Youth & Society

http://yas.sagepub.com/

Young Girls' and Caretakers' Reports of Problem Behavior: Comprehension and Concordance Across Age, Race, and Behavior

Lee Ann Slocum, Sally S. Simpson, Alison E. Hipwell and Rolf Loeber

Youth Society published online 27 August 2010
DOI: 10.1177/0044118X10379128

The online version of this article can be found at:
http://yas.sagepub.com/content/early/2010/08/25/0044118X10379128

Published by:
SAGE
http://www.sagepublications.com

Youth & Society
XX(X) 1–31

DOI: 10.1177/0044118X10379128
http://yas.sagepub.com

Young Girls' and Caretakers' Reports of Problem Behavior: Comprehension and Concordance Across Age, Race, and Behavior

Lee Ann Slocum[1], Sally S. Simpson[2], Alison E. Hipwell[3], and Rolf Loeber[3]

Abstract

The article discusses a research instrument developed and utilized by the Pittsburgh Girls Study that asked young girls (ages 7 and 8) and their caretakers to report on the girls' involvement in a variety of problem behaviors. In this article, the authors evaluate whether comprehension, prevalence, and caretaker–child concordance of problem behaviors varied by child age and race. Results indicate that the girls understood most questions (except for some related to drug use) and that comprehension and reported involvement increased with age. Findings show that nonwhites showed greater comprehension and reported more involvement in problem behavior than Whites. Overall, the authors find modest concordance between reports from the girls and their caretakers, with greater agreement for nonwhites and older children. The authors conclude that a more comprehensive

[1]University of Missouri, St. Louis
[2]University of Maryland, College Park
[3]Western Psychiatric Institute and Clinic, Pittsburg, PA

Corresponding Author:
Lee Ann Slocum, Assistant Professor, Department of Criminology and Criminal Justice, University of Missouri, St. Louis, 324 Lucas Hall, One University Boulevard, St. Louis, MO 63121-4400
E-mail: slocuml@umsl.edu

understanding of youth problem behavior is gained when both caretakers and children provide reports.

Keywords

reliability, Antisocial Behavior Scale, girls

Developmental studies have found that the precursors to delinquency can be observed in early childhood (Loeber, 1982; Moffitt, Caspi, Rutter, & Silva, 2001; Thornberry & Krohn, 2001). For example, Tremblay and colleagues (2004) documented that the initiation of physical aggression begins in infancy. Although Krohn and colleagues (Krohn, Thornberry, Rivera, & LeBlanc, 2001) suggested that only a small number of children demonstrate problem behaviors at a very early age, the relationship between early conduct problems and later delinquency has been replicated using data taken from various time periods, countries, and cultures (for a review, see Loeber, 1982). This has made it increasingly important for researchers to accurately measure these behaviors in young children.

For the most part, problem behavior in young children has been measured using observational data or informant reports, typically taken from teachers and primary caretakers (Thornberry & Krohn, 2000). Other sources of data that are generally used to document the illegal activities of older youth and adults have not been used for young children. Official records, for example, are not available for young children because their actions rarely come to the attention of criminal justice authorities (Loeber, 1987, but see Farrington, Jolliffe, Loeber, & Homish, 2007). Self-report data, which are considered an adequately valid and reliable source of information on illegal behavior for adolescents and adults, rarely have been collected from young children. The assumption has been that young children are unable to provide accurate reports because they will not understand the questions or the behaviors contained in the questions (Loeber, Stouthamer-Loeber, van Kammen, & Farrington, 1989). This assumption, however, has been challenged.[1] There is some indication that children as young as 7 can provide reliable reports on their own problem behavior. For example, Loeber and colleagues found a surprising level of agreement between boy and caretaker reports of antisocial behavior for a sample of young boys studied in the Pittsburgh Youth Study (PYS). Aside from this study, relatively little is known about the ability of young children, especially girls, to provide reliable reports on their own participation in problem behaviors.

The goal of this study is to build upon Loeber and colleagues' (1989) earlier work. Specifically, using data from the Pittsburgh Girls Study (PGS) we

first examined the extent to which young girls understand questions regarding their problem behavior. We then estimated the prevalence of these behaviors using both caretaker and child self-reports, comparing the two sources as a means to assess the interrater reliability. Throughout, our emphasis was on examining differences across age and race groups.

Review of Literature

The fact that young children often engage in acts that are associated with future delinquency challenges researchers to develop instruments that can accurately measure problem behavior in young children. One such instrument, the Antisocial Behavior Scale (ABS), was developed as part of the PYS (Loeber et al., 1989) and again utilized in the PGS (Hipwell et al., 2002). This scale measures reports of antisocial (or problem) behavior collected from children as young as age 7.

The original ABS was assessed by Loeber and colleagues (1989) using a sample of boys and their caretakers. Contrary to assumptions, they found that most boys understood the questions, although comprehension was lower for items involving substance use. Moreover, they found a surprising level of agreement between caretaker and child self-reports. Concordance was highest for violent behaviors (in contrast to theft) and increased with respondent age. More important, the researchers claimed that information from adults (in addition to self-reports) is key to providing a more comprehensive view of youthful misconduct because there is a tendency for youth to underreport acts, either intentionally or unintentionally.

Although this study provided preliminary evidence of the reliability of the ABS, there were several limitations. First, because the PYS only sampled boys, it is not known whether the instrument elicits reliable information for girls; however, studies using older samples have found gender differences in the reliability of self-report data (see Weis, 1986). Second, this study did not disaggregate the findings by race, leaving open the possibility that the scale reliability was not equally high for Whites and nonwhites. This is important because there is some evidence of differential reliability in self-report data by race (e.g., Grills & Ollendick, 2002; Lau et al., 2004; but see Huizinga & Elliott, 1986). Third, although Loeber and colleagues (1989) concluded that there is high a level of convergence between parent and child reports, no tests were conducted to determine whether the observed agreement differed significantly from the levels of convergence that would be expected by chance. Finally, boys and their parents were not asked the same set of questions, making across-informant comparisons difficult.

Age, Race, and Interrater Agreement

Huizinga and Elliott (1986, p. 323) cautioned that the "quality of measures cannot be taken for granted" and that reliabilities for a general population may not carry over into subpopulations. Indeed, there are a number of factors that can affect the quality of self-report measures, including the age (Grills & Ollendick, 2002) and race of the respondent (Hindelang, Hirschi, & Weis, 1981). Empirical work in this area, however, has been inconclusive (Grills & Ollendick, 2002).

A meta-analysis on the agreement between parent and child reports of child problem behavior measured using the Achenbach Child Behavior Check List (Achenbach, McConaughy, & Howell, 1987) found higher rates of agreement for younger children (ages 6-11) than older children (ages 12-19); however, Loeber and colleagues' (1989) initial evaluation of the ABS found that boys studying in fourth grade had higher rates of agreement with their parents than did boys in first grade. Moreover, other studies suggest that the age–agreement relationship may vary by sample type, with more agreement for younger children in clinical samples and for older children in community samples (van der Meer, Dixon, & Rose, 2008).

Race is another factor that might affect interrater agreement. For example, Lau and colleagues (2004) found that African American youth said they engaged in more externalizing behavior than their parents reported and White youth reported less than their parents. In contrast, using data from the PYS, Youngstrom, Loeber, and Stouthamer-Loeber (2000) failed to find evidence of differential reliability by race for externalizing behaviors when comparing youth self-reports to caretaker reports. Others have observed differences in reliability by race and ethnicity that vary according to the type of behavior being examined (Grills & Ollendick, 2002).

To the extent that race differences in reliability exist, Weis suggests that they may be a consequence of communication errors (i.e., related to content or wording). Specifically, Whites may be more likely to comprehend the questions because they are better educated or because of cultural biases in the question wording (Chaiken & Chaiken, 1982; Weis, 1986). Differences in comprehension may also explain variation in parent–child concordance by age.

The Current Study

Results from this literature suggest several research hypotheses regarding levels of comprehension, prevalence, and reliability of problem behavior in our study. First, if the differences in reliability observed across demographic

groups are indeed due to differences in communication errors, then we expect to see lower levels of comprehension for younger children and nonwhites. Second, we hypothesize that older children and nonwhites will be more involved in problem behaviors and this will be reflected in both youth self-reports and caretaker reports. Third, consistent with Loeber and colleagues' (1989) work, we predict that children will underreport their involvement in problem behavior compared to their caretakers' reports, but we expect variation in interrater reliability by race and age. Finally, we anticipate that reports will be more reliable for older children and Whites and that reliability will vary by offense type. Specifically, we hypothesize that reports of overt behaviors (e.g., aggression) will be more reliable than those less likely to come to the attention of caretakers (e.g., theft or substance use).

