

Defining Physical Aggression: A Developmental Perspective

At first glance, physical aggression seems to refer to a straightforward group of behaviors that most people would agree on. However, when trying to define physical aggression in very young children for the purpose of observational research, several issues are yet to be addressed in detail. Early childhood is characterized by developmentally appropriate limitations in motor skills and cognitive abilities, as well as age-specific play behaviors that need to be taken into account when defining physical aggression for this developmental stage. In other words, we need a *developmental* definition of physical aggression to capture the age-specific manifestations of this behavior, and to obtain a detailed description of which behaviors should and should not be considered aggressive.

The most salient issue in trying to formulate such a developmental definition is that of “intent.” Most definitions of (physical) aggression include the *intention* to inflict hurt or harm to others [e.g. Brook et al., 2001; Estrem, 2005; Ostrov et al., 2004]. However, not only are intentions very hard to assess at any age [Hartup, 2005], they are particularly problematic when referring to behaviors in very young children. The ability to oversee the consequences of one’s behavior and to understand other people’s feelings does not develop fully until the end of the preschool years [e.g. Zahn-Waxler et al., 1992]. Before that age, children may have learned the practical consequences of aggression, such as punishment, attention, or even the termination of a parental demand [Sroufe, 1995], but they are unable to fully gauge the effect on other people’s feelings, physical or emotional. They are physically able to show behaviors that may cause harm to others, as shown by several studies [Alink et al., 2006; Tremblay et al., 1999], regardless of the presence of an intention to hurt another person. And as Sroufe [1995] stated, young children who use aggressive behavior may not do so with the intention to hurt others, but those who are in some way disposed to be hostile and aggressive will have the tools needed to show such behaviors at a later age. This idea is confirmed by studies reporting significant longitudinal stability of early aggressive behaviors [Cummings et al., 1989; Keenan and Shaw, 1994], even in 1-year-old children [Alink et al., 2006]. Therefore, even if the behaviors are unintentional, early aggression appears to be developmentally relevant, supporting the choice to exclude *intent* from a definition of physical aggression in young children.

A second important issue in determining the manifestations of aggression in toddlers and preschoolers is that of age-appropriate behaviors. Behaviors caused by age-specific motor limitations and certain play or exploration behaviors may be mistaken for aggression. Young children can be very heavy-handed in their manipulation of objects or their interaction with other people, only because they have limited motor control. In addition, young children typically explore their environment to learn more about the functions and characteristics of objects. This may involve behaviors such as pushing, shaking, or even hitting, but the *context* of these behaviors is one reflecting play and exploration, rather than aggression. A developmental definition of physical aggression in early childhood needs to take these aspects of young children’s normative behavior into consideration. Consequently, observational measures of the aggression in young children need to take into account the specific context of relevant behaviors.

The third aspect of young children’s behavior that is relevant to a developmental definition of physical aggression concerns the overlap with other externalizing behaviors. For instance, temper tantrums are characterized by behaviors that are very similar to physical aggression, such as stamping feet and flailing arms. However, these specific behaviors usually are not aimed at anyone or anything in particular. They may be meant to convey some type of message to another person (e.g. for the mother to give in to the child’s demands), but the behaviors themselves are not necessarily physically directed at that person (picture an angry preschooler on the supermarket floor). Because temper tantrums are quite common in early childhood [Van Zeijl et al., 2006a], it is important to distinguish these behaviors from aggression that is physically aimed at and may actually harm another person.

Taking these developmental issues into account, we propose to define physical aggression in early childhood as behavior that is aimed at and may cause harm to people, animals, or objects, and is not because of motor limitations, or part of age-appropriate play and exploration.

The Observation of (Physical) Aggression in Early Childhood

Most studies of early childhood aggression rely on parent and teacher reports [Crick et al., 1997; Estrem, 2005; Russell et al., 2003; Tremblay et al., 1999, 2004]. Relatively few studies have used observational methods to assess (physical) aggression in young children, and even fewer have

observed aggression in a setting of mother–child interaction [Del Vecchio and O’Leary, 2006]; most observation studies focused on aggression among peers [Cummings et al., 1989; Ostrov et al., 2004; Strayer and Roberts, 2004]. Although peer-related aggression is certainly an important topic, theory and research suggest that coercive and aggressive interactions are likely to originate in the family. Early experiences with a rejecting, unresponsive, or uninvolved parent, as well as an insecure attachment relationship have been found to be related to early-onset conduct problems [e.g. McCartney et al., 2004; Shaw and Winslow, 1997]. In addition, parental reinforcement of aversive behaviors and the use of negative discipline strategies have been found to predict antisocial behavior [e.g. Del Vecchio and O’Leary, 2006; Eddy et al., 2001; Snyder and Stoolmiller, 2002]. These early experiences of parent–child interactions are thought to constitute a blueprint for social exchanges, which influences the child’s behavior in other social settings [Greenberg, 1999; Ramsey et al., 1990].

Consistent with this emphasis on early parent–child interactions in the understanding of conduct problems such as aggression, Keenan and Shaw [1994] developed an instrument to observe early childhood physical aggression in a laboratory situation involving mothers and their children: the System for Coding Early Physical Aggression (SCEPA). Although they did not provide an explicit definition of physical aggression, their coding instructions reflect several implicit assumptions regarding the developmental issues discussed above [Shaw, personal communication, August 22, 2003]. First, the instructions state that intent should not be inferred, and that only behaviors should be coded. Second, the manual emphasizes that it is important to determine whether the child is playing rather than acting aggressively. Third, behaviors that are part of temper tantrums are only coded as aggressive if behaviors such as kicking or hitting are explicitly aimed at something or someone in particular. The instrument, therefore, includes specific instructions regarding each of the salient developmental issues in defining physical aggression in early childhood. Keenan et al. [1998] reported that aggression observed with the SCEPA at age 18 months significantly predicted observed aggression at age 24 months. In addition, aggression observed at 24 months predicted mother-reported externalizing problems at 36 months, but only for boys [Shaw et al., 1994]. However, it is unclear whether the SCEPA shows different associations with aggressive versus nonaggressive externalizing problem beha-

viors as reported by parents. This issue needs to be addressed to establish the convergent and discriminant validity of the observation instrument.

Finally, results regarding gender differences in the rate of physical aggression in young children have been equivocal. Several studies using parent reports have found that boys show higher levels of physical aggression than girls before the age of 4 years [Alink et al., 2006; Baillargeon et al., 2005; Koot et al., 1997; Tremblay et al., 1999]. Conversely, studies using observational data failed to find significant gender differences in physical aggression in 2- and 3-year-olds [Cummings et al., 1989; Shaw et al., 1994]. This discrepancy may be because of biases in parents’ report as a result of gender-specific social expectations. However, it must be noted that both of the observational studies were based on small sample sizes, which may have limited their power to detect potential gender differences.

