Early childhood aggression

Alink, L.R.A.

3

The observation of early childhood physical aggression: A psychometric study

Mesman, J., Alink, L.R.A., Van Zeijl, J., Stolk, M.N., Juffer, F., Koot, H.M., Bakermans-Kranenburg, M.J., & Van IJzendoorn, M.H.
Manuscript in preparation

Abstract

The reliability and convergent as well as discriminant validity of an observational measure of physical aggression in toddlers and preschoolers, originally developed by Shaw, Keenan, and Vondra (1994), was investigated. The observation instrument was based on a developmental definition of aggression. Physical aggression was observed twice in a laboratory setting: the first time when children were 1 to 3 years old, and again 1 year later. Observed physical aggression was significantly related to concurrent mother-rated physical aggression for 2- to 4-year-olds, but not to maternal ratings of nonaggressive externalizing problems. We did not find significant 1-year stability of observed physical aggression in any of the age groups, whereas mother-rated physical aggression was significantly stable for all ages. The observational measure shows promise, but may have assessed *state* rather than *trait* aggression in our study.

Introduction

Physical aggression in children has been found to emerge around the 1st birthday, showing a sharp increase in frequency during the second year of life with a peak at ages 2 and 3 years, and a decrease from the third birthday onward (Alink et al., 2006b [chapter 2]; Cummings, Iannotti, & Zahn-Waxler, 1989; Tremblay et al., 1999, 2004). Although most children apparently learn to inhibit this socially unaccepted behavior, there is evidence that individual differences in the level of physical aggression in early childhood show stability across time and that high levels predict later maladaptation (Alink et al., 2006b [chapter 2]; Broidy et al., 2003; NICHD, 2004b). Both genetic and environmental factors have been implicated in the development and stability of aggression (Arsenault et al., 2003; Dionne, Tremblay, Boivin, Laplante, & Pérusse, 2003). Two parenting factors that have been found to contribute to the environmental effect on aggression in children are parental sensitivity (NICHD, 2004b; Olson, Bates, Sandy, & Lanthier, 2000) and harsh or inconsistent discipline (Shaw, Gilliom, & Giovannelli, 2000; Snyder, Edwards, McGraw, Kilgore, & Holton, 1994).

Despite the increasing research interest in early childhood physical aggression, salient developmental issues regarding the definition and assessment of the pertinent behaviors have been largely ignored. The current study reports on an observational measure for the assessment of physical aggression in early childhood, originally developed by Shaw, Keenan, and Vondra (1994). We address relevant developmental issues and investigate the reliability and stability of observed early childhood physical aggression, as well as its association with parent reports of aggressive versus nonaggressive externalizing problem behaviors (i.e., convergent versus discriminant validity).

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 have 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, Zheng, Whiteman, & Brook, 2001; Estrem, 2005; Ostrov, Woods, Jansen, Casas, & Crick, 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, Radke-Yarrow, Wagner, & Chapman, 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., 2006b [chapter 2]; 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 & Shaw, 1994), even in 1-year-old children (Alink et al., 2006b [chapter 2]). So 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.

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 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 (Koot & Verhulst, 1991; Van Zeijl et al., 2006), 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, objects, or animals, and is not due to 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, Casas, & Mosher, 1997; Estrem, 2005; Russell, Hart, Robinson, & Olsen,

2003; Tremblay et al., 1999, 2004). Relatively few studies have used observational methods to assess (physical) aggression in young children, and most of these focused on aggression among peers (Cummings et al., 1989; Ostrov et al., 2004; Strayer & Roberts, 2004). While 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, Owen, Booth, Clarke-Stewart, & Vandell, 2004; Shaw & 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., Eddy, Leve, & Fagot, 2001; Snyder & Stoolmiller, 2002). These early experiences of parent-child interactions are thought to constitute a blueprint for social exchanges that influences the child's behavior in other social settings (Greenberg, 1999; Ramsey, Walker, & Patterson, 1990).

