33, 113–130.
- Shaffer, D., Lucas, C.P., & Richters, J.E. (1999). *Diagnostic assessment in child and adolescent psychopathology*. New York: Guilford Press.
- Stanger, C., & Lewis, M. (1993). Agreement among parents, teachers, and children on internalizing and externalizing behavior problems. *Journal of Clinical Child Psychology*, 22, 107–115.

Manuscript accepted 4 April 2007