This Study

In this study, we examine the reliability and validity of the SCEPA [Keenan and Shaw, 1994], an observational method to assess physical aggression in a large sample of 1- to 4-year-olds. We investigate (1) the association between observed physical aggression and maternal ratings of aggressive and nonaggressive externalizing problem behaviors, (2) the associations of age and gender with rates of observed physical aggression, and (3) the 1-year stability of observed and mother-rated physical aggression.

METHOD

The SCRIPT Study

The SCRIPT study (Screening and Intervention of Problem behavior in Toddlerhood) is a collaboration between Leiden University (Centre for Child and Family Studies) and the Vrije Universiteit Amsterdam (Department of Developmental Psychology). The study investigates the effectiveness of an early intervention program aimed at reducing externalizing problems in 1- to 3-year-old children by enhancing parental sensitivity and discipline strategies [Van Zeijl et al., 2006b]. The data for this study were derived from the pretest (Time 1) and posttest (Time 2) laboratory sessions.

Sample

Participants were recruited from community records of several cities and towns in the western

region of the Netherlands. Children born in a specific time period were selected in order to obtain a group of 1-, 2-, and 3-year-old children. Children were not eligible to participate in the screening phase if they had non-Dutch first names as well as non-Dutch family names (implying a possible lack of familiarity with the Dutch language and meeting exclusion criteria for the intervention phase regarding ethnic background). In the screening phase, parents of 4,615 children were sent questionnaire booklets by mail. We obtained 2,408 questionnaires from primary caregivers (response rate 52%). To ensure a homogeneous sample, only children living with two parents (with the biological mother as the primary caregiver and a father figure—biological or stepfather—as the second caregiver) were eligible for the intervention study (95% of the sample). This selection and the application of several other exclusion criteria (e.g. twins, serious medical condition in child or mother) resulted in the exclusion of 454 cases, leaving a target selection sample of 1,954 children. For each age group, children with scores above the 75th percentile on the Preschool Child Behavior Checklist [CBCL/1½-5; Achenbach and Rescorla, 2000] Externalizing Problems scale (age 1 year: scores ≥ 13 ; age 2 years: scores ≥ 19 ; age 3 years: scores ≥ 20) were selected for the intervention study.

Of the 438 selected families, parents of 246 children (56%) agreed to participate in the intervention study. During the intervention phase, 9 families withdrew from the study, leaving 237 children and their mothers in the final sample (87 1-year-olds, 75 2-year-olds, 75 3-year-olds). Fifty-six percent of the children were boys, over half of the children had siblings (59%), and 55% were first-borns. Mean age of the mothers was 33 years and half of the mothers had a high educational level (Bachelor's or Master's degree). There were no significant differences between selected families who agreed to participate in the entire intervention phase and those who did not regarding the initial level of child externalizing problems ($P = .99$), child and maternal age ($P = .18$ and $.07$), child sex ($P = .84$), and the presence of siblings ($P = .98$). The only statistically significant difference was that participating parents had a somewhat higher educational level than nonparticipating parents, $F(1, 434) = 12.70$, $P < .01$.

For the 237 children in the final sample, data regarding aggression were obtained at two time points. At Time 1 the mean ages of the children were 15.61 months for the first age group ($SD = 1.17$, range = 13.58–18.84), 27.61 for the second age group ($SD = 1.17$, range = 25.87–30.34), and 39.58

for the third age group ($SD = 1.05$, range = 37.11–41.91). At Time 2 the mean ages were 27.81 months for the first age group ($SD = 1.54$, range = 25.31–33.40), 40.28 for the second age group ($SD = 1.70$, range = 36.79–46.91), and 51.98 for the third age group ($SD = 1.30$, range = 49.87–56.97).

Measures

Observation of physical aggression. Physical aggression in toddlers and preschoolers was defined as behavior that is aimed at and may cause harm to objects or people, or animals (the last category being irrelevant to the present observations). The observation instrument SCEPA (see Appendix A) was based largely on the work by Keenan and Shaw [1994]. Specific behaviors coded as aggression included the following: hitting, kicking, biting, pinching, scratching, shaking, pushing, stamping, throwing, and physically threatening to perform any of these behaviors. The behaviors needed to be distinguished from (a) behavior caused by motor limitations, such as using force to place a heavy toy in a basket, and (b) play and exploration, such as shaking things to find out what happens. The intent to hurt or harm someone or something was not a requisite for coding aggression, as it cannot be observed directly and constitutes an aspect of aggression that is not age-appropriate for young children. Several factors need to be taken into account to establish whether the behavior should not be ascribed to motor limitations or play, including the context, the force, and the appropriateness of the behavior as well as the child's facial and verbal expressions. Each combination of the different aspects of these factors may yield a different judgement of the behavior. For example, a surprised facial expression combined with excited and positive vocalizations when hitting two wooden blocks against each other in the absence of frustration or anger would not be considered aggressive. On the other hand, hitting a wooden block against the chair with excessive force after having been told to clean up the toys in combination with an angry facial expression and/or angry vocalizations would be counted as aggressive. Behaviors not coded as physical aggression included screaming or cursing, temper tantrums or disobedience without explicit aggression, simply dropping objects without force, behaviors not aimed at anything or anyone in particular (such as flailing arms or stamping on the floor), and aggressive acts aimed at the child's own body. Consecutive aggressive behaviors were only recorded as separate behaviors if (a) there were 2 sec

or more between behaviors, and/or (b) the behaviors reflected different types of aggressive behaviors (e.g. hitting and kicking at the same time). Aggression was also coded if part of the behavior was not visible on videotape, but only audible (e.g. observing the hand being raised, followed by an audible but invisible bang). When visibility was inadequate to the extent that there was reasonable doubt about the nature of the behavior, aggression was not coded. If the behavior met all of these criteria, it was always coded as aggression, even if mother seemed to condone or encourage the behavior.

Physical aggression was observed in a laboratory setting at Time 1 and Time 2 during four episodes, including one neutral episode, and three potentially challenging or frustrating episodes. The neutral episode was the break in which mother and child had a drink and a snack (duration 5 min after which coding ended, even if the break was longer). The first challenging episode consisted of a "clean-up" task (duration 1–4 min: the episode was ended after 4 min, or when the child finished the task). The second challenging task was a "don't" task in which the child was not allowed to touch toys for 2 min, after which the child was only allowed to touch the least attractive toy for another 2 min. For 1-year-olds the duration was two times 1.5 min, instead of 2 min (total duration of don't task: 3 or 4 min). In the third challenging episode, mother and child were asked to solve tasks that were somewhat difficult considering the age of the child, using different but functionally similar play material for each age group. Dyads were given three problem-solving tasks at Time 1 and two tasks at Time 2 consisting of a construction task (Times 1 and 2), a puzzle (Times 1 and 2), and a sorting task (only at Time 1) for 5 min per task. Mothers were instructed to help their children in the way they would normally do. For each episode, the frequency of object-directed and mother-directed aggression was computed. These were summed to form total aggression frequencies across episodes for object- and mother-directed aggression separately. Because the duration of the clean-up task and the attractive toys task varied, the raw frequencies of aggression were divided by the actual duration of the task in minutes and multiplied by four (the standard duration of each of the two tasks). This way, the total observation time was the same for all children: 28 min for Time 1 and 23 min for Time 2. Because the frequency scores do not account for variations in the intensity of the aggressive behaviors, we also assigned global ratings based on the frequency as well as the intensity of the aggressive behaviors, using a scale from 1 (*not*

aggressive) to 5 (*very aggressive*) for both object- and mother-directed aggression.