Consistent with this emphasis on early parent-child interactions in the understanding of conduct problems such as aggression, Shaw and colleagues (1994) developed an instrument to observe early childhood physical aggression in a laboratory situation involving mothers and their children. 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. Shaw et al. (1994) reported that aggression observed with this instrument at age 18 months significantly predicted observed aggression at age 24 months, but only in girls. In addition, aggression observed at 24 months predicted mother-reported externalizing problems at 36 months, but only for boys. The lack of consistent associations between observed aggression at different times, and between observed aggression and later maternal ratings of externalizing problems as reported by Shaw and colleagues may have been due to the small sample sizes (results were presented separately for 51 boys and 37 girls). Further, it is unclear whether the instrument developed by Shaw et al. shows different associations with aggressive versus nonaggressive externalizing problem behaviors 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 in early childhood boys show higher levels of physical aggression than girls (Alink et al., 2006b [chapter 2]; Baillargeon, Tremblay, & Willms, 2005; Koot, Van den Oord, Verhulst, & Boomsma, 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 due to 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.

The present study

In the current study, we examined the reliability and validity of an observational method to assess physical aggression (based on the work by Shaw et al., 1994) in a large sample of 1- to 3-year-olds. We investigated (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, (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., in press). The data for the current study were derived from the pretest (Time 1) and posttest (Time 2) laboratory sessions.

Sample

In the screening phase of the study, 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 homogenous 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 Child Behavior Checklist for 1½- to 5-year-olds (CBCL/1½-5; Achenbach & 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 sample. Fifty-six percent of the children were boys and over half of the children had siblings (59%). Mean age of the mothers was 33 years and the majority of the parents had a high educational level (one or both parents with Bachelor's or Master's degree in 64% of the sample). There were no significant differences between selected families who agreed to participate in the entire intervention phase and those who did not regarding initial level of child externalizing problems (p = .99), child and maternal age (p = .18 and p = .07), child sex (p = .84), and presence of siblings (p = .98). The only statistically significant difference was that participating parents had a somewhat higher educational level than non-participating parents, $F(1, 434) = 12.70$, $p < .01$.

Procedure

Participating families were invited for a pretest in the laboratory. The mean ages of the children at the pretest were 15.60 months for the first age group (SD = 1.23, range = 13.58 – 18.84), 27.63 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). During the 1½-hour laboratory session, mother and child completed several tasks (coded afterwards from videotapes with observational measures, by independent coders unaware of experimental condition and other data regarding the participants) and mothers were asked to fill in some questionnaires. After the pretest, families were randomly assigned to either the control (n = 117) or the intervention (n = 120) group. There were no differences between both groups regarding initial level of child externalizing problems (p = .13), parental educational level (p = .46), child and maternal age (p = .85 and p = .97), and presence of siblings (p = .67). The only statistically significant difference was the percentage of girls, which was higher in the intervention group (51%) as compared to 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, using the same procedures as the pretest. Mean age of the children at the posttest was 39.41 months (SD = 10.11, range = 25.31 – 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 people, objects, or animals (the last category being irrelevant to the present observations). The observation instrument was based largely on the work by Shaw et al. (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 context of the behavior as well as the child's facial and verbal expressions were taken into account to establish whether the behavior should not be ascribed to motor limitations or play. The intent to hurt or harm someone or something was not a requisite for coding aggression. Behaviors not coded as physical aggression included screaming or cursing, temper tantrums 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 seconds 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 three episodes, including one neutral episode and two potentially frustrating episodes. The neutral episode was a break in which mother and child had a drink and a snack (duration 5 minutes after which coding ended, even if the break was longer). The first frustration episode consisted of a "clean-up" task (duration 1 to 4 minutes: the episode was ended after 4 minutes, or when the child finished the task). The second frustration task was a "don't" task in which the child was not allowed to touch attractive toys for 2 minutes , after which he or she was only allowed to touch the least attractive toy for another 2 minutes . For 1-year-olds the duration was two times 1.5 minutes , instead of 2 minutes (total duration of don't task: 3 or 4 minutes).