This work is important for several reasons. First little is known about girls' ability to comprehend questions about their problem behavior, and even less is known about how this might vary with race. In addition, establishing the situations under which there is poor convergence between parent and child self-reports can help researchers determine when they should collect data from multiple sources (Lau et al., 2004; Loeber, Green, & Lahey, 1990). If comprehension or convergence varies by race or age, this has potential implications for tests of theory. Specifically, conclusions about the correlates of misconduct for young girls may vary depending on who provides information on the children's behavior. It also has implications for theory development. For example, because concordance between child and parent reports is based on "the behavior of the child and the lens through which adults view child behavior" (Weisz, McCarty, Eastman, Chaiyasit, & Suwanlert, 1997, p. 569), identifying racial and age variability in interrater reliability may provide insight into group differences in how parents view their children's delinquent behavior (Barker, Bornstein, Putnick, Hendricks, & Suwalsky, 2007; Lau et al., 2004). This may be especially relevant for labeling theory in which the perceptions of parents and others are hypothesized to affect youth involvement in delinquency (e.g., Matsueda, 1992).

Method

Sample

The data were collected as part of the PGS, a multiple cohort, longitudinal study designed to examine the development and causes of conduct disorder and delinquency in girls. All girls in the PGS sample were 5 to 8 years old at the time of the initial interview in 2000. The sample was drawn from 103,238

households in Pittsburgh and included all households located in the poorest 23 neighborhoods and 50% of households in other neighborhoods. As a result, there are a disproportionate number of disadvantaged girls (mainly nonwhite) included in the study. Of the households sampled, 3,118 met the qualifications for participation in the study.[2] Of these, 85%, or 2,451, were included in the final PGS sample. Because the younger cohorts were not administered the ABS questionnaire, the sample used in the current study consists of the 1,188 girls who were between the ages of 7 and 8 years at the initial interview. The girls were interviewed and their caretakers also completed interviews and questionnaires. For the most part, caretakers were the girls' biological mothers (92%) followed by grandparents (3%) and adopted parents (3%).

As expected because of the sampling strategy, most of the girls were nonwhite (58%) and many of them lived in families that received financial assistance (37%). The majority of the girls' caretakers were employed (70%), and most girls lived in a two-parent household (58%).

ABS

Child problem behavior was measured using the ABS, a 31-item self-report questionnaire. Loeber and colleagues designed this scale for use with the first- and fourth-grade cohorts in the PYS because other popular scales, such as the one used in the National Youth Survey (see Elliott, Huizinga, & Ageton, 1985), are considered inappropriate for young children. Specifically, traditional instruments often ask about behaviors that are rarely engaged in by young children and use words that children may not understand. The ABS was designed to rectify these concerns. This survey instrument only includes questions about behaviors that are age appropriate for young children. For example, there are no questions about being drunk in public or lying about one's age to purchase alcohol (Loeber et al., 1989).

In addition, the self-report ABS questionnaire uses more basic vocabulary than surveys designed for older youth. For example, the girls were asked about "taking something that does not belong to you" rather than "stealing." Similarly, the questions are more concrete in that they specify situations where the behavior may have taken place (e.g., "Did you take something from a store?" "Did you take something from home that did not belong to you").

Measures

Comprehension. To ensure that children understood the questions, the instrument was designed to assess each child's comprehension of the antisocial behavior measures. Before young respondents were asked to report whether

they had engaged in a particular behavior, the interviewer used the behavior in a sentence and asked the child if she understood the meaning. If the child said yes, she was then required to demonstrate her understanding by providing an example. Children who gave a correct response were then asked questions regarding their participation in that type of behavior. If the child did not understand the behavior, the interviewer provided an example of the behavior and then once again asked the child to provide an illustration. If the child still could not, the interviewer skipped any questions involving that behavior.[3] These comprehension questions asked about vandalism, theft, hitting, cheating, skipping school, avoiding payment, smoking marijuana, and sniffing glue. For each behavior, we coded comprehension as 1 if the child understood the behavior and 0 if not.

Child self-reported problem behavior. The ABS asked girls to report about their past-year involvement in 31 different behaviors, including vandalism, theft, substance use, hitting, running away, being sent home from school, and cheating on tests.[4] Complete listings of the items are presented in Tables 2 through 5. For each behavior, girls who had not engaged in the behavior in the past year were given a score of 0 and those who engaged in the behavior once or more in the past year were given a score of 1. Because of low base rates, for some investigations we aggregated the behaviors into five general categories of problem behavior: vandalism, theft, aggression, status offenses, and drug use.

Children who did not comprehend a question could be excluded from subsequent analyses of that behavior or, alternately, one could assume that children who did not understand the question did not engage in the behavior. To remain consistent with the original reliability check by Loeber and colleagues (1989), we assumed that children who did not understand the question did not engage in that particular behavior. We are cognizant that this coding scheme may understate the level of problem behavior and negatively affect parent–child concordance; therefore, we repeated the analyses excluding children who did not understand the behavior. We note when findings vary according to the coding scheme used.

Caretaker-reported problem behavior. Caretakers were asked to report on their children's behavior over the past year for the same set of behaviors that the girls were asked to report on. As with the girls' data, these responses were recoded so that if the caretaker reported his or her child engaged in the behavior during the past year it was scored as a 1 and a lack of involvement was scored as a 0.

Analyses

The analyses are divided into four main parts. First, to assess comprehension we computed the percent of children who understood each question. These

results were disaggregated by age and race, and we assessed whether there were significant differences across subgroups using a *t* tests for two proportions (Bachman & Paternoster, 2004).

Second, we examined the prevalence of problem behaviors. We first computed the percent of all girls and caretakers who reported that the girl had engaged in each behavior during the past year and then disaggregated these findings by age and race. To determine whether the differences in levels of participation varied significantly by age and race, *t* tests for two proportions were computed using first the child self-reports and then the caretaker reports. Dependent sample *t* tests for two proportions were also computed to determine whether there were significant differences in reporting levels between caretakers and children. This provides information on whether we would reach the same conclusions about age and race differences using caretaker reports as we would using child self-reports. All tests were conducted using the aggregate behavior measures because of low base rates for individual items. To take into account the large number of comparisons, a Bonferroni correction was used.

Third, we examined the level of interrater agreement between caretaker and child self-reports by age and race. Concordance was measured using three different methods, all of which used pairs of reports (dyads) as the unit of analysis. A detailed discussion of the computation of all measures is presented in the appendix. First, we computed the percent of children who reported engaging in a behavior given the caretaker reported the girl had engaged in the behavior. Then we computed the percent of caretakers who reported that their children engaged in a behavior given that the child reported she had engaged in the behavior. Although these two measures seem similar, the former assumes that if the behavior has come to the attention of the caregiver then it actually occurred and, therefore, should be reported by the youth. In contrast, the latter measure addresses the extent to which caregivers are aware of their children's behavior because behaviors reported by youth may not have come to the attention of their caregivers.

Finally, interrater reliability was assessed by calculating relative improvement over chance scores (RIOC; see Copas & Loeber, 1990; Farrington & Loeber, 1989; Loeber & Dishion, 1983). RIOC is a measure of interrater agreement that indicates the concordance between child and caretaker reports, relative to the number of matches expected based on chance and the maximum number of matches possible (Farrington & Loeber, 1989).[5] It ranges from -1 to 1, with negative scores indicating that there are fewer matches than expected based on chance, positive scores indicating that there are more matches than expected by chance, and 0 indicating that the number of matches is what is

expected by chance. For each RIOC score, we assessed whether the number of matches observed was significantly different than the number of matches expected by chance[6] and, when comparisons were possible, we evaluated whether there were significant differences in RIOC across cohorts and races using tests described by Copas and Loeber (1990; see appendix).

