A longitudinal analysis of parent and teacher ratings of problem behavior in boys with and without developmental delays

Meredith N Will
Seattle Pacific University, USA

Beverly J Wilson
Seattle Pacific University, USA

Date accepted: 16 February 2014

Abstract
This study investigated parent and teacher ratings of behavior problems in children with and without intellectual disabilities at three time points over the course of 9 years. The group of children with intellectual disabilities (IDs) had higher behavior problem ratings than the group of children with no IDs (NIDs) across the three time points. Parents and teachers generally agreed on behavior, with the exception of attention problems and externalizing problems. The ratings of problem behavior remained stable over the three time points. Our findings imply that children with ID may be more likely to be perceived as having greater externalizing behaviors by teachers than by parents and that parents may perceive children as having greater attention problems than teachers. For the majority of the subscales, the lack of differences between parent and teacher ratings over time implies relative stability of ratings over a 9-year period from young childhood through adolescence for individuals with and without ID.

Keywords
intellectual disability, behavioral measurement methods, parents, mental health

Corresponding author:
Meredith N Will, Seattle Pacific University, 2917 Erie Ave, Cincinnati, OH 45208, USA.
Email: meredithnwill@gmail.com
Introduction

Behavior problems in childhood negatively influence children’s school performance, social functioning, and adult outcomes (Kaiser, 2007; Kroes et al., 2002; Neilsen and McEvoy, 2004). Children with intellectual disabilities (IDs) have higher rates of behavioral problems than the children with no developmental delays or IDs (NIDs) (Baker et al., 2003; de Ruiter et al., 2007). For example, Baker and colleagues (2003) reported that preschool-aged children with developmental delays were three times more likely to have behavioral problems than their non-delayed peers. In an older sample of children (age 7–12), Dekker et al. (2002) found that 49% of children with ID were rated by parents as in the borderline to clinical range for problem behavior, compared to only 18% of the control group. They also examined teacher ratings of behavior problems and found similar results. In addition, previous research indicates that behavior problems in individuals with developmental and IDs tend to be relatively stable over time (Emerson et al., 2011; Gray et al., 2012).

Collection of data from multiple informants is the current gold standard for behavior ratings used in research and clinical assessment (Hunsley and Mash, 2007; Mash and Barkley, 2007). A multi-informant approach to behavior rating allows for a more comprehensive evaluation by providing evidence for problems across different settings (Finch et al., 2012). Studies investigating agreement across reporters have generally found low inter-rater concordance but have differed in their interpretation of differences between reporters. In a classic meta-analysis of multi-informant ratings of child behavior, low cross-informant agreement was found between parents and teachers, with a mean $r$ of 0.27 (Achenbach et al., 1987). Although this may be interpreted as a problematic measurement error, Achenbach et al. (1987) argued that each rater provides a valid description of the child’s behavior, thus supporting the notion that behaviors may vary across settings. More recently, Dumenci et al. (2011) provided further evidence that multi-informant discrepancies are attributable to variability in behavior across contexts.

In contrast to the interpretations of Achenbach and colleagues (Achenbach et al., 1987; Dumenci et al., 2011), other researchers have attributed multi-informant discrepancies to rater bias and strive to investigate methods of resolving discrepancies between raters (De Los Reyes and Kazdin, 2005; Ferdinand et al., 2003; Gresham et al., 2010). In addition, a number of studies have investigated whether ratings by parents or teachers provide more valid and reliable information. For example, Power et al. (1998) found teacher ratings had greater predictive validity than parent ratings for children’s ADHD symptoms. One potential reason for this advantage is that teachers typically have greater opportunity than parents to observe the behavior of multiple same-age children, such as in the classroom context. It is also possible that teachers are more likely to notice externalizing problems such as disruptive and hyperactive behaviors because these behaviors can disrupt the learning environment in classrooms. In contrast, it may be more difficult for teachers to notice internalizing behaviors in a classroom because these behaviors may be less apparent and are less disruptive than externalizing behaviors (Abikoff et al., 1993; Achenbach et al., 1987).