For each episode, the frequency of object-directed and mother-directed aggression was computed. These were summed across episodes to form total aggression frequencies for object- and mother-directed aggression separately. Because the duration of the clean-up task and the don't 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). In addition, ratings on a scale from 1 (not aggressive) to 5 (very aggressive) were assigned for both object- and mother-directed aggression. These ratings were

based on the frequency and intensity of the aggressive behaviors. Intercoder reliabilities (intraclass correlations, single rater, absolute agreement) were computed separately for 1-, 2-, and 3-year-olds (3 times 15 children) for object-directed frequency and rating, mother-directed frequency and rating, and total frequency and the average rating of aggression. Intraclass correlations were computed for a total of 19 separate pairs of coders, and ranged from .70 to .98, with 92% of correlations higher than .80.

Mother-rated physical aggression

The Physical Aggression Scale for Early Childhood (PASEC; Alink et al., 2006b [chapter 2]) was completed by mothers at Time 1 and Time 2 at the end of the laboratory session. 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 separately for each age group for both Time 1 and Time 2. Cronbach's alphas ranged from .71 to .87.

Mother-rated nonaggressive externalizing problems

The Child Behavior Checklist for 1½- to 5-year-old children (CBCL/1½-5; Achenbach & Rescorla, 2000) was used to assess nonaggressive externalizing problems and was obtained at Time 1 and Time 2. The previous version of the CBCL/1½-5 (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. Recently, 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. (2006). For the present paper, we used a Nonaggressive Externalizing Problems scale obtained by summing only the items from the Oppositional and Overactive narrowband scales. These scales did not include any items referring to physical aggression. Internal consistencies of the Nonaggressive Externalizing Problems scale were computed separately for each age group for both Time 1 and Time 2. Cronbach's alphas ranged from .80 to .88.

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 experimental condition. Similar results were obtained when missing data were excluded from the analyses.

Results

Preliminary analyses

The descriptive statistics for the six observation variables are summarized in Table 3.1. The statistics show that mother-directed physical aggression occurs infrequently compared to object-directed aggression, with 80% of children not showing any physical aggression directed at 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, $t(236) = 8.88$, $p < .01$, and the global ratings, $t(236) = 9.87$, $p < .01$. The correlations between observed object-directed and mother-directed aggression were .39 at Time 1 and .20 at Time 2 (both *p*s < .01). Because the very low occurrence of mother-directed aggression is likely to hamper reliable conclusions about this subtype of aggression, we decided to focus on the total frequency and the average rating of 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 .90 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 to the real numbers of aggressive acts. Therefore, analyses in the present paper will be based only on the total frequencies of physical aggression.

Outliers were found for observed physical aggression at Time 1 and Time 2 ($n = 4$ at each assessment, but not the same children), for Time 2 mother-rated physical aggression ($n = 2$) and nonaggressive externalizing problems ($n = 1$). These outliers ($|z| > 3.29$) were winsorized (i.e., "moved in close to the good data"; Hampel, Ronchetti, & Rousseeuw, 1986, p. 69) by replacing the outlying scores with the next highest value of the remaining distribution.

Table 3.1

Descriptive statistics for all physical aggression observation variables at Time 1

N = 237	Range	M (SD)	% with score '0'
Frequencies			
Object-directed	0 – 15	1.79 (2.53)	42
Mother-directed	0 – 11	0.43 (1.19)	80
Total	0 – 24	2.22 (3.19)	37
Global ratings			
Object-directed	0 – 4	0.91 (1.08)	42
Mother-directed	0 – 4	0.25 (0.62)	81
Average	0 – 4	0.58 (0.72)	37

* p < .05. ** p < .01.

Age and gender differences

Table 3.2 shows the means and standard deviations for observed and mother-rated physical aggression by age group and gender at Time 1. Separate ANOVAs were performed for each measure, examining the main effects of age group and gender, as well as the interaction between age group and gender. In addition, three separate ANOVAs were performed to investigate gender differences for each age group. We found significant main effects of gender for both measures of physical aggression, with boys showing higher levels of physical aggression than girls. The main effect of age group was only significant for mother-rated physical aggression, with post hoc tests showing that 1-year-olds had significantly lower scores compared to 2- and 3-year-olds ($p < .01$ in both comparisons). The separate ANOVAs for gender effects per age group showed that the difference between boys and girls (i.e., boys higher than girls) was only significant for both the observational measure and mother reports in 3-year-olds. In 1-year-old children, boys also showed higher levels of aggression than girls, but only for mother ratings. However, there were no significant interaction effects between gender and age group for either of the measures.