The final part of the analysis assessed reliability across caretaker and child reports regarding the girl's *relative level* of involvement in problem behavior. Even if caretaker and child reports differed on the girl's absolute level of involvement in antisocial behavior, it is still possible that the informants converged in terms of the child's relative level of involvement compared to other girls in the sample. For this analysis, we first rank-ordered the girls by the number of different problem behaviors they self-reported and then we repeated this ranking procedure using caretaker reports. We assessed the concordance in these rankings by determining what percent of children who were ranked as the most delinquent (top 5%) on the basis of caretaker reports were also ranked as most delinquent on the basis of self-reports. Spearman correlations were also computed to assess the relationship between parent and child rankings. These analyses were repeated disaggregating the rankings by age and race.

Results

Comprehension

Table 1 shows the percents of girls who understood the questions, sorted by behavior type and disaggregated by race and age. Overall, the girls' comprehension for all behaviors was high, with the exception of substance use. The behaviors most often understood were cheating on school tests (97%) and hitting (95%). In contrast, the questions least understood asked about smoking marijuana (47%) and sniffing glue (31%).

Although this general pattern held when the data were disaggregated by age, older girls demonstrated significantly greater comprehension of all acts, with the exception of sniffing glue. Breaking the sample down by race, nonwhites were generally *more* likely to understand the questions, but only one difference—marijuana use—was statistically significant.

Prevalence: Past Year

Data comparing the levels of involvement observed in the caretaker reports to those found in the self-reports are presented for the entire sample in Table 2 and then disaggregated by age and race in Tables 2 and 3. The results for both

Table 1. Comprehension by Age and Race

Behavior	All girls (%)	Age 7 (%)	Age 8 (%)	Age comparison *t* score[a]	White (%)	Nonwhite (%)	Race comparison *t* score[a]
Damage on purpose	90.7	87.1	94.2	−4.26**	89.6	91.6	−1.16
Cheat on school tests	96.5	94.7	98.2	−3.27**	95.0	97.5	−2.23
Hit, slapped, shoved, scratched, or pulled hair of someone	94.6	91.9	97.2	−4.03**	93.6	95.4	−1.30
Skip school	90.7	86.6	94.6	−4.75**	90.0	91.1	−0.68
Avoid paying	90.2	86.6	93.9	−4.45**	91.4	89.3	1.23
Smoke marijuana	47.1	38.1	55.8	−6.19**	39.7	52.5	−4.43**
Sniff glue	31.0	27.9	33.9	−2.22	28.9	32.5	−1.35
N	1,188	580	608	—	499	689	—

a. Two-tailed *t* test for two population proportions (variance assumed to be unequal except for skipping school race comparison)
*$p < .05$ (Bonferroni correction: $\alpha = .05/7 = .007$). **$p < .01$ (Bonferroni correction: $\alpha = .01/7 = .001$).

self- and caretaker reports confirm that the precursors to delinquency begin at a young age. The most commonly reported behavior type was aggression: 48% of children and 53% of caretakers reported that the child had engaged in at least one form of aggression (most typically hitting other children) in the past year. Interestingly, sibling aggression was more prevalent in child self-reports, whereas peer aggression was more prevalent in caretaker reports. This may reflect the fact that parents view sibling fights as normative and, therefore, are less likely to report them. Vandalism was also relatively common in both self- and caretaker reports (13% and 17%, respectively). Substance use was the least prevalent behavior; according to both sets of informants, less than 3% of the participants used any type of illicit substance in the past year, including alcohol and tobacco. Other behaviors were prevalent according to the caretaker reports, but rarer according to self-reports. For example, caretaker

Table 2. One-Year Prevalence Rates of Problem Behavior as Reported by Girls and Their Caretakers for the Full Sample and Disaggregated by Age

Behavior	Full sample		Age 7		Age 8		Cross-cohort comparisons	
	Child self-reports (%)	Caretaker reports (%)	Child self-reports (%)	Caretaker reports (%)	Child self-reports (%)	Caretaker reports (%)	*t* score[b] (self, age 7 vs. 8)	*t* score[b] (caretakers, age 7 vs. 8)
Vandalism	13.4	16.9	12.1	17.4	14.6	16.4	−1.30	.44
Broken family possessions	9.9	12.6	8.6	12.9	11.0	12.3		
Broken school possessions	1.4	1.0	1.2	0.9	1.6	1.2		
Broken other possessions	3.2	2.0	2.6	1.6	3.8	2.5		
Spray-painted graffiti	1.3	5.1	1.2	5.5	1.3	4.6		
Set fires	0.3	0.2	0.2	0	0.3	0.3		
t score[a] (self vs. caretaker)	−2.40		−2.55		−.85			
Theft	16.2	26.9	14.3	28.4	17.8	25.3	−1.61	1.23
Shoplifted	6.1	6.8	6.9	6.4	5.4	7.2		
Stolen money from home	5.5	15.3	4.3	16.9	6.6	13.8		
Stolen anything (beside money) from home	5.2	11.9	5.4	12.1	5.1	11.7		
Stolen from school	2.9	4.0	2.6	3.6	3.1	4.3		
Stolen from other's home	1.9	4.2	1.7	3.8	2.0	4.6		
Stolen from car	1.0	1.1	1.0	1.4	1.0	.8		
Avoided paying	1.2	0.8	0.3	0.7	2.0	1.0		
Stolen a purse	0.9	0.3	0.7	0.3	1.2	0.3		
t score[a] (self vs. caretaker)	−7.03**		−6.45**		−3.54**			
Aggression	48.2	53.0	43.6	52.6	51.5	53.3	−2.81*	−0.30
Hit adult at school	1.2	0.9	1.0	0.7	1.3	1.2		
Hit parent/caregiver	0.8	5.5	1.0	6.7	0.5	4.3		

(continued)

Table 2. (continued)

Behavior	Full sample		Age 7		Age 8		Cross-cohort comparisons	
	Child self-reports (%)	Caretaker reports (%)	Child self-reports (%)	Caretaker reports (%)	Child self-reports (%)	Caretaker reports (%)	*t* score[b] (self, age 7 vs. 8)	*t* score[b] (caretakers, age 7 vs. 8)
Hit sibling	33.8	22.5	31.0	23.3	35.9	21.7		
Hit other kids	21.8	35.9	18.1	35.0	25.2	36.7		
Acted up in public	10.1	6.5	7.9	7.1	12.2	5.9		
Hidden weapon	0.8	0.3	0.7	0.2	1.0	0.3		
Thrown rocks	4.6	1.6	4.1	1.7	5.1	1.5		
t score[a] (self vs. caretaker)	−2.81*		−3.33**		−0.58			
Status	16.6	16.6	13.4	14.1	19.6	18.9	−2.87*	−2.22
Cheated on school test	6.3	1.5	4.5	0.7	8.1	2.3		
Trespassed in other's yard	6.5	9.0	5.5	7.8	7.4	10.2		
Run away	1.9	1.0	2.1	1.0	1.6	1.0		
Skipped school	0.8	0.9	0.7	0.9	0.8	1.0		
Sent home from school	5.7	6.1	4.5	4.7	6.9	7.4		
t score[a] (self vs. caretaker)	0.06		−0.40		0.43			
Substance use	2.8	3.0	3.6	3.3	3.5	5.3	0.16	−1.69
Sipped beer	0.4	1.4	0.5	2.2	1.3	2.3		
Sipped wine	0.6	1.2	0.3	1.0	0.5	1.2		
Sipped liquor	0.2	0.4	0.3	0.2	0.5	0.5		
Smoked cigarette/chewed tobacco	0.4	1.0	0.3	0.7	1.3	2.5		
Smoked marijuana	0	0	0	0	0	0		
Sniffed glue	1.6	0.0	2.1	0	1.2	0.2		
t score[a] (self vs. caretaker)	−1.01		0.32		−1.72			
N	1,188		580		608			

a. Paired-samples *t* test (two tailed).
b. Independent samples *t* test for two proportions (two tailed).
*$p < .05$ (Bonferroni correction: $\alpha = .05/5 = .01$). **$p < .01$ (Bonferroni correction: $\alpha = .01/5 = .002$).

reports indicated significantly higher prevalence of thefts than self-reports. Caretakers also reported that their children had significantly higher levels of involvement in aggression.[7]

Age

We observed patterns similar to those for the full sample when we disaggregated by age: Regardless of informant, aggression was the most commonly reported behavior in both cohorts and substance use was the least prevalent. When differences across cohorts were examined using self-reports only, there was limited indication of variation in involvement by age (see the first column of the cross-cohort comparison). Although most acts were more prevalent among the 8-year-olds compared with the 7-year-olds, only 2 of the summary measures, aggression and status offenses, differed significantly by age.[8] Caretakers' reports provided even less evidence to suggest an age effect (see the last column of the cross-cohort comparison); only about half of behaviors were more likely to be reported by the caretakers of the older girls, and none of the cohort differences for the summary measures was significant.