When we first developed the instrument, only the two frustration episodes (clean-up and not touching toys) and the neutral break episode were used. Computed across these episodes, the first and second author coded 15 tapes from this study for each of the three age groups (1-, 2-, and 3-year-olds, corresponding to 17–20% of the total number of tapes for each age group). They reached intercoder reliabilities (intraclass correlations, single rater, absolute agreement) for the frequencies and global ratings (object- and mother-directed) ranging from .75 to .97. The first author coded another set of 30 tapes from this study to use as training tapes. Seven Master's students were trained in eight weekly sessions, during which videotapes were discussed that had been coded as "homework" by the training participants (three tapes per session). After the eight training sessions, the participants were assigned a specific age group and coded the same 15 tapes of children of that age that were coded by both the first author in an earlier stage. Intraclass correlations (single rater, absolute agreement) were computed for the frequencies (object-, and mother-directed and total) and global ratings (object- and mother-directed), yielding 5 correlations per coding pair. One student was assigned to the 1-year-olds (an age group only represented in the Time 1 assessment) and reached reliabilities of .80 to .91 with the first author. Two students were assigned to the 2-year-olds (represented in both Time 1 and Time 2) and reached reliabilities of .70 to .97 with the first author and each other (3 pairs of coders). Five students were assigned to the 3-year-olds (represented in both Time 1 and Time 2, and also applied to the 4-year-olds at Time 2) and reached reliabilities of .73 to .98 with the first author and each other (15 pairs of coders).

After the first group of trained coders had completed their work on the two frustration episodes and the neutral break episode, we decided that a longer observation period could enhance the quality of the data. To maximize added value, we chose to include the task-episode, which was less frustrating than the clean-up and don't-touch episodes, but more challenging than the neutral break. The second author proceeded to code these episodes for 45 tapes (again 15 per age group) and subsequently trained five new Master's students for this episode according to the same format described above. Two students coded 1-year-olds (3 pairs, intraclass correlations .70 to .90), three students coded 2-year-olds (6 pairs, $r_s = .78$ to 1.00), and four

students coded 3-year-olds (10 pairs, $r_s = .63$ to 1.00 , with only two $< .70$).

In the final data set, the codes for all tapes used for training and reliability were those given by the first or second author. In those rare cases when trained coders were unsure about how to code a certain behavior after the reliability phases, they showed the tapes to the first or second author to discuss the behavior and decide on a final coding together.

Mother-rated physical aggression. The Physical Aggression Scale for Early Childhood [Alink et al., 2006] was completed by mothers at Time 1 and Time 2 at the end of the lab sessions. The questionnaire consisted of 11 items concerning physical aggression, including behaviors such as hitting, biting, and destroying things. Parents were asked whether their child had shown these behaviors during the past 2 months. The items were scored on a 3-point Likert scale (0 = not true, 1 = somewhat or sometimes true, 2 = very true or often true). A physical aggression score was computed by summing the item scores (potential score range 0–22). Internal consistencies of the total physical aggression score were computed for both Time 1 and Time 2, separately for each age group. Cronbach's α s ranged from .71 to .87.

Mother-rated nonaggressive externalizing problems. The CBCL for 1½- to 5-year-old children [CBCL/1½-5; Achenbach and Rescorla, 2000] was used to assess nonaggressive externalizing problems and was obtained at Time 1 and Time 2. The previous version of the CBCL/11½-5 (the CBCL/2-3) was tested in a Dutch population of 2- to 3-year-olds by Koot et al. [1997] who identified a broadband Externalizing Problems syndrome (31 items) consisting of three narrowband syndromes: Oppositional (19 items), Aggressive (7 items), and Overactive (5 items). Koot et al. reported good reliability and validity. Evidence for the reliability and validity of the CBCL/1½-5 in 1-year-old children under age 18 months was presented by Van Zeijl et al. [2006a]. For this paper, we used a Nonaggressive Externalizing Problems scale obtained by summing only the items from the Oppositional and Overactive narrowband scales (e.g. can't sit still, can't wait, stubborn, sulks). These scales did not include any items referring to physical aggression. Internal consistencies of the Nonaggressive Externalizing Problems scale were computed for both Time 1 and Time 2, separately for each age group. Cronbach's α s ranged from .80 to .88. We used the Physical Aggression Scale for Early Childhood rather than the CBCL Aggressive syndrome to assess mother-rated aggressive behavior, because

the CBCL Aggressive syndrome also includes a number of nonaggressive behaviors in both the Dutch (three out of nine items) and the American version (15 out of 19 items).

Procedure

Participating families were invited for a pretest (Time 1) in the laboratory. During the 1.5 hr laboratory session, mother and child completed several tasks. These sessions were videotaped with cameras that were fixed to the walls and were operated from behind the one-way screen so that no third person was present in the room. The videotapes were coded afterwards by coders unaware of experimental condition. During the lab session, mothers were asked to fill in some questionnaires. After the pretest, families were randomly assigned to either a control ($n = 117$) or an intervention ($n = 120$) group. There were no differences between the two groups regarding the initial level of child externalizing problems ($P = .13$), parental educational level ($P = .46$), child and maternal age ($P = .85$ and $.97$), and the presence of siblings ($P = .67$). The only statistically significant difference was the percentage of girls, which was higher in the intervention group (51%) compared with the control group (38%), $\chi^2(1, N = 237) = 4.20, P < .05$. Families in the intervention group received six home visits and, parallel in timing, families in the control group received six telephone calls. Approximately 1 year after the pretest ($M = 12.41$ months, $SD = 1.14$, range = 8.25–19.49), families from both the intervention and control group visited the laboratory for the posttest (Time 2), which used the same procedures as the pretest. All participating parents signed an informed consent form.

Statistical Analyses

There were some missing data (1 case for Time 2 observed aggression, 2 different cases for Time 1 and Time 2 mother-rated externalizing problems). These missing data were substituted with the mean score on the variable for children with the same sex, age, parental educational level, and for Time 2 variables were also matched for experimental condition.