Table 3.2

Means and standard deviations for Time 1 observed and mother-rated physical aggression: Effects of age and gender

			Observation	Mother rating
Age 1 year	*n*			
Total	87	*M (SD)*	1.77 (2.56)	3.49 (2.52)
Boys	48	*M (SD)*	2.07 (2.69)	4.02 (2.69)
Girls	39	*M (SD)*	1.40 (2.36)	2.84 (2.15)
		Gender difference	$F(1,85) = 1.48$	$F(1,85) = 4.93$*
Age 2 years	*n*			
Total	75	*M (SD)*	2.63 (3.71)	6.23 (3.16)
Boys	49	*M (SD)*	2.80 (4.02)	6.43 (3.09)
Girls	26	*M (SD)*	2.31 (3.11)	5.85 (3.31)
		Gender difference	$F(1,73) = 0.29$	$F(1,73) = 0.45$
Age 3 years	*n*			
Total	75	*M (SD)*	2.20 (2.51)	5.43 (3.79)
Boys	35	*M (SD)*	2.89 (3.17)	6.43 (3.09)
Girls	40	*M (SD)*	1.60 (1.55)	5.85 (3.31)
		Gender difference	$F(1,73) = 5.23$*	$F(1,73) = 9.16$**
Main effect Age			$F(2,231) = 1.53$	$F(2,231) = 16.75$**
Main effect Gender across ages			$F(1,231) = 4.35$*	$F(1,231) = 12.16$**
Age * Gender			$F(2,231) = 0.38$	$F(2,231) = 1.89$

* $p < .05$. ** $p < .01$.

Associations between measures

To investigate the association between observed and mother-rated physical aggression, we computed cross-sectional correlations for the total sample as well as per age group, for both Time 1 and Time 2. We also examined the association between observed physical aggression and mother-rated nonaggressive ex-

ternalizing problems, to find out whether our observational measure showed discriminant validity. Table 3.3 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.

Table 3.3

Cross-sectional correlations between observed physical aggression and mother ratings for Time 1 and Time 2, per age group

	Mother ratings	
Observed Physical Aggression	Physical Aggression	Nonaggressive Externalizing
All Age Groups (N = 237)		
Time 1 (age 1 to 3 years)	.29**	.12
Time 2 (age 2 to 4 years)	.37**	.07
Age Group 1 (n = 87)		
Time 1 (age 1 year)	.15	.14
Time 2 (age 2 years)	.40*	.11
Age Group 2 (n = 75)		
Time 1 (age 2 years)	.35**	.07
Time 2 (age 3 years)	.34*	.10
Age Group 3 (n = 75)		
Time 1 (age 3 years)	.28*	.09
Time 2 (age 4 years)	.37**	.14

* $p < .05$. ** $p < .01$.

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 per age group. No significant longitudinal correlations for observed physical aggression were found for 1-year-olds, $r(87) = .00$, $p = 1.00$, for 2-year-olds, $r(75) = .02$, $p = .84$, or 3-year-olds, $r(75) = .15$, $p = .20$. The 1-year stability of mother-rated physical aggression was significant for all ages (all ps < .01), with correlations of .59 for 1-year-olds, .64 for 2-year-olds, and .56 for 3-year-olds. We also examined the stability of aggression separately for boys and girls in each age group and across age groups. Again, we only found significant longitudinal correlations for mother-reported aggression (all with $p < .01$, except for 1-year-old girls with $p < .05$), and not for the observational measure (all ps > .26). Because half of the families received an intervention between Time 1 and Time 2, we repeated our stability analyses correcting for experimental condition. These analyses did not yield different results.