When self-reported levels of involvement were compared to those from caregiver reports within each cohort, the findings from the entire sample generally held (e.g., problem behavior was generally more prevalent in caretaker compared to child reports). The only difference to emerge was that for the 8-year-olds, there was no significant difference in the level of aggression according to girl and caretaker reports.

Race

Moving next to race differences (see Table 3), aggression again was the most prevalent behavior whereas substance use was the least prevalent. However, these data showed stronger evidence of race effects than age effects. Using only the child self-reports, we observed that nonwhites were significantly more involved in vandalism and status offenses than Whites (see the first column of the cross-race comparison). Caretaker reports indicated more race differences in offending than did child self-reports. In addition to providing evidence of racial disparities in vandalism and status offenses, caretaker reports indicated a higher percent of nonwhite girls engaged in theft and aggression during the past year compared to White girls.

There were also important differences between child and caretaker prevalence reports. Caretakers of White girls reported significantly more involvement of their children in theft than did the children themselves, whereas nonwhite caretakers reported higher prevalence rates for vandalism, theft, and aggression than did child self-reports.[9]

Table 3. One-Year Prevalence Rates of Problem Behavior as Reported by Girls and Their Caretakers, Disaggregated by Race

	White		Nonwhite		Comparisons across race	
Behavior	Child self-reports (%)	Caretaker reports (%)	Child self-reports (%)	Caretaker reports (%)	*t* score (self, Whites vs. Nonwhite)[b]	*t* score (caretaker, Whites vs. Nonwhite)[b]
Vandalism	8.8	9.6	16.7	22.0	−4.15**	−5.79**
Broken family possessions	6.6	8.8	12.2	15.4		
Broken school possessions	0.2	0.2	2.3	1.6		
Broken other possessions	2.2	1.2	3.9	2.6		
Spray-painted graffiti	0.6	1.6	1.7	7.5		
Set fires	0	0	0.4	0.3		
Vandalism *t* score[a] (self vs. caretaker)	−0.45		−2.56*			
Theft	13.7	21.4	17.9	30.8	−1.97	−3.48**
Shoplifted	2.6	3.2	8.6	9.4		
Stolen money from home	5.4	10.9	5.4	18.5		
Stolen anything (beside money) from home	6.0	12.1	4.7	11.7		
Stolen from school	1.2	2.0	4.1	5.4		
Stolen from other's home	1.2	3.0	2.3	5.1		
Stolen from car	0.8	0.4	1.2	1.6		
Avoided paying	0.2	0.0	1.9	1.5		
Stolen a purse	0.4	0.4	1.3	0.3		
Theft *t* score[a] (self vs. caretaker)	−3.39**		−6.30**			
Aggression	46.0	45.2	49.6	58.5	−1.19	−4.65**
Hit adult at school	0.6	0.4	1.6	1.3		
Hit parent/caregiver	1.2	11.1	0.4	1.5		
Hit sibling	39.4	11.0	29.7	30.8		
Hit other kids	10.1	31.3	29.9	39.4		
Acted up in public	8.0	5.0	11.7	7.6		
Hidden weapon	0.0	0.2	1.5	0.3		
Thrown rocks	2.0	0.4	6.3	2.5		
Aggression *t* score[a] (self vs. caretaker)	0.30		−4.04**			

(continued)

Table 3. (continued)

Behavior	White		Nonwhite		Comparisons across race	
	Child self-reports (%)	Caretaker reports (%)	Child self-reports (%)	Caretaker reports (%)	*t* score (self, Whites vs. Nonwhite)[b]	*t* score (caretaker, Whites vs. Nonwhite)[b]
Status	11.6	11.6	20.2	20.1	−4.09**	−3.93**
Cheated on school test	5.2	2.0	7.1	1.2		
Trespassed in other's yard	5.0	8.7	7.6	9.3		
Run away	2.4	1.6	1.5	0.6		
Skipped school	0.4	0.2	1.0	1.5		
Sent home from school	0.6	0.6	9.4	9.9		
Status *t* score[a] (self vs. caretaker)	0		0.08			
Substance use	2.8	3.0	4.1	5.2	−1.16	−1.86
Sipped beer	0.4	1.4	1.3	2.9		
Sipped wine	0.6	1.2	0.3	1.0		
Sipped liquor	0.2	0.4	0.6	0.3		
Smoked cigarette/chewed tobacco	0.4	1.0	1.2	2.0		
Smoked marijuana	0	0	0	0		
Sniffed glue	1.6	0.0	1.6	0.1		
Substance use *t* score[a] (self vs. caretaker)	−0.20		−1.09			
N	499		687			

a. Paired-samples *t* test (two tailed).
b. Independent samples *t* test (two tailed)
*$p < .05$ (Bonferroni correction: $\alpha = .05/5 = .01$). **$p < .01$ (Bonferroni correction: $\alpha = .01/5 = .002$).

Interrater Reliability

In the previous section, we examined whether self-reports and caretaker reports lead to similar conclusions regarding girls' involvement in problem behavior. In contrast, the results presented in this section used the girl–caretaker dyad as the unit of analysis and compared each girl's report to her caretaker's report to assess interrater reliability. Several different methods were used to measure reliability, and findings are presented for the full sample and by age and race (Tables 4 and 5).

Entire Sample

Percent of children self-reporting the behavior given the caretaker reported the behavior. One way to measure agreement is to determine what percent of girls report engaging in a behavior given the caretaker reported the child engaged in the behavior. This method is more useful for behaviors that are likely to be observed by or reported to the caretaker, such as hitting a caretaker, stealing from home, breaking family possessions, and running away.

For the entire sample, the prevalence of self-reports conditional on the caretaker reporting the behavior was greatest for aggression (62%), especially hitting siblings, other children and adults at school (46%, 37%, and 36%, respectively) and carrying a hidden weapon (33%; see Table 4). The other behaviors with relatively high levels of this type of concordance fall into the category of status offenses. More than 33% of children reported that they had cheated on a school test, run away, or been sent home from school conditional on their caretakers reporting the behavior, and, overall, conditional prevalence was almost 40% for status offenses. This suggests that aggression and status offenses are salient behaviors that girls are willing to report.

Overall, children were least likely to report engaging in substance use given that their caretaker reported the girl had engaged in substance use, although these results are difficult to interpret because of the girls' limited involvement with these behaviors and relatively poor comprehension of these questions. There was, however, a relatively high level of agreement for tobacco use. Of the 19 caretakers who reported that their girls had used tobacco in the past year, more than 26% of their girls also reported using tobacco. Behaviors that were more common yet had relatively low levels of conditional prevalence (less than 10%) included spray painting graffiti, stealing anything (besides money) from home, stealing from school, and hitting a caregiver. The low level of concordance for the last behavior—hitting a caregiver—was surprising, given that this seems likely to be a salient event for the child. It is possible that this divergence is the result of perceptual

Table 4. Conditional Prevalence and RIOC for the Full Sample and Disaggregated by Age