RESULTS

Preliminary Analyses

The descriptive statistics for the observation and mother-reported variables are summarized in Table I. The statistics show that mother-directed

physical aggression occurs infrequently compared with object-directed aggression, with 74% of children not showing any physical aggression directed at the mother at all. Paired samples *t*-tests showed that the difference in means between object- and mother-directed physical aggression was significant for both the frequencies, $t(236) = 9.48, P < .01$, and the global ratings, $t(236) = 10.76, P < .01$. The correlations between observed object-directed and mother-directed aggression were .38 at Time 1 and .21 at Time 2 (both P s < .01). We also dichotomized the object-directed and mother-directed variables to differentiate between children who showed no aggression at all and children who showed at least one aggressive act. Cross tabulations of the dichotomous variable for object- and mother-directed aggression revealed significant associations between the two at Time 1 (Odds ratio = 2.95, 95% confidence interval: 1.44–6.05) and for Time 2 (Odds ratio = 5.20 95% confidence interval: 2.16–12.53). The very low occurrence of mother-directed aggression is likely to hamper reliable conclusions about this subtype of aggression, and because object-directed and mother-directed aggression were significantly interrelated, we decided to focus on the total frequency of observed physical aggression and the average of the ratings for object- and mother-directed physical aggression. We computed cross-sectional correlations between these two variables for both Time 1 and Time 2. Results showed that the correlations between the total frequencies and ratings of total physical aggression were .94 for Time 1 and .91 for Time 2. Because of these very high correlations, only one of these measures was used for further analyses. We felt that the total frequencies would be most informative, because these refer closely to the real numbers of aggressive acts and are therefore easier to interpret. Thus, analyses in this paper will be

TABLE I. Descriptive Statistics for All Variables at Time 1 (N = 237)

	Range	M (SD)	% with score "0"
Frequency observations			
Object-directed	0–20	2.23 (2.96)	33
Mother-directed	0–13	0.54 (1.33)	74
Total	0–26	2.77 (3.68)	29
Global rating observations			
Object-directed	0–4.0	0.65 (0.72)	33
Mother-directed	0–2.5	0.19 (0.40)	74
Average	0–2.8	0.42 (0.47)	29
Mother reports			
Aggression	0–17	4.97 (3.36)	33
Nonaggressive externalizing	4–41	19.92 (6.72)	0

based on the total frequencies of observed physical aggression.

Outliers ($|z| > 3.29$) were found for observed physical aggression at Time 1 ($n = 6$) and Time 2 ($n = 4$), for Time 2 mother-rated physical aggression ($n = 2$) and nonaggressive externalizing problems ($n = 1$). All 13 outliers represented 13 different children. The outliers were winsorized [i.e. “moved in close to the good data”; Hampel et al., 1986; p 69] by replacing the outlying scores with the next highest value of the remaining distribution.

Age and Gender Differences

Table II shows the means and standard deviations for observed and mother-rated physical aggression by age group and gender at Time 1. A multivariate analysis of variance (MANOVA) was performed, examining the main effects of age group and gender, as well as an age group \times gender interaction term on the aggression variables. In addition, three separate MANOVAs were performed to investigate gender differences for each age group. Across age groups, we found significant main effects of gender, $F(3, 229) = 5.05, P < .01$, with boys showing higher levels of physical aggression than girls for both measures of aggression. There was also a main effect of age group, $F(6, 458) = 6.28, P < .01$, but the between-subject results indicated that this was only significant for mother-rated physical aggression, with post hoc tests showing that 1-year-olds had significantly lower scores compared with 2- and 3-year-olds

TABLE II. Means and Standard Deviations for Time 1 Observed and Mother-Rated Physical Aggression: Effects of Age and Gender

			Observation	Mother rating
Age 1 year				
	<i>n</i>			
Total	87	<i>M</i> (SD)	2.44 (2.87)	3.49 (2.52)
Boys	48	<i>M</i> (SD)	2.94 (3.02)	4.01 (2.69)
Girls	39	<i>M</i> (SD)	1.82 (2.59)	2.85 (2.15)
	Gender difference		$F = 3.37$	$F = 4.83^*$
Age 2 years				
	<i>n</i>			
Total	75	<i>M</i> (SD)	3.13 (3.48)	6.23 (3.16)
Boys	49	<i>M</i> (SD)	3.20 (3.54)	6.43 (3.09)
Girls	26	<i>M</i> (SD)	3.01 (3.43)	5.85 (3.31)
	Gender difference		$F = 0.05$	$F = 0.58$
Age 3 years				
	<i>n</i>			
Total	75	<i>M</i> (SD)	2.29 (2.53)	5.43 (3.79)
Boys	35	<i>M</i> (SD)	2.95 (3.17)	6.77 (4.06)
Girls	40	<i>M</i> (SD)	1.71 (1.64)	4.25 (3.14)
	Gender difference		$F = 4.69^*$	$F = 9.16^{**}$
Main effect gender			$F = 4.68^*$	$F = 12.09^{**}$
Main effect age			$F = 1.53$	$F = 16.75^{**}$
Age \times Gender			$F = 0.67$	$F = 1.89$

* $P < .05$; ** $P < .01$.

($P < .01$ in both comparisons). The separate MANOVAs revealed main effects of gender for 1-year-olds, $F(3, 83) = 3.28$, $P < .05$, and for 3-year-olds, $F(4, 70) = 3.09$, $P < .05$, but not 2-year-olds, $F(3, 71) = 0.26$, $P = .86$. The gender-effects (i.e. boys scoring higher than girls) were significant for both the observational measure and the mother reports in 3-year-olds, but only for mother ratings in 1-year-old children. However, there were no significant interaction effects between gender and age group for either of the measures.

Convergent and Discriminant Validity

To investigate convergent and discriminant validity of the SCEPA, we examined the correlations between three variables: observed aggression, mother-reported aggression, and mother-reported nonaggressive externalizing problems. Each correlation addresses a different issue regarding method variance and validity. First, the associations between observed and mother-reported aggression test the convergent validity as it establishes the association between scores for the same construct as measured by two different methods. Second, the associations between observed aggression and mother-reported nonaggressive externalizing problems reflect a test of both a methodological and a construct difference. Third, the associations between mother-reported aggression and nonaggressive externalizing behaviors reflect a test of construct differences as the methods do not vary. We computed cross-sectional correlations for the total sample as well as per age group, for both Time 1 and Time 2. Table III shows that the total frequency of observed physical aggression was significantly related to mother-rated physical aggression in almost all subgroups (except for 1-year-olds at Time 1), and not at all related to mother-rated nonaggressive externalizing problems. We did find significant correlations between mother-reported aggression and nonaggressive externalizing problems for the total sample and for each age group ($r_s = .48-.61$, all $P_s < .01$).