Other research indicates that parents tend to report more child problematic behaviors than teachers (Loeber et al., 1991). Loeber and colleagues (1991) examined parent and teacher ratings of problem behavior in 7- to 12-year-old boys and found that parents tended to report a greater prevalence of problem behaviors than teachers. However, when asked to rank the prevalence of specific problem behaviors, parents and teachers had similar rankings. Although parents and teachers may report differences in frequency of problem behavior, Loeber et al. (1991) argue that parents and teachers agree on which problem behaviors tend to occur more or less often for children.
**Teacher and parent reports for children with IDs**

Little parallel work comparing the ratings of teachers and parents has been completed for children with ID. In one of the few studies examining the concordance of parent and teacher ratings of problem behavior in school-age children with developmental delays, Keogh and Bernheimer (1998) found an average correlation of 0.43, suggesting a higher level of accordance between parents and teachers in children with developmental delays than in children with typical development. This study measured parent and teacher ratings of behavior using the Teacher–Child Rating Scale (T-CRS) and the Parent–Child Rating Scale (P-CRS; Hightower et al., 1988). Rather than two broadband internalizing and externalizing scales, as found in other assessments such as the Achenbach scales of measurement (Achenbach and Rescorla, 2001), the T-CRS and P-CRS define three clusters of behavior: acting out, shy/anxious, and learning problems. Parent–teacher correlations for the acting out scale were larger than for the shy/anxious scale (0.52 and 0.18, respectively). One potential explanation for higher agreement on the acting out scale may be that these behaviors are more prevalent in the delayed population, are easier for both parents and teachers to notice, and this possibly leads to greater agreement on behavior. In another study examining parent and teacher ratings of problem behavior in youth with ID, de Ruiter et al. (2008) found parents to have more stable ratings of behavior over a 5-year period compared with teachers whose ratings of behavior problems increased with child’s age. The authors discussed situational differences between home and school as a possible explanation for these differences in stability of ratings between parents and teachers.

**Purpose**

The purpose of the current study was to investigate parent and teacher ratings of behavioral problems for children with and without ID, across three times points spanning 9 years. The Teacher’s Report Form (TRF; Achenbach, 1991a; Achenbach and Rescorla, 2001) and parent report Child Behavior Checklist (CBCL; Achenbach 1991b; Achenbach and Rescorla, 2001) were administered at all three time points. We hypothesized that children with ID would have higher adult-rated behavior problems overall than children with NIDs. We also expected that parents would be more likely to report higher rates of behavior problems than teachers overall (Loeber et al., 1991). Further, we hypothesized that ratings of behavior problems would remain stable over time (Emerson et al., 2011; Gray et al., 2012).

**Method**

**Participants**

The current sample was originally recruited as part of a larger research project investigating the social skills and play interactions of preschool boys during play groups (Guralnick et al., 1996). The larger study did not investigate parent and teacher ratings in a longitudinal fashion. Therefore, the current study was drawn from the larger data set to address this issue. The original sample of children without delays was recruited from private and public preschools, day care providers, and newspaper advertisements. The sample of boys with ID was recruited from local diagnostic clinics and community service providers. The original study recruited boys only for two reasons: (a) boys were more readily recruited from the local clinics and (b) boys have higher rates of ID than girls do (Simpson et al., 2003).
**Participant criteria for original study.** Participants in the original study needed to meet the following criteria: (a) no major physical impairment, (b) no significant hearing loss or blindness, (c) lived with their current caregiver for at least 1 year, and (d) no significant motor deficits. Children were also excluded from the original study for having three or more siblings within 3 years of age and if they were already socially acquainted with another child in the study.

**Group assignment for the original study.** Children in the NID group were required to have Full Scale IQ scores on the Wechsler Preschool and Primary Scales of Intelligence–Revised (WPPSI; Wechsler, 1989) between 90 and 130. Exclusion criteria for the NID group included (a) standard scores for verbal IQ or performance IQ on the WPPSI-R <90, (b) total standardized scores on the Test for Auditory Comprehension of Language–Revised (TACL-R; Carrow-Woolfolk, 1985) <98, (c) total problem behaviors score from the CBCL (Achenbach, 1991b) in the 90th percentile or above, (d) enrollment in a preschool program in which more than 15% of the children had documented disabilities, or (e) having a sibling with a documented disability.