Discussion

The results of the present study showed that physical aggression in 1- to 3-year-olds can be reliably assessed in a laboratory setting, using the observational measure originally developed by Shaw et al. (1994). For 2- and 3-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 includes clear rules about which behaviors should and should not be considered as such, taking into account developmentally relevant issues. Based on 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-settings or in multiple-risk samples), both subtypes of aggression should ideally be analyzed separately. Further, the ratings of physical aggression were discarded in favor of the frequency measure because of the high correlation between these two measures at both assessments (.94 and .90), which made the inclusion of both measures superfluous. A re-examination of our rating scale revealed that it relied too heavily on the frequency of aggression, with each scale point defined in terms of a certain frequency range. The inclusion of an estimation of the severity of the behaviors did not yield enough additional information to provide a clear distinction between the rating and the frequency measure.

The concurrent convergent validity of our observational measure of physical aggression 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 (one 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 due to 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. Parents may feel disinclined to assign this type of behavior to such young children, because they feel the terminology is inappropriate for this age group. Indeed, some mothers of 1-year-olds 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), while others may have taken the items at face-value, without trying to infer intent. These discrepancies between mothers of 1-year-olds may have lead 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 observational instrument in young children.

In addition to evidence for the convergent validity of the observational instrument, the results showed its discriminant validity. Observed physical aggression was not related to mother reports of nonaggressive externalizing problems such as oppositional and overactive behaviors in any of the age groups. 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. Our observational measure has proven to specifically measure physical aggression rather than oppositional or overactive behaviors.

For mother-reported physical aggression, 1-year-olds showed significantly less physical aggression than 2- and 3-year-olds (see also Alink et al., 2006b [chapter 2]), while 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 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 child's sight and reach. For 2- and 3-year-old children, the frustration tasks are more likely to resemble the 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. In the laboratory 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 observational measure 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 the group of 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 were 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., 2006b [chapter 2]; Baillargeon et al., 2005; Tremblay et al., 1999) may reflect true differences in aggression and not be completely due to informant biases.

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. To date, only a few studies have reported on the stability of (physical) aggression in toddlers and preschoolers. In Shaw et al.'s (1994) study, aggression as observed with the instrument also used in the current study was significantly stable from 18 to 24 months of age, but only for girls. Cummings et al. (1989) reported significant stability of observed aggression in a peer setting between the ages of 2 and 5 years. Significant 1-year stability has also been found for parent ratings of physical aggression in toddlers and preschoolers for both girls and boys (Alink et al., 2006b [chapter 2]; Van Beijsterveldt, Bartels, Hudziak, & Boomsma, 2003).

One explanation for the lack of stability found for the observation of physical aggression in the present study may be the relatively short duration of the laboratory episodes used for this measure. Whereas mothers reported on behaviors over a period of 2 months, the observations were based on approximately 13 minutes, compared to 22 minutes in the study by Shaw et al. (1994). In the Cummings et al. (1989) study, observation time was 66 minutes, and concerned children in a peer-setting, thus providing more "targets" for aggressive behaviors than the mother-child sessions employed in the current study and in the research by Shaw and colleagues. 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 13 minutes 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, whereas 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. The stability of mother-rated aggression is also likely to be partly due to mothers' traits in terms of consistent informant biases across time (Rowe & Kandel, 1997). However, in a previous study we found comparable 1-year stabilities using the mean of mother- and father-rated aggression in order to reduce possible informant effects (Alink et al., 2006b [chapter 2]).

To establish trait aggression in mother-child interactions by means of observation, studies 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 hour on each of 10 nonconsecutive days (Snyder et al., 1994), and in the home during two 2-hour sessions (McFadyen-Ketchum, Bates, Dodge, & Pettit, 1996), or a single 1-hour session (Eddy et al., 2001). These procedures may be more likely to elicit trait-like physical aggression in young children than relatively short laboratory sessions. Although we failed to find significant stability of observed physical aggression, longitudinal results of our study reported by Alink et al. (2006a [chapter 4]), showed that observed physical aggression was predicted by parenting behaviors in a theoretically meaningful way. These results suggest that our observational instrument does measure behaviors that are developmentally relevant in the context of parent-child interactions.

In conclusion, the observational measure of physical aggression in toddlers and preschoolers as designed by Shaw et al. (1994) shows promise: Intercoder reliability is high and it 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.