Stability of Observed and Mother-Rated Physical Aggression

The 1-year stability of physical aggression was examined by computing correlations between Time 1 and Time 2 measures across and per age group. Because half of the families received an intervention between Time 1 and Time 2, only the control group was used for these analyses. No significant longitudinal correlations for observed physical aggression were found across age groups. $r(117) = -.05$,

TABLE III. Cross-Sectional Correlations between Observed Physical Aggression (Object-Directed+Mother-Directed) and Mother Ratings for Time 1 and Time 2, per Age Group

	Mother ratings	
	Physical aggression	Nonaggressive externalizing
All age groups ($n = 237$)		
Time 1 (age 1-3 years)	.28**	.10
Time 2 (age 2-4 years)	.24**	.08
Age group 1 ($n = 87$)		
Time 1 (age 1 year)	.17	.14
Time 2 (age 2 years)	.23*	.18
Age group 2 ($n = 75$)		
Time 1 (age 2 years)	.35**	.10
Time 2 (age 3 years)	.28*	.11
Age group 3 ($n = 75$)		
Time 1 (age 3 years)	.29*	.10
Time 2 (age 4 years)	.30**	.14

* $P < .05$; ** $P < .01$.

$P = .61$, for 1-year-olds, $r(44) = -.11$, $P = .48$, for 2-year-olds, $r(38) = .03$, $P = .87$, or 3-year-olds, $r(35) = -.05$, $P = .79$. The 1-year stability of mother-rated physical aggression was significant for all ages (all $P < .01$), with correlations of .59 for 1-year-olds, .52 for 2-year-olds, and .51 for 3-year-olds. We also examined the stability of aggression separately for boys and girls. Again, we only found significant longitudinal correlations for mother-reported aggression (boys $P < .01$, girls $P < .05$), and not for the observational measure ($P_s > .38$). We repeated our stability analyses using the total sample correcting for experimental condition. These analyses did not yield different results.

DISCUSSION

The results of this study show that physical aggression in 1- to 4-year-olds can be reliably assessed in a laboratory setting, using the observational measure originally developed by Keenan and Shaw [1994]. For 2- to 4-year-olds, observed physical aggression was significantly related to mother-rated physical aggression. This was not the case for 1-year-olds. For all age groups, observed physical aggression was not related to maternal ratings of nonaggressive externalizing problems. Significant 1-year stability was found for mother-rated but not for observed physical aggression.

The observation instrument for the assessment of physical aggression in toddlers and preschoolers used in this study, the SCEPA, includes clear rules about which behaviors should and should not be considered as such, taking into account developmentally relevant issues. On the basis of these rules, we formulated an explicit definition of physical aggression in early childhood, something that has been lacking in the literature to date. The definition and rules were successful in that the intercoder reliability was high, also for the youngest age groups. Apparently, the coders managed to distinguish between physical aggression on the one hand, and play and nonaggressive externalizing behaviors on the other hand. Furthermore, the exclusion of "intent" from the definition of physical aggression may have facilitated intercoder agreement. Although intercoder reliabilities were high for all separate measures (i.e. for both frequencies and ratings, as well as for mother-directed and object-directed aggression), we decided to perform our analyses using only the total frequency measure. This decision was based on the fact that mother-directed aggression was very rare and therefore showed little variance. Nevertheless, we feel that the distinction between mother-directed (or more generally person-directed) and object-directed physical aggression is important in that they may be associated with different causes and consequences. In settings or samples where higher rates of person-directed physical aggression are expected (e.g. in peer-setting or in multiple-risk samples), both subtypes of aggression should ideally be analyzed separately. Further, the global ratings of physical aggression were discarded in favor of the frequency measure because of the high correlation between these two measures at both assessments ($>.90$), which made the inclusion of both measures superfluous. A re-examination of our global rating scale revealed that it relied heavily on the frequency of aggression, with each scale point defined in terms of a certain frequency range. This may be why the inclusion of severity in the global score did not add much unique information. It is also possible that the frequency and severity of aggression are highly correlated and that knowing the frequency of the behavior is also a strong indication of its severity. A more detailed examination of these issues is needed to decide whether a (modified) global score is a useful addition to the SCEPA.

The concurrent convergent validity of the SCEPA was established for 2- to 4-year-olds, but not for 1-year-olds. Because of the cross-sequential nature of the data, we were able to replicate some of our

findings for Time 1 by examining the results for Time 2 (1 year after Time 1, with ages partly overlapping with those of Time 1). Thus, the results for the 2- and 3-year-olds were established not only at Time 1, but also at Time 2 for children originally aged 1 year and 2 years. Using the cross-sequential data, we were also able to show that the lack of association between observed and mother-reported physical aggression in 1-year-olds was not because of characteristics of the subsample of 1-year-olds. The associations between observations and mother ratings found at Time 2 when these children were 2 years old were similar to those found for 2-year-olds at Time 1. The lack of association between observations and mother ratings in 1-year-olds is therefore likely to reflect other issues than selective sample characteristics.

Mothers of 1-year-old children may have differed in their rating of physically aggressive behaviors such as hitting and kicking for this age group. Parents may feel disinclined to assign this type of behavior to such young children, because they feel the terminology is inappropriate to this age group. Some mothers of 1-year-olds, indeed, wrote down remarks to this effect on the questionnaires. Some stated that the behaviors were not applicable because the child did not do them on purpose, or that the behaviors were much too severely stated to apply to such young children. It seems that the description of physical aggression in 1-year-olds leads to major differences in interpretation by mothers, but not by independent and trained observers. For instance, some mothers may have applied the rule of intent to their rating of aggressive behaviors (even though this was not mentioned in the instructions), whereas others may have taken the items at face value, without trying to infer intent. These discrepancies between mothers of 1-year-olds may have led to the absence of a significant association between observed and mother-rated physical aggression in this age group. Additional instructions for mothers regarding the interpretation of the items of the questionnaire in terms of intent may enhance the convergent validity of the SCEPA in young children.

In addition to evidence for the convergent validity of the SCEPA, the results also indicated discriminant validity, as our observational measure of aggression correlated only with parent reports of aggression and not oppositional or overactive behavior. This means that of the two types of differences reflected in the comparison of observed aggression and mother-reported nonaggressive externalizing problems (i.e. a difference in method

and in construct), it is likely to have been the construct difference that is reflected in the lack of a significant correlation between the two variables. We did find a significant correlation between mother-reported aggressive and nonaggressive externalizing problems. Given this significant association using the same method, it is especially noteworthy that observed aggression was related only to mother-reported aggression. The often-found association between the two constructs may therefore be partly because of method variance. Discriminant validity is especially relevant in the case of physical aggression, which is hypothesized to have more severe negative outcomes than other forms of externalizing problems [Broidy et al., 2003]. In order to identify the specific developmental pathways, and risk and protective factors for physical aggression, measures need to be able to distinguish these behaviors from other related externalizing behaviors. The SCEPA seems to specifically measure physical aggression rather than oppositional or overactive behaviors. An additional test of the measure's discriminant validity would be to examine correlations with observations of non-aggressive externalizing behaviors.