	Full sample			Age 7			Age 8		
	% SR given CR[a]	% CR given SR[b]	RIOC[c] (N = 1,188)	% SR given CR[a]	% CR given SR[b]	RIOC[c] (n = 580)	% SR given CR[a]	% CR given SR[b]	RIOC[c] (n = 608)
Vandalism	20.1 (199)	25.2 (159)	0.077[d]	14.9 (101)	21.4 (70)	0.050	25.0 (100)	28.1 (89)	0.141[d]
Broken family possessions	17.3 (150)	22.2 (117)	0.110[d]	10.7 (75)	16.0 (50)	0.035	24.0 (75)	26.9 (67)	0.166[d]
Broken school possessions	8.3 (12)	5.9 (17)	—	0 (5)	0 (7)	—	14.3 (7)	10.0 (10)	—
Broken other possessions	0 (24)	0 (38)	—	0 (9)	0 (15)	—	0 (15)	0 (23)	—
Spray-painted graffiti	6.7 (60)	26.7 (15)	—	9.4 (32)	42.9 (7)	—	3.6 (28)	12.5 (8)	—
Set fires	0.0 (2)	0 (3)	—	(0)	0 (1)	—	0 (2)	0 (2)	—
Theft	27.4 (317)	45.5 (191)	0.256[d]	23.6 (165)	47.0 (83)	0.259[d]	31.2 (154)	44.4 (108)	0.257[d]
Shoplifted	23.5 (81)	26.4 (72)	0.210[d]	29.7 (37)	27.5 (40)	0.245[d]	18.2 (44)	25.0 (32)	0.191[d]
Stolen money from home	16.0 (181)	45.3 (64)	0.354[d]	11.2 (98)	45.8 (24)	0.348[d]	21.4 (84)	45.0 (40)	0.363[d]
Stolen anything (beside money) from home	6.4 (140)	14.5 (62)	0.030	7.1 (70)	16.1 (31)	0.046	5.7 (71)	12.9 (31)	—
Stolen from school	8.5 (47)	11.76 (34)	—	4.8 (21)	6. 7 (15)	—	11.5 (26)	15.8 (19)	—
Stolen from other's home	12.0 (50)	27.3 (22)	0.241[d]	13.6 (22)	30.0 (10)	—	10.7 (28)	25.0 (12)	—
Stolen from car	0 (13)	0 (12)	—	0 (8)	0 (6)	—	0 (5)	0 (6)	—
Avoided paying	20.0 (10)	14.3 (14)	—	25.0 (4)	50.0 (2)	—	16.7 (6)	8.3 (12)	—
Snatched purse/picked pocket	25.0 (4)	9.1 (11)	—	50.0 (2)	25.0 (4)	—	0 (2)	0 (7)	—
Aggression	62.2 (621)	68.5 (564)	0.329[d]	54.8 (305)	66.0 (253)	0.284[d]	67.6 (324)	70.4 (311)	0.366[d]
Hit adult at school	36.4 (11)	28.6 (14)	—	25.0 (4)	16.7 (6)	—	42.9 (7)	37.5 (8)	—
Hit parent/caregiver	6.2 (65)	44.4 (9)	—	7.7 (39)	50.0 (6)	—	3.8 (26)	33.3 (3)	—
Hit sibling	46.0 (265)	30.7 (398)	0.185[d]	40.0 (135)	30.0 (180)	0.135	51.5 (132)	31.2 (218)	0.241[d]
Hit other kids	37.1 (426)	61.7 (256)	0.402[d]	31.0 (203)	60.0 (105)	0.384[d]	42.6 (223)	62.9 (151)	0.413[d]
Acted up in public	22.1 (77)	14.2 (120)	0.133[d]	19.5 (41)	17.4 (46)	0.125	25.0 (36)	12.2 (74)	0.145
Hidden weapon	33.3 (3)	10 (10)	—	0 (1)	0 (4)	—	50.0 (2)	16. 7 (6)	—
Thrown rocks	10.5 (19)	3.8 (53)	—	20.0 (10)	8.7 (23)	—	0 (9)	0 (30)	—

(continued)

Table 4. (continued)

	Full sample			Age 7			Age 8		
	% SR given CR[a]	% CR given SR[b]	RIOC[c] (N = 1,188)	% SR given CR[a]	% CR given SR[b]	RIOC[c] (n = 580)	% SR given CR[a]	% CR given SR[b]	RIOC[c] (n = 608)
Status	39.3 (196)	39.1 (197)	0.272[d]	35.4 (82)	37.2 (78)	0.268[d]	41.7 (115)	40.3 (119)	0.280[d]
Cheated on school test	38.9 (18)	9.3 (75)	0.348[d]	50.0 (4)	7.7 (26)	—	35.7 (14)	8.2 (49)	0.301[d]
Trespassed in other's yard	15.9 (107)	22.1 (77)	0.143[d]	13.3 (45)	18.8 (32)	0.119	17.7 (62)	24.4 (45)	0.158[d]
Run away	33.3 (12)	18.2 (22)	—	50.0 (6)	25.0 (12)	—	16.7 (6)	10.0 (10)	—
Skipped school	27.3 (11)	33.3 (9)	—	20.0 (5)	25.0 (4)	—	33.3 (6)	40 (5)	—
Sent home from school	49.3 (71)	51.5 (68)	0.484[d]	40.7 (27)	42.3 (26)	0.395[d]	53.3 (45)	57.1 (42)	0.538[d]
Substance use	13.7 (51)	16.7 (42)	0.129[d]	5.3 (19)	4.8 (21)	—	18.3 (32)	28.6 (21)	0.246[d]
Sipped beer	3.7 (27)	9.1 (11)	—	0 (13)	0 (3)	—	7.1 (14)	12.5 (8)	—
Sipped wine	0.0 (13)	0 (5)	—	0 (6)	0 (2)	—	0 (7)	0 (3)	—
Sipped liquor	0.0 (4)	0 (5)	—	0 (1)	0 (2)	—	0 (3)	0 (3)	—
Smoked cigarette/ chewed tobacco	26.3 (19)	50 (10)	0.492[d]	25.0 (4)	50.0 (2)	—	26.7 (15)	50.0 (8)	—
Smoked marijuana	(0)	(0)	—	(0)	(0)	—	(0)	(0)	—
Sniffed glue	100.0 (1)	5.3 (19)	—	(0)	0 (12)	—	100 (1)	14.3 (7)	—

Note: RIOC = relative improvement over chance; SR = child self-report; CR = caretaker report. No significant differences (α = .05) were found in RIOC scores across age groups.

a. Number of caretakers reporting the behavior in parentheses.

b. Number of children reporting the behavior in parentheses.

c. Only calculated when cell sizes were greater than 5.

d. RIOC differs significantly from 0 at α = .05/17 = .0029 for all girls, α = .05/13 = .0038 for 7-year-olds, and α = .05/14 = .0036 for 8-year-olds.

differences between what adults and children consider hitting or unwillingness on the part of children to report this behavior.

Percent of caretakers reporting the behavior given the child self-reported the behavior. Measuring concordance by computing the percent of caretakers who reported a behavior given that the child reported that behavior provides an indication of the extent to which caretakers know what their children are doing and are willing to report the behavior. For the entire sample of girls, the conditional prevalence for caretaker reports was highest for aggression (69%) and theft (46%). As expected, the items with the highest conditional prevalence rates were primarily overt behaviors likely to come to the attention of caregivers—hitting other children (62%), being sent home from school (52%), stealing money from home (45%), and hitting a caretaker (44%). The summary measure with the lowest levels of concordance was substance use.

For some behaviors, such as stealing money from home, spray-painting graffiti, and hitting a caregiver, we observed relatively large differences between the levels of conditional prevalence given parent reports and the conditional prevalence given child reports and, in most of these cases, conditional prevalence given child reports was higher. This suggests that at this relatively young age, parent reports may be more complete than child reports.

RIOC. We also measured agreement by calculating RIOC scores, which are listed in Table 4 for the entire sample. RIOC was not calculated for behaviors in which any of the cells used to calculate concordance contained five or fewer observations.[10] For most behaviors, the level of agreement between child and caretaker reports was significantly higher than would be expected by chance alone. RIOC scores were highest for the summary measure of aggression, where there were almost 33% more matches than expected based on chance. The individual items with the highest RIOC scores were smoking cigarettes (.492) and being sent home from school (.484). RIOC was lowest for the summary measure of vandalism (.077) followed by substance use (.129). Moreover, there were no more matches than would be expected based on chance for stealing anything (besides money) from home. The mean RIOC for the 12 individual items where this score was computed was .261 indicating that, on average, we observed over 26% more matches than we expected to see on the basis of chance alone.