Participants in the ID group had cognitive delays that were usually of unknown etiology with three exceptions, one child had received radiation and chemotherapy treatments for cancer and two children had Down syndrome. The range of intellectual functioning of children within our ID group was clearly specified, in that all participants in the ID group had Full Scale IQ scores between 52 and 80 and were excluded for any of the following reasons: (a) WPPSI-R performance IQ >80, (b) total problems score from the CBCL greater than the 90th percentile, or teacher report of substantial disruptive behavior, or (c) TACL-R total score >90 or <55.

The two groups were matched on family socioeconomic status and chronological age (within 2 months). Additional information about this previous study and the extensive matching procedures can be found in articles published by Guralnick and colleagues (Guralnick et al., 1996).

**Current sample.** Participants from the original study were recontacted for the first assessment in the current study in 1993 (Time 1) and included 40 boys aged 6 to 8 years (20 with ID and 20 with NID; Guralnick et al., 1996). Participants were recontacted a second time and data were collected in 1997 (Time 2), and the third data collection point occurred in 2002 (Time 3). Ten participants were unable to participate in Time 2 or Time 3 assessments. Therefore, the final sample included 100% response rates for all three time points from 30 boys, 15 with IDs and 15 boys with NIDs. At the first data collection point, participants ranged in age from 6 years 1 month to 8 years 1 month ($M = 7$ years 1 month, $SD = 6$ months). Because several children were not included in the final sample, the groups were rematched on chronological age (within 4 months). Overall, demographic variables for families were very similar across the two groups. All child participants were Caucasian males. The median family income, which was measured at Time 1 in 1993 for both groups, was between US$40,000 and US$59,000 (range = US$10,000–US$80,000), and the average educational level for parents was 16 years (range = 12–22 years). Also, 90% of parents were married, and the average number of children living in the home was two with a range of one to five.

**Procedure**

Ethical approval for all procedures and measures used in this study was granted by the institutional review board of University of Washington, Seattle, Washington, USA. Children’s teachers and parents completed a questionnaire about children’s behavior in 1993, 1997, and 2002. Additional procedures completed by children for the larger study are described in previous publications (Guralnick et al., 1996).
Measure of problem behavior. At all three time points, children’s behavior was rated by teacher report using the TRF–school-age form (Ages 6–18) (Achenbach, 1991a; Achenbach and Rescorla, 2001) and parents using the CBCL–school-age form (ages 6–18) (Achenbach, 1991b; Achenbach and Rescorla, 2001). The TRF and CBCL provide one total problem behavior score, two second-order broadband scores (internalizing and externalizing behavior), and eight clinical subscales (anxious/depressed, somaticizing, withdrawn, social problems, thought problems, attention problems, delinquent behavior, and aggressive behavior). The internalizing broadband score is comprised of specific items from the withdrawn, somaticizing, and anxious/depressed subscales. The externalizing broadband score is comprised of specific items from the delinquent behavior and aggressive behavior subscales). All scores are represented by t scores in which <65 is the normal range, 65–69 is the borderline clinical range, and >69 is the clinically significant range.

The 1991 versions of the TRF and CBCL (Achenbach, 1991a, 1991b) were used at Time 1 and Time 2, and the updated 2001 (Achenbach and Rescorla, 2001) versions were used at Time 3. Reliability and validity information are well documented for the TRF and CBCL (Achenbach, 1991a, 1991b; Achenbach and Rescorla, 2001). For the 2001 forms, the internal consistency estimates for 6- to 18-year-old boys ranged from 0.72 to 0.97 and 0.71 to 0.97 for the TRF and CBCL forms, respectively. Test–retest reliability ranged from 0.85 to 0.90 and 0.88 to 0.90 for the TRF and CBCL, respectively. Inter-rater reliability estimates between the TRF and CBCL index and subscales are low, ranging from 0.15 to 0.36. Pearson correlations between the 1991 and 2001 versions of the subscales range from 0.97 to 1.0 for the CBCL and 0.92 to 1.0 for the TRF (Achenbach and Rescorla, 2001).

Results
Analyses were conducted using Statistical Package for the Social Sciences version 18.0 (SPSS Inc., Chicago, Illinois, USA). We first examined the total problem behavior score, followed by the second-order broadband scores (internalizing and externalizing), and finally the eight clinical subscales of the CBCL and TRF. A descriptive examination of the data is presented in Table 1.