For mother-reported physical aggression, 1-year-olds showed significantly less physical aggression than 2- and 3-year-olds, whereas no age effects were found for observed aggression. This finding may be related to the difference between daily family life and the frustration tasks in the laboratory setting. The demands made on the 1-year-olds' frustration tolerance during our laboratory observations may have exceeded those that they experience at home. For instance, many mothers of 1-year-olds indicated that their children were not used to having to clean up their toys at home. In daily life, mothers of 1-year-old children may also be more likely to place forbidden objects out of sight, whereas in the laboratory the forbidden toys were within the children's sight and reach. For 2- and 3-year-old children, the frustration tasks are more likely to resemble challenges that they face at home, implying higher ecological validity for this age group than for younger children. Thus, 1-year-olds may be less likely than older children to show high rates of aggression at home because their mothers do not yet put great demands on their frustration tolerance [see also Alink et al., 2006]. In the laboratory session, however, children in each age group were faced with the same challenges (although for a shorter time in the 1-year-old age group), which may have led to our finding of similar rates of physical aggression. This also indicates that 1-year-olds are just as

capable of showing physical aggression as 2- and 3-year-olds, if the situation is challenging enough. However, one may argue that the ecological validity of the observational measure for 1-year olds might be strengthened when the laboratory tasks are somewhat less challenging, and more similar to related tasks in the natural setting. Thus, in addition to a developmental definition of physical aggression, a developmental approach to task selection for the SCEPA may be particularly important for the youngest age group.

Regarding gender differences, boys were found to show higher levels of physical aggression than girls in 3-year-olds and the total sample for both mother ratings and observations. For mother-rated aggression, this gender difference was also found for 1-year-olds. There was no significant age by gender interactions for either measure. It must be noted that our sample was selected for showing high levels of externalizing problems, regardless of gender. This may have diminished the likelihood of finding strong and consistent gender differences in this study. On the other hand, the gender differences that we did find were therefore all the more salient. They show that even within a group of young children who were all reported by their mothers to display elevated levels of externalizing problems, boys show more physical aggression than girls. For the total sample of 1- to 3-year-olds, this was true for both mother-reported and observed physical aggression. This suggests that previous findings of gender differences in early childhood physical aggression based on parent reports [Alink et al., 2006; Baillargeon et al., 2005; Tremblay et al., 1999] may reflect true differences in aggression and not be completely because of informant biases. For 1- and 2-year-olds, it remains unclear whether gender differences are specific to parent reports, because there are relatively few observational studies of aggression at such a young age. Further research is needed to resolve this issue.

We did not find significant stability of observed physical aggression across a 1-year period for any of the age groups, in contrast to the highly significant 1-year stabilities for mother-rated physical aggression found in all age groups. One explanation for the lack of longitudinal stability may be that early childhood aggressive behavior in mother-child interactions simply is not very stable, at least not for all dyads. The only evidence for stability of *observed* aggression in the mother-child context comes from Keenan et al. [1998], who reported significant stabilities of .30 for girls and .25 for boys (both $P < .05$) from age 18 to 24 months. Two observational studies of early

childhood aggression in a peer setting showed mixed results. Cummings et al. [1989] found strong stability of physical aggression in a peer setting for boys ($r = .59$, $P < .01$), but not for girls ($r = .36$, $P < .10$) from age 2 to 5 years. However, in a study by Howes and Phillipsen [1998], observed aggression in a peer setting was not significantly stable from 12 to 48 months. More consistent evidence for the stability of early aggressive behavior comes from studies using parent reports [e.g. Alink et al., 2006; Koot et al., 1997; Van Beijsterveldt et al., 2003]. These results, however, need to be treated with caution because of the potential influence of stable rater bias. Overall, we conclude that there is relatively little evidence that aggression in early childhood is actually stable across time.

From a developmental perspective, a lack of stability in this type of behavior is understandable, especially in the context of mother-child interaction. The process of a growing need for autonomy that takes place in early childhood may be accompanied by (some) aggressive behavior in parent-child interactions. Whether this behavior is successfully dealt with by the parent is likely to influence the stability of aggression [Tremblay, 2003]. Thus, child aggressive behavior may be stable only for certain dyads and not for others. If child aggression in some dyads has not subsided at the end of the preschool years as a result of inadequate parenting strategies, this behavior is likely to become more predictable over time. Therefore, we need to investigate parenting moderators of the association between aggression at different ages in early childhood. Similarly, for early screening purposes, we may need to look at child aggression in combination with ineffective parental discipline.

Another explanation for the lack of stability found for the observation of physical aggression in this study may be the relatively short duration of the laboratory episodes used for this measure. Although mothers report on behaviors over a period of 2 months, the observations are based on approximately 25 min. The observation time is similar to the 22 min in the study by Shaw et al. [1994]. In the Cummings et al. [1989] study, the observation time was 66 min. Although the majority of observation time in our study was spent on frustration tasks, these were restricted to the presence of the mother, thus excluding the aggression of children shown in the presence of peers. Children who are highly aggressive in daily life across a 2-month period according to their mothers (or during an hour spent with a group of children with several available "targets") are likely to show

some stability in this behavior across time. However, a child who has shown aggression in the 25 min of the laboratory episodes at Time 1 may not necessarily show aggression again during the same small window of time and specific setting 1 year later, and vice versa. Thus, although the mother-rated aggression in our study is likely to refer to *trait* aggression, our observations may reflect mostly *state* aggression, and by definition the first is more stable than the latter.

To establish trait aggression in mother-child interactions by means of observation, studies may need more observation time and possibly a naturalistic setting. Other studies have observed parent-child interactions (including aggression) in a room made to resemble a family room for 1 hr on each of 10 nonconsecutive days [Snyder et al., 1994], and in the home during two 2-hr sessions [McFadyen-Ketchum et al., 1996], or a single 1-hr session [Eddy et al., 2001]. These procedures seem more likely to elicit trait-like physical aggression in young children than relatively short laboratory sessions. Observing aggression in the home with close-age siblings present may also enhance the measurement of trait aggression in the family context, especially when they have to share toys [Garcia et al., 2000]. Although we failed to find significant stability of observed physical aggression, longitudinal results of our study reported by Alink et al. [in press] showed that observed physical aggression was predicted by parenting behaviors in a theoretically meaningful way. These results suggest that the SCEPA does measure behaviors that are developmentally relevant in the context of parent-child interactions.