Age

Percent of children self-reporting the behavior given the caretaker reported the behavior. For both cohorts, the results from the summary measures mirrored those for the total sample. Aggression and status offenses were most likely to

Table 5. Conditional Prevalence and RIOC Disaggregated by Race

	White			Nonwhite		
	% SR given CR[a]	% CR given SR[b]	RIOC[c] (n = 499)	% SR given CR[a]	% CR given SR[b]	RIOC[c] (n = 689)
Vandalism	14.6 (48)	15.9 (44)	0.070	21.9 (151)	28.7 (115)	0.086
Broken family possessions	11.4 (44)	15.2 (33)	0.069	19.8 (106)	25.0 (84)	0.114
Broken school possessions	0 (1)	0 (1)	—	9.1 (11)	6.3 (16)	—
Broken other possessions	0 (6)	0 (11)	—	0 (18)	0 (27)	—
Spray-painted graffiti	0 (8)	0 (3)	—	7.7 (52)	33.3 (12)	—
Set fires	(0)	(0)	—	0 (2)	0 (3)	—
Theft	21.7 (106)	33.8 (68)	0.158	30.3 (211)	52.0 (123)	0.307[d]
Shoplifted	6.3 (16)	7.7 (13)	—	27.7 (65)	30.5 (59)	0.233[d]
Stolen money from home	16.7 (54)	33.3 (27)	0.252[d]	15.7 (127)	54.1 (37)	0.436[d]
Stolen anything (beside money) from home	6.7 (60)	13.3 (30)	—	6.3 (80)	15.6 (32)	0.045
Stolen from school	0 (10)	0 (6)	—	10.8 (37)	14.3 (28)	—
Stolen from other's home	20.0 (15)	50.0 (6)	—	8.6 (35)	18.8 (16)	—
Stolen from car	0 (2)	0 (4)	—	0 (11)	0 (8)	—
Avoided paying	(0)	0 (1)	—	20.0 (10)	15.4 (13)	—
Snatched purse/picked pocket	50.0 (2)	50.0 (2)	—	0 (2)	0 (9)	—
Aggression	60.2 (221)	59.1 (225)	0.262[d]	63.3 (400)	74.6 (339)	0.389[d]
Hit adult at school	0 (2)	0 (3)	—	44.4 (9)	36.4 (11)	—
Hit parent/caregiver	5.5 (55)	50.0 (6)	—	10.0 (10)	33.3 (3)	—
Hit sibling	46.3 (54)	12.9 (194)	0.115	46.0 (211)	47.5 (204)	0.242[d]
Hit other kids	18.7 (155)	58.0 (50)	0.389[d]	47.6 (271)	62.6 (206)	0.383[d]
Acted up in public	16.0 (25)	10.0 (40)	—	25.0 (52)	16.3 (80)	0.151[d]
Hidden weapon	0 (1)	(0)	—	50.0 (2)	10 (10)	—
Thrown rocks	0 (2)	0 (10)	—	11.8 (17)	4.7 (43)	—

(continued)

Table 5. (continued)

	White			Nonwhite		
	% SR given CR[a]	% CR given SR[b]	RIOC[c] (n = 499)	% SR given CR[a]	% CR given SR[b]	RIOC[c] (n = 689)
Status	34.5 (58)	34.5 (58)	0.258[d]	41.3 (138)	41.0 (139)	0.264[d]
Cheated on school test	30.0 (10)	11.5 (26)	—	50.0 (8)	8.2 (49)	—
Trespassed in other's yard	14.0 (43)	24.0 (25)	0.168[d]	17.2 (64)	21.2 (52)	0.131[d]
Run away	37.5 (8)	25.0 (12)	—	25.0 (4)	10 (10)	—
Skipped school	100 (1)	50 (2)	—	20.0 (10)	28.6 (7)	—
Sent home from school	66.7 (3)	66.7 (3)	—	48.5 (68)	50.8 (65)	0.454[d]
Substance use	13.3 (15)	14.3 (14)	—	13.9 (36)	17.9 (28)	0.133[d]
Sipped beer	0 (7)	0 (2)	—	5.0 (20)	11.1 (9)	—
Sipped wine	0 (6)	0 (3)	—	0 (7)	0 (2)	—
Sipped liquor	0 (2)	0 (1)	—	0 (2)	0 (4)	—
Smoked cigarette/chewed tobacco	40.0 (5)	100 (2)	—	21.4 (14)	37.5 (8)	—
Smoked marijuana	(0)	(0)	—	(0)	(0)	—
Sniffed glue	(0)	0 (8)	—	100 (1)	9.1 (11)	—

Note: RIOC = relative improvement over chance; SR = child self-report; CR = caretaker report. No significant differences (α = .05) were found in RIOC scores by race.

a.Number of caretakers reporting the behavior in parentheses.

b.Number of children self-reporting the behavior in parentheses.

c.Only calculated when cell sizes were greater than 5.

d.RIOC differs significantly from 0 at α = .05/9 = .0056 for Whites and α =.05/13 = .0038 for nonwhites.

be reported by the child given the caretaker reported the behavior and substance use was the least likely. In the younger cohort, the individual items with the greatest concordance were running away, cheating on school tests, and snatching a purse or wallet; however, all of these behaviors were rare according to caretaker reports. Behaviors that were more prevalent and still had a high level of concordance in the younger cohort were being sent home from school (41%), hitting a sibling (40%), hitting other kids (31%), and shoplifting (30%). For the older cohort, getting sent home from school (53%), hitting a sibling (52%), carrying a hidden weapon (50%), hitting an adult at school (43%), and hitting other kids (43%) had reasonably high levels of conditional prevalence. The conditional prevalence of self-reports given caretaker reports was generally higher for the older cohort.

Percent of caretakers reporting the behavior given the child self-reported the behavior. For both age cohorts, the percent of caretakers reporting a behavior given the child's self-reporting that behavior was highest for aggression and lowest for substance use. For the younger cohort, however, status offenses and not theft had the second highest conditional prevalence. The conditional prevalence of caretaker reports was slightly higher for older girls, except for theft.

RIOC. When disaggregated by age, the average RIOC for the 7-year-olds over the 9 individual behaviors where RIOC could be calculated was .204. RIOC scores were highest for the summary measures of aggression, followed by status offenses and theft. The individual items with the highest RIOC were being sent home from school, hitting other kids, and stealing money from home. In contrast, low reliability was observed for the summary measure, *vandalism*, where there were no more matches than would be expected based on chance. This was also true for several individual behaviors, including breaking family possessions, stealing anything (besides money) from home, hitting siblings, acting up in public, and trespassing.

The mean RIOC for the 8-year-olds, over the 9 individual behaviors for which this measurement could be calculated, was higher (.280) than that of the younger children. Acting up in public was the only behavior to have no more matches than expected by chance. Significance tests, however, indicated no significant differences in RIOC scores by age.

Race

Percent of children self-reporting the behavior given the caretaker report of that behavior. For the summary measures, race-specific findings for the conditional prevalence of self-reports were similar to those found for the total sample—convergence was highest for aggression and status offense and

lowest for substance use. For the summary measures as well as the majority of items, conditional prevalence was higher for nonwhites than for Whites, although the differences were small for aggression and substance use.

Percent of caretakers reporting behavior given the child self-reported the behavior. When disaggregated by race, the conditional prevalence for caretaker reports mirrored those from the total sample for Whites only. For nonwhites, conditional prevalence was highest for aggression followed by theft, not status offenses. As observed for self-reports, the conditional prevalence of caretaker reports for the summary measures was always greater for nonwhites.

RIOC. A similar pattern was observed when calculating RIOC for these groups. For Whites, RIOC scores were highest for aggression, followed by status offenses. In contrast, the highest scores for nonwhites were for aggression and theft. Vandalism had the lowest RIOC for both groups. When comparisons could be made, contrary to our hypothesis, nonwhites typically had higher RIOC scores than Whites. However, none of these differences was statistically significant; some of them (e.g., stealing money from home and theft) were large, however.

Convergence of Delinquency Rankings

In our last set of analyses, we ranked the girls on the basis of self- and caretaker reports of problem behavior and explored whether the instruments identified the same set of children as "high risk" across informants. The Spearman correlation between the children's rankings and the caregiver's rankings was .32 ($p < .001$). Approximately 19% of the 65 girls who fell in the top 5% of the "variety" ranking according to caretaker reports also ranked near the top according to self-reports. Although the agreement is small in magnitude, a chi-square test of independence indicated that this relationship is significant, $\chi^2(1) = 23.06$, $p < .001$.

When using data disaggregated by age and race, the Spearman correlations showed weaker associations between caretaker and child rankings for 7-year-olds and Whites ($r = .31$ and $r = .27$, respectively) and slightly stronger associations for 8-year-olds and nonwhites ($r = .33$ and $r = .32$, respectively). These findings were replicated when the variety scores were dichotomized. Approximately, 13% of Whites and 7-year-olds who ranked in the top 5% based on caretaker rankings also ranked in the top 5% based on self-reports. For nonwhites and 8-year-olds, this number was approximately 22%. For all comparisons, self-reported rankings were significantly related to caretaker rankings.