Total problem behavior score
A 2 (developmental status: ID vs. NID) × 2 (rater: parent vs. teacher rating) × 3 (time: 1993, 1997, and 2002) analysis of variance (ANOVA) was conducted to evaluate the effects of developmental status and rater on the total behavior problem score of the CBCL and TRF across three time points. Results indicated a significant main effect for developmental status, $F(1, 168) = 21.89, p < 0.001$, partial $\eta^2 = 0.12$; more specifically, the ID group had higher total problem behavior scores than the NID group, as hypothesized. No main effects were found for reporter for total problem behavior scores, $F(1, 168) = 0.08, p = 0.78$, partial $\eta^2 = 0.00$, or across time regarding this score $F(2, 168) = 1.60, p = 0.20$, partial $\eta^2 = 0.02$. No significant interactions were found among these variables.

Internalizing and externalizing behavior broadband scores
Two separate $2 \times 2 \times 3$ ANOVAs were conducted to evaluate the effects of developmental status, rater, and time on both the internalizing and externalizing scores. A Bonferroni correction was used for both ANOVAs to control for Type I error. The ANOVA for the internalizing score indicated no significant main effects or interactions for status, rater, or time (see Table 2). The ANOVA for the externalizing score indicated no significant main effects for status, rater, or time; however, there
Table 1. Means (standard deviations) for ID and NID groups between parent and teacher ratings using the CBCL and TRF, across three time points.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Parent</td>
<td>Teacher</td>
<td>Parent</td>
</tr>
<tr>
<td>Total problems</td>
<td>57.3 (10.8)</td>
<td>53.7 (10.7)</td>
<td>58.5 (9.3)</td>
</tr>
<tr>
<td>Internalizing</td>
<td>50.3 (9.9)</td>
<td>52.6 (10.5)</td>
<td>51.9 (10.6)</td>
</tr>
<tr>
<td>Externalizing</td>
<td>51.0 (12.3)</td>
<td>53.3 (9.3)</td>
<td>53.0 (9.8)</td>
</tr>
<tr>
<td>Anxious/ depressed</td>
<td>53.1 (4.2)</td>
<td>55.7 (6.8)</td>
<td>54.6 (5.4)</td>
</tr>
<tr>
<td>Somatic complaints</td>
<td>56.6 (6.7)</td>
<td>54.7 (6.9)</td>
<td>54.3 (6.6)</td>
</tr>
<tr>
<td>Withdrawn</td>
<td>54.3 (6.5)</td>
<td>53.7 (5.3)</td>
<td>55.8 (6.0)</td>
</tr>
<tr>
<td>Social problems</td>
<td>63.1 (10.6)</td>
<td>53.9 (6.0)</td>
<td>59.9 (7.5)</td>
</tr>
<tr>
<td>Aggression</td>
<td>56.2 (7.2)</td>
<td>52.0 (14.1)</td>
<td>56.9 (6.9)</td>
</tr>
<tr>
<td>Delinquent behavior</td>
<td>53.5 (5.7)</td>
<td>56.2 (6.7)</td>
<td>54.1 (5.3)</td>
</tr>
<tr>
<td>Attention problems</td>
<td>63.1 (8.6)</td>
<td>55.3 (6.3)</td>
<td>61.4 (10.0)</td>
</tr>
<tr>
<td>Thought problems</td>
<td>58.2 (10.7)</td>
<td>58.6 (8.4)</td>
<td>57.8 (8.5)</td>
</tr>
</tbody>
</table>

ID: intellectual disability; NID: no intellectual disability; CBCL: Child Behavior Checklist; TRF: Teacher’s Report Form.
A significant interaction between status and rater $F(1, 168) = 4.71, p = 0.03$, partial $\eta^2 = 0.03$.

Because the interaction was significant, we examined the rater simple main effects—meaning, the difference between raters for the ID and NID groups separately. To control for Type I error across the two simple main effects, we set the $\alpha$ level for each at 0.025. The simple main effects test indicated that for the ID group, differences between parent and teacher ratings of externalizing behavior approached significance, although this was not significant at the 0.025 level, $F(1, 168) = 4.27, p = 0.04$, partial $\eta^2 = 0.04$, with teachers rating the ID group higher on externalizing behavior. No difference between parent and teacher ratings on externalizing behavior was found for the NID group, $F(1, 168) = 1.01, p = 0.32$, partial $\eta^2 = 0.01$. No other significant main effects or interactions were identified.