This study has some limitations. The first is the relatively low response rates in both the screening and the intervention phase, with selective nonresponse by families from lower socio-economic backgrounds. Further, the sample was selected for showing high levels of externalizing problems. These issues point to limited generalizability of the findings regarding the occurrence of physical aggression. However, the low response rates are less likely to have affected the results with respect to the convergent and discriminant validity of the SCEPA. Finally, the low occurrence of mother-directed aggression precluded separate analyses of this form of aggressive behavior. As mother-directed aggression is probably very rare in the general population, a clinical sample may be needed to successfully investigate this type of aggression.

In conclusion, the observational measure of physical aggression in toddlers and preschoolers (the SCEPA) as designed by Keenan and Shaw

[1994] shows promise: Intercoder reliability is high and the instrument distinguishes between mother-rated physical aggression and nonaggressive externalizing problems. Future studies are needed to examine whether longer observation times in naturalistic or laboratory settings will yield estimates of the rate of physical aggression that are stable across time. The emphasis on an age-specific definition of physical aggression in young children is particularly relevant to the field of developmental psychopathology, because it allows for the investigation of early pathways of aggression.

APPENDIX A SCEPA MANUAL

A.1. Definition and Specific Behaviors

Physical aggression is defined as behavior that is aimed at and may cause harm to objects or people (and if applicable: animals). An aggressive act needs to be distinguished from play, task oriented behavior, age-appropriate communication, or developmental motor limitations by:

- (a) the use of excessive/unnecessary force, and/or
- (b) a context that includes a direct or indirect cause for aggression, and/or
- (c) facial expressions or vocalizations that are compatible with aggression.

Behaviors coded as aggression	Behaviors NOT coded as aggression (if not accompanied by any of the behaviors on the left)
Hitting	Pulling Throwing a ball or other appropriate use of toys
Kicking	Shaking Grabbing or trying to grab an object
Spitting	Pushing Resisting or trying to "escape" physical restraining
Biting	Stamping Dropping something without throwing
Pinching	Throwing Undirected kicking of the legs or swinging of the arms; Disobedience, rowdiness, hyperactivity, or anger
Scratching	Aggression aimed at the child's own body
Physically threatening the above	

A.2. Additional Coding Rules

- Multiple aggressive acts are only counted separately if:
 - (a) there are at least 2 sec in between the behaviors,
 - (b) the child shows two different aggressive acts at the same time or close to each other.
- Aggression may also be coded if part of the behavior is invisible but audible (e.g. observing only the beginning of a hitting movement and hearing but not observing the resulting bang). If visibility of the behavior leaves room for doubts about the nature of the act, this act is not coded as aggression.
- If a behavior meets the criteria for an aggressive act, it is counted whether or not the mother seems to approve of the behavior.
- The total count of aggressive acts represents the frequency measure of physical aggression. This frequency can be counted separately for object- and mother-directed aggression.
- Every aggressive act that is coded is labeled to reflect the intensity of the behavior: 1 = low/moderate intensity, 2 = high intensity. These labels are not used for the total frequency count, but only for the global ratings (see below).

A.3. Global Ratings

The global ratings of aggression are based on a combination of the frequencies and intensities of aggressive acts. This can be done separately for object- and mother-directed aggression.

1. *Not aggressive*: The child shows no aggression at all. The child may be passive or uncooperative, but it does not show any aggressive behaviors.
2. *Slightly aggressive*: The child shows one or two aggressive behaviors, but the behavior has a low or moderate intensity, and is inconspicuous or mild.
3. *Somewhat aggressive*: The child shows some aggressive behavior, but the behavior is fleeting. There may be one high-intensity aggressive act or more than two aggressive acts of low to moderate intensity.
4. *Moderately aggressive*: The child shows two or more aggressive acts of high intensity, or multiple aggressive acts of mostly moderate intensity. However, the behaviors are generally short-lived

and do not dominate the interaction between mother and child.

5. *Aggressive*: The child shows several aggressive behaviors and gives the overall impression of being aggressive rather than nonaggressive, either because of the frequency or because of the intensity of the behaviors.
6. *Strongly aggressive*: The child shows many aggressive acts with a generally high intensity. Much of the total behavioral display is aggressive in nature.