Conclusions

Our primary goal in this research was to assess the reliability of a unique data collection instrument, the ABS, which was used in the PGS to collect information on (a) young girls' self-reports of problem behavior, and (b) reports on child behavior from the girls' primary caretakers. Our research investigated the level of comprehension children exhibited when queried about a variety of problem behaviors and whether this knowledge varied by age and race. Results suggested that while comprehension did improve with age, children as young as age 7 had a relatively good grasp of the meaning of most behaviors, with the exception of substance use.

Contrary to our expectations about differences in comprehension by race, nonwhites generally exhibited better comprehension of the antisocial behaviors than did Whites, although these differences only obtained significance for marijuana use. Our expectation that Whites would show higher levels of comprehension was based on the assumption that more advantaged children (which in this sample would be disproportionately Whites) would be better educated, and therefore more knowledgeable, about antisocial behaviors. Instead, our findings imply that differences in the comprehension of antisocial behaviors might be tied to experience and exposure rather than education. This predication should be explored in future research. Nevertheless, our results suggest that the observed differences in reliability by race that have been found in some self-report studies were probably not due to differences in comprehension per se.

A second research task involved comparing involvement in problem behavior by age and race. Consistent with predictions, older girls tended to be more involved in problem behavior; however, these differences were significant for aggression and status offenses only. Caregiver reports painted a different picture, showing no significant age differences in involvement.

When considering race differences in prevalence, the opposite pattern of differences was found; there were more differences between groups when using caretaker rather than child reports. Child self-reports indicated that nonwhites were significantly more likely to be involved in damaging and status offenses, whereas caretaker reports indicated nonwhites were significantly more likely to engage in all behaviors except for substance use. Therefore, estimates of involvement by race only converged for damaging and status offenses. This could be driven by racial differences in caretakers' detection of their children's problem behaviors or their willingness to report the behavior. It may also be due to differences between Whites and nonwhites in their interpretation of what constitutes unacceptable behavior. These differences could

be due to racial differences in family structure, employment, or community characteristics, but additional research is needed to explore these ideas.

These results suggest that researchers interested in estimating the prevalence of behaviors in young children would reach different conclusions regarding race differences depending on the informant. Therefore, whenever possible, decisions regarding what groups to target when allocating services or resources should be based on reports from multiple informants (Lau et al., 2004). Moreover, these discrepancies suggest that the correlates of misconduct may differ depending on whether parent or child reports are used to construct the outcome.

Our last research task was to examine agreement in reported behavior between youths and their caregivers. Overall, the greatest concordance was achieved in the area of aggression and the least in substance use. This is consistent with predictions that more overt behaviors will be reported more reliably than covert behaviors. This pattern held regardless of age, race, and method used to calculate concordance. We also found greater concordance for the older cohort and for nonwhites, though none of these differences emerged as significant.

Finally, we found that for the most part, child and caretaker reports did not identify the same girls as high risk. Less than a quarter of girls who were involved in the most types of delinquency according to caretakers were also involved in the most types of delinquency according to self-reports. Given the relative importance placed on early identification of this group, such findings challenge researchers to learn more about the source of these differences.

Given the discrepancy between caretaker and child reports, it is important that future work examine methods of integrating the two sources of data. For example, some work has treated parent reports and child reports as two separate measures of delinquency, including both in the study (e.g. Gottfredson et al., 2006). Furthermore, Loeber and colleagues (1989) claim that information from adults (in addition to self-reports) is key to providing a more comprehensive view of youth antisocial behavior because there is a tendency for youth to underreport acts. More recently, some researchers have begun to develop theories regarding how data from multiple informants can be combined (see Kraemer et al., 2003).

As with most studies, our research has several limitations that could affect our results and conclusions. First, the analyses presented here are primarily descriptive and, therefore, do not address the mechanisms leading to the observed patterns of congruence. Although description is a necessary first step in understanding a phenomenon, researchers should explore the factors that might affect agreement between the reports of caretakers and

their young daughters, such as poverty, caretaker depression, or family characteristics. Future work is also needed that uses more refined measures of race and ethnicity.

Race differences in comprehension and concordance also deserve more systematic attention. For example, family and neighborhood characteristics might affect caretakers' and children's views as to whether certain behaviors are deviant or normal child's play. Such perceptions might lead to differences in comprehension and prevalence of antisocial behavior as well as differences in the reliability of reports by race. Future research should examine these potential causal mechanisms.

Additional research in this area is critical for learning more about nascent delinquent behavior and the degree to which multiple observer reports converge. Findings from this study suggest that our understanding of delinquency will improve when information is collected from both caretakers and children (Loeber et al., 1989). Using information from only one informant may lead to different conclusions regarding race and age differences in behavior than might be obtained using different or multiple informants. In addition, this type of information has practical utility. As Grietens and associates (2004, p. 145) suggest, information from multiple sources "helps to obtain a more comprehensive, reliable, and valid picture of the child" for clinical purposes.

Appendix: Computation of Reliability Measures

Computation of reliability measures are described using the following contingency table where "yes" indicates the informant reported the child engaged in the behavior and "no" indicates the informant reported the child did not engage in the behavior.

	Informant 2		
Informant 1	Yes	No	Total
Yes	a	b	e
No	c	d	n-e
Total	f	n-f	n

Conditional Probability

Percent of children self-reporting the behavior given the caretaker reported the behavior. Here Informant 1 is the parent and Informant 2 is the child, P(child|parent) = a/e × 100%.

(continued)

Appendix (continued)

Percent of caretakers reporting the behavior given the child reported the behavior. Here Informant 1 is the parent and Informant 2 is the child, P(parent|child) = a/f × 100%.

Relative Improvement Over Chance Scores (RIOC)

Formulas presented below are taken from Copas and Loeber (1990) and Farrington and Loeber (1989). RIOC is calculated as (na – ef)/(nf – ef), where e ≥ f. To test whether there are significant differences in RIOC across cohorts and races we use the following formula:

$$\frac{RIOC_1 - RIOC_2}{\sqrt{s_1^2 - s_2^2}}$$

RIOC1 and RIOC2 represent the RIOC scores from each sample and s_1^2 and s_2^2 are the variances of the RIOC scores from each sample. Variances are computed as

$$S_R^2 = \frac{nc(nf(n-e) + c(ne - ef - 2nf - n^2) + 2nc^2}{(n-e)^3 f^3}$$

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the authorship and/or publication of this article.

Funding

This research was supported by grants MH056630, MH73941, and K01 Award MH071790 from the National Institute of Mental Health; DA012237 from the National Institute of Drug Abuse; and 2005-JK-FX-0001 from the Office of Juvenile Justice and Delinquency Prevention.

Notes

1. Researchers have also been challenging this notion with regards to obtaining self-report data on victimization from young children (e.g., Hamby & Finkelhor, 2001).

2. Qualified households had English-speaking caretakers and 5-8 years old girls who were not severely developmentally delayed or deaf with no sign language skills.
3. There was an exception to this protocol: If the child did not understand the questions regarding marijuana use or glue sniffing, the question was immediately skipped to avoid raising the child's curiosity about these behaviors (Loeber et al., 1989).
4. Girls were also asked to report about their lifetime involvement. We focus on past year involvement because children are less likely to remember behaviors that occurred when they were very young.
5. RIOC scores are equivalent to kappa and phi scores that have been adjusted to account for the maximum number of matches possible (Farrington & Loeber, 1989).
6. This is identical to a χ^2 test of independence for a 2×2 table (Copas & Loeber, 1990).
7. If the 64 girls who did not understand the meaning of "hitting" are dropped from the analysis, the difference in reported prevalence between self- and caretaker-reports for aggression is no longer significant.
8. When girls who did not understand the question are dropped, the difference for aggression is no longer significant.
9. Differences in caretaker and child reports for vandalism did not remain significant when children who did not understand the behavior were dropped.
10. With low cell frequencies the variance of the RIOC (relative improvement over chance) can become large and confidence limits may take on values outside of 0 through 1 (Copas & Loeber, 1990).

References

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

Bachman, R., & Paternoster, R. (2004). *Statistics for criminology and criminal justice* (2nd ed.). New York: McGraw-Hill.

Barker, E. T., Bornstein, M. H., Putnick, D. L., Hendricks, C., & Suwalsky, J. T. D. (2007). Adolescent-mother agreement about adolescent problem behaviors: Direction and predictors of disagreement. *Journal of Youth and Adolescence, 36*, 950-962.

Chaiken, J. M., & Chaiken, M. R. (1982). *Varieties of criminal behavior.* Santa Monica, CA: RAND.

Copas, J. B., & Loeber R. (1990). Relative improvement over chance (RIOC) for 2x2 tables. *British Journal of Mathematical and Statistical Psychology, 43*, 293-307.

Elliott, D. S., Huizinga, D., & Ageton, S. S. (1985). *Explaining delinquency and drug use.* Beverly Hills, CA: Sage.

Farrington, D. P., Jolliffe, D., Loeber, R., & Homish, D. L. (2007). How many offenses are really committed per juvenile court offender? *Victims and Offenders, 2*, 227-249.

Farrington, D. P., & Loeber, R. (1989). Relative improvement over chance (RIOC) and phi as measures of predictive efficiency and strength of association in 2 × 2 tables. *Journal of Quantitative Criminology, 5*, 201-213.

Gottfredson, D., Kumpfer, K., Polizzi-Fox, D., Wilson, D., Puryear, V., Beatty, P., et al. (2006). The Strengthening Washington D.C. Families Project: A randomized effectiveness trial of family-based prevention. *Prevention Science, 7*(1), 57-74.

Grietens, H., Onghena, P., Prinzie, P., Gadeyne, E., van Assche, V., Ghesquière, P., et al. (2004). Comparison of mothers', fathers', and teachers' reports on problem behavior in 5- to 6-year-old children. *Journal of Psychopathology and Behavioral Assessment, 25*(2), 137-146.

Grills, A. E., & Ollendick, T. H. (2002). Issues in parent-child agreement: The case of structured diagnostic interviews. *Clinical Child and Family Psychology Review, 5*(1), 57-83.

Hamby, S. L., & Finkelhor, D. (2001). The victimization of children: Recommendations for assessment and instrument development. *Journal of the American Academy of Child and Adolescent Psychiatry, 39*, 829-840.

Hindelang, M. J., Hirschi, T., & Weis, J. G. (1981). *Measuring delinquency*. Beverly Hills, CA: Sage.

Hipwell, A. E., Loeber, R., Stouthamer-Loeber, M., Keenan, K., White, H. R., & Kroneman, L. (2002). Characteristics of girls with early onset disruptive and delinquent behavior. *Criminal Behavior and Mental Health, 12*(1), 99-118.

Huizinga, D., & Elliott, D. S. (1986). Reassessing the reliability and validity of self-report delinquency measures. *Journal of Quantitative Criminology, 2*, 293-327.

Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J., Boyce, W. T., & Kupfer, D. J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing methods and matching contexts and perspectives. *The American Journal of Psychiatry, 160*, 1566-1577.

Krohn, M. D., Thornberry, T. P., Rivera, C., & LeBlanc, M. (2001). Later delinquency careers. In R. Loeber & D. P. Farrington (Eds.), *Child delinquents: Development, intervention, and service needs* (pp. 67-93). Thousand Oaks, CA: Sage.

Lau, A. S., Garland, A. F., Yeh, M., McCabe, K. M., Wood, P. A., & Hough, R. L. (2004). Race/ethnicity and inter-informant agreement in assessing adolescent psychopathology. *Journal of Emotional and Behavioral Disorders, 12*, 145-156.

Loeber, R. (1982). The stability of antisocial and delinquent child behavior: A review. *Child Development, 53*, 1431-1446.

Loeber, R. (1987). The prevalence, correlates, and continuity of serious conduct problems in elementary school children. *Criminology, 25*, 615-642.

Loeber, R., & Dishion, T. (1983). Early predictors of male delinquency: A review. *Psychological Bulletin, 94*, 68-99.

Loeber, R., Green, S. M., & Lahey, B. B. (1990). Mental health professionals' perception of the utility of children, mothers, and teachers as informants on child psychopathology. *Journal of Clinical and Child Psychology, 19*(2), 136-143.

Loeber, R., Stouthamer-Loeber, M., van Kammen, W. B., & Farrington, D. P. (1989). Development of a new measure of self-reported antisocial behavior for young children: Prevalence and reliability. In M. W. Klein (Ed.), *Cross-national research in self-reported crime and delinquency* (pp. 203-225). Dordrecht, Netherlands: Kluwer Academic.

Matsueda, R. L. (1992). Reflected appraisals, parental labeling, and delinquency: Specifying a symbolic interactionist theory. *American Journal of Sociology, 97*, 1577-1611.

Moffitt, T. E., Caspi, A., Rutter, M., & Silva, P. A. (2001). *Sex differences in antisocial behavior*. Cambridge, UK: Cambridge University Press.

Thornberry, T. P. & Krohn, M. D. (2000). The self-report method for measuring delinquency and crime. In D. Duffee, R. D. Crutchfield, S. D. Mastrofski, L. G. Mazerolle, D. McDowall, & B. Ostrom (Eds.), *Measurement and analysis of crime and justice* (pp. 33-83). Washington, DC: U.S. Department of Justice.

Thornberry, T. P., & Krohn, M. D. (2001). The development of delinquency: An interactional perspective. In S. O. White (Ed.), *Handbook of youth and justice* (pp. 289-306). New York: Plenum.

Tremblay, R. E., Nagin, D. S., Seguin, J. R., Zoccolillo, M., Zelazo, P. D., Boivin, M., et al. (2004). Physical aggression during early childhood: Trajectories and predictors. *Pediatrics, 114*(1), 43-50.

Van der Meer, M., Dixon, A., & Rose. D. (2008). Parent and child agreement on reports of problem behaviors obtained from a screening questionnaire, the SDQ. *European Child and Adolescent Psychiatry, 17*, 491-497.

Weis, J. J. (1986). Issues in the measurement of criminal careers. In A. Blumstein, J. Cohen, J. Roth, & C. Visher (Eds.), *Criminal careers and career criminals* (Vol. 2, pp. 1-51). Washington, DC: National Academy Press.

Weisz, J. R., McCarty, C. A., Eastman, K. L., Chaiyasit, W., & Suwanlert, S. (1997). Developmental psychopathology and culture: Ten lessons from Thailand. In S. S. Luthar, J. A. Burack, D. Cicchetti, & J. R. Weisz (Eds.), *Developmental psychology: Perspectives on adjustment, risk, and disorder* (pp. 568-592). Cambridge, UK: Cambridge University Press.

Youngstrom, E., Loeber, R., & Stouthamer-Loeber, M. (2000). Patterns and correlates of agreement between parent, teacher, and male adolescent ratings of externalizing and internalizing problems. *Journal of Consulting and Clinical Psychology, 68*, 1038-1050.

Bios

Lee Ann Slocum is assistant professor of criminology and criminal justice at the University of Missouri, St. Louis. Her research focuses on intraindividual change and stability in offending and substance use over time. Her work has been published in *Criminology, Journal of Quantitative Criminology, and Journal of Research in Crime and Delinquency*.

Sally S. Simpson is professor and chair of criminology and criminal justice at the University of Maryland, College Park. Her research interests include corporate crime, criminological theory, and the intersection between gender, race, class, and crime. She is the chair (08/09) for the crime, law, and deviance section of the American Sociological Association; past president of the White-Collar Crime Research Consortium; and recipient of the American Society of Criminology's Herbert Bloch Award. She is an elected fellow of the ASC and received the Distinguished Scholar Award from the ASC Division on Women and Crime. Her current research examines corporate environmental crime and regulatory compliance and women's experience of violence (WEV).

Alison E. Hipwell is an associate professor of psychiatry and psychology at the University of Pittsburgh. She is a coinvestigator of the Pittsburgh Girls Study, and her research interests focus on the development of conduct problems and depression in girls, parenting influences, teenage sexual behaviors, pregnancy, and postpartum psychopathology.

Rolf Loeber, PhD, is distinguished university professor of psychiatry, and professor of psychology, and epidemiology at the University of Pittsburgh, and professor of juvenile delinquency and social development at the Free University, Amsterdam, Netherlands. He is codirector of the Life History Program and is principal investigator of three longitudinal studies, the Pittsburgh Youth Study, the Developmental Trends Study, and the Pittsburgh Girls Study. He is an elected member of the Koninklijke Academie van Wetenschappen (Royal Academy of Sciences) in the Netherlands and the Royal Irish Academy in Ireland. He has published widely in the fields of juvenile antisocial behavior and delinquency, substance use, and mental health problems.