REFERENCES

- Achenbach TM, Rescorla LA. 2000. Manual for ASEBA Preschool Forms & Profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth & Families.
- Alink LRA, Mesman J, Van Zeijl J, Stolk MN, Juffer F, Koot HM, Bakermans-Kranenburg MJ, Van IJzendoorn MH. 2006. The early childhood aggression curve: Development of physical aggression in 10- to 50-month-old children. *Child Dev* 77:954–966.
- Alink LRA, Mesman J, Stolk MN, Van Zeijl J, Juffer F, Koot HM, Bakermans-Kranenburg MJ, Van IJzendoorn MH. (In press). Maternal sensitivity moderates the relation between negative discipline and aggression in early childhood. *Soc Dev*.
- Arseneault L, Moffitt TE, Caspi A, Taylor A, Rijdsdijk FV, Jaffee SR, Ablow JC, Measelle JR. 2003. Strong genetic effects on cross-situational antisocial behaviour among 5-year-old children according to mothers, teachers, examiner-observers, and twins' self-reports. *J Child Psychol Psychiatry* 44:832–848.
- Baillargeon RH, Tremblay RE, Willms JD. 2005. Gender differences in the prevalence of physically aggressive behaviors in the Canadian population of 2- and 3-year-old children. In Pepler DJ, Madsen KC, Webster C, Levene KS (eds). *The Development and Treatment of Girlhood Aggression*. Mahwah, NJ: Lawrence Erlbaum Associates, pp 55–74.
- Broidy LM, Nagin DS, Tremblay RE, Bates JE, Brame B, Dodge KA, Fergusson D, Horwood JL, Loeber R, Laird R, Lynam DR, Moffitt TE, Pettit GS, Vitaro F. 2003. Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: A six-site, cross-national study. *Dev Psychol* 39:222–245.
- Brook JS, Zheng L, Whiteman M, Brook DW. 2001. Aggression in toddlers: Association with parenting and marital relations. *J Genet Psychol* 162:228–241.
- Crick NR, Casas JF, Mosher M. 1997. Relational and overt aggression in preschool. *Dev Psychol* 33:579–588.
- Cummings EM, Iannotti RJ, Zahn-Waxler C. 1989. Aggression between peers in early childhood: Individual continuity and developmental change. *Child Dev* 60:887–895.
- Del Vecchio T, O'Leary SG. 2006. Antecedents of toddler aggression: Dysfunctional parenting in mother–toddler dyads. *J Clin Child Adolesc Psychol* 35:194–202.
- Dionne G, Tremblay RE, Boivin M, Laplante D, Pérusse D. 2003. Physical aggression and expressive vocabulary in 19-month-old twins. *Dev Psychol* 39:261–273.
- Eddy JM, Leve LD, Fagot BI. 2001. Coercive family processes: A replication and extension of Patterson's coercion model. *Aggr Behav* 27:14–25.
- Estrem TL. 2005. Relational and physical aggression among preschoolers: The effect of language skills and gender. *Early Educ Dev* 16: 207–231.
- Garcia MM, Shaw DS, Winslow EB, Yaggi KE. 2000. Destructive sibling conflict and the development of conduct problems in young boys. *Dev Psychol* 36:44–53.
- Greenberg MT. 1999. Attachment and psychopathology in childhood. In Cassidy J, Shaver PR (eds). *Handbook of Attachment. Theory, Research, and Clinical Applications*. New York: Guilford Press, pp 469–496.
- Hampel ER, Ronchetti EM, Rousseeuw PJ, Stahel WH. 1986. *Robust Statistics: The Approach Based on Influence Functions*. New York: Wiley.
- Hartup WW. 2005. The development of aggression. Where do we stand? In Tremblay RE, Hartup WW, Archer J (eds). *Developmental Origins of Aggression*. New York: The Guilford Press, pp 3–22.
- Howes C, Phillipsen L. 1998. Continuity in children's relations with peers. *Soc Dev* 7:340–349.
- Keenan K, Shaw DS. 1994. The development of aggression in toddlers: A study of low-income families. *J Abnorm Child Psychol* 22:53–77.
- Keenan K, Shaw D, Delliquadri E, Giovannelli J, Walsh B. 1998. Evidence for the continuity of early problem behaviors: Application of a developmental model. *J Abnorm Child Psychol* 26:221–454.
- Koot HM, Van den Oord EJCG, Verhulst FC, Boomsma DI. 1997. Behavioral and emotional problems in young preschoolers: Cross-cultural testing of the validity of the Child Behavior Checklist/2-3. *J Abnorm Child Psychol* 25:183–196.
- McCartney K, Owen MT, Booth CL, Clarke-Stewart A, Vandell DL. 2004. Testing a maternal attachment model of behavior problems in early childhood. *J Child Psychol Psychiatry* 45:765–778.
- McFadyen-Ketchum SA, Bates J, Dodge K, Pettit GS. 1996. Patterns of change in early childhood aggressive-disruptive behavior: Gender differences in predictions from early coercive and affectionate mother–child interaction. *Child Dev* 67:2417–2433.
- NICHD Early Childcare Research Network. 2004. Trajectories of physical aggression from toddlerhood to middle childhood. *Monogr Soc Res Child Dev* 69:1128.
- Olson SL, Bates JE, Sandy JM, Lanthier R. 2000. Early developmental precursors of externalizing behavior in middle childhood and adolescence. *J Abnorm Child Psychol* 28:119–133.
- Ostrov JM, Woods KE, Jansen EA, Casas JF, Crick NR. 2004. An observational study of delivered and received aggression, gender, and social-psychological adjustment in preschool: "This White Crayon Doesn't Work...". *Early Child Res Q* 19:355–371.
- Ramsey E, Walker HM, Patterson GR. 1990. Generalization of the antisocial trait from home to school settings. *J Appl Dev Psychol* 11:209–223.
- Russell A, Hart CH, Robinson CC, Olsen SF. 2003. Children's sociable and aggressive behaviour with peers: A comparison of the US and Australia, and contributions of temperament and parenting styles. *Int J Behav Dev* 27:74–86.
- Shaw DS, Winslow EB. 1997. Precursors and correlates of antisocial behavior from infancy to preschool. In Stoff D, Breiling J, Maser JD (eds). *Handbook of Antisocial Behavior*. New York: Wiley, pp 148–158.
- Shaw DS, Keenan K, Vondra JI. 1994. Developmental precursors of externalizing behavior: Ages 1 to 3. *Dev Psychol* 30:355–364.
- Shaw DS, Gilliom M, Giovannelli J. 2000. Aggressive behavior disorders. In Zeanah H. (ed). *Handbook of Infant Mental Health*. New York: Guilford Press, pp 397–411.

- Snyder J, Stoolmiller M. 2002. Reinforcement and coercion mechanisms in the development of antisocial behavior: The family. In Reid JB, Patterson GR, Snyder J (eds). *Antisocial Behavior in Children and Adolescents. A Developmental Analysis and Model for Intervention*. Washington, DC: American Psychological Association, pp 101–122.
- Snyder J, Edwards P, McGraw K, Kilgore K, Holton A. 1994. Escalation and reinforcement in mother-child conflict: Social processes associated with the development of physical aggression. *Dev Psychopathol* 6:305–321.
- Sroufe LA. 1995. *Emotional Development. The Organization of Emotional Life in the Early Years*. Cambridge: Cambridge University Press.
- Strayer J, Roberts W. 2004. Empathy and observed anger and aggression in five-year-olds. *Soc Dev* 13:1–13.
- Tremblay RE. 2003. Why socialization fails. The case of chronic physical aggression. In Lahey BB, Moffitt TE, Caspi A (eds). *Causes of Conduct Disorder and Juvenile Delinquency*. New York: Guilford Press, pp 182–224.
- Tremblay RE, Japel C, Pérusse D, Boivin M, Zoccolillo M, Montplaisir J, McDuff P. 1999. The search for the age of “onset” of physical aggression: Rousseau and Bandura revisited. *Crim Behav Ment Health* 9:8–23.
- Tremblay RE, Nagin DS, Séguin JR, Zoccolillo M, Zelazo PD, Boivin M, Pérusse D, Japel C. 2004. Physical aggression during early childhood: Trajectories and predictors. *Pediatrics* 114:43–50.
- Van Beijsterveldt CEM, Bartels M, Hudziak JJ, Boomsma DI. 2003. Causes of stability of aggression from early childhood to adolescence: A longitudinal genetic analysis in Dutch twins. *Behav Genet* 33:591–605.
- Van Zeijl J, Mesman J, Stolk MN, Alink LRA, Van IJzendoorn MH, Bakermans-Kranenburg MJ, Juffer F, Koot HM. 2006a. Terrible ones? Assessment of externalizing behaviors in infancy with the Child Behavior Checklist. *J Child Psychol Psychiatry* 47:801–810.
- Van Zeijl J, Mesman J, Van IJzendoorn MH, Bakermans-Kranenburg MJ, Juffer F, Stolk MN, Koot HM, Alink LRA. 2006b. Attachment-based intervention for enhancing sensitive discipline in mothers of 1- to 3-year-old children at risk for externalizing behavior problems: A randomized controlled trial. *J Consult Clin Psychol* 74:994–1005.
- Zahn-Waxler C, Radke-Yarrow M, Wagner E, Chapman M. 1992. Development of concern for others. *Dev Psychol* 28:126–136.