Eight clinical subscales

**Main effects.** A $2 \times 2 \times 3$ multivariate ANOVA (MANOVA) was conducted to evaluate the effects of developmental status, rater, and time on the eight clinical subscales of the Achenbach scales. A significant main effect was found for status, Wilks’s $\Lambda = 0.63, F(8, 161) = 11.82, p < 0.001$, partial $\eta^2 = 0.37$.

A series of one-way ANOVAs on the dependent variables were conducted to follow-up the significant MANOVA. Using the Bonferroni method, each ANOVA was conservatively tested at the 0.006 level. Follow-up analyses indicated that the ID group was rated significantly higher than the NID group on the social problems, $F(1, 168) = 63.68, p < 0.001$, partial $\eta^2 = 0.28$; attention problems, $F(1, 168) = 39.80, p < 0.001$, partial $\eta^2 = 0.19$; and thought problems, $F(1, 168) = 22.64, p < 0.001$, partial $\eta^2 = 0.12$, subscales (see Table 1).

A significant main effect was also found for rater, Wilks’s $\Lambda = 0.82, F(8, 161) = 4.56, p < 0.001$, partial $\eta^2 = 0.19$. Follow-up ANOVAs indicated that parents rated children significantly higher than teachers on the attention problems subscale, $F(1, 168) = 9.42, p = 0.003$, partial $\eta^2 = 0.05$ (see Table 1). No significant main effect was found for time, Wilks’s $\Lambda = 0.87, F(16, 322) = 1.48, p = 0.15$, partial $\eta^2 = 0.07$.

### Table 2. Summary of analysis of variance for status, rater, and time for the internalizing subscale.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>$F$</th>
<th>$p$</th>
<th>partial $\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Status</td>
<td>1</td>
<td>341.69</td>
<td>341.69</td>
<td>3.21</td>
<td>0.08</td>
<td>0.02</td>
</tr>
<tr>
<td>Rater</td>
<td>1</td>
<td>235.756</td>
<td>2.217</td>
<td>0.138</td>
<td>0.14</td>
<td>0.01</td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>130.067</td>
<td>1.223</td>
<td>0.297</td>
<td>0.30</td>
<td>0.01</td>
</tr>
<tr>
<td>Status × Rater</td>
<td>1</td>
<td>35.556</td>
<td>0.334</td>
<td>0.564</td>
<td>0.56</td>
<td>0.00</td>
</tr>
<tr>
<td>Status × Time</td>
<td>2</td>
<td>280.289</td>
<td>2.635</td>
<td>0.075</td>
<td>0.08</td>
<td>0.03</td>
</tr>
<tr>
<td>Rater × Time</td>
<td>2</td>
<td>20.556</td>
<td>0.193</td>
<td>0.824</td>
<td>0.82</td>
<td>0.00</td>
</tr>
<tr>
<td>Status × Rater × Time</td>
<td>2</td>
<td>2.422</td>
<td>0.023</td>
<td>0.977</td>
<td>0.98</td>
<td>0.00</td>
</tr>
<tr>
<td>Error</td>
<td>168</td>
<td>17,867.33</td>
<td>106.35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>180</td>
<td>521,792.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$df$: degrees of freedom; SS: sum of squares; MS: mean squares.
Interaction effects. Examination of the eight clinical subscales indicated a significant interaction between status and rater, Wilks’s $\Lambda = 0.86$, $F(8, 161) = 3.16$, $p = 0.002$, partial $\eta^2 = 0.14$. This interaction was significant for the delinquent behavior subscale, $F(1, 168) = 8.68$, $p = 0.004$, partial $\eta^2 = 0.05$. Because the interaction was significant, we examined the rater simple main effects—meaning, the difference between raters for the ID and NID groups separately. To control for Type I error across the two simple main effects, we used a procedure suggested by Green and Salkind (2004), whereby the previously adjusted $\alpha$ level of 0.006 was divided by 2, resulting in an $\alpha$ of 0.003 for each analyses. The simple main effects test indicated that there was a trend toward significance for the difference between parent’s and teacher’s ratings of delinquent behavior for the NID group; however, this was not significant at this conservative level ($M = 2.42$, 95% confidence interval [0.53, 4.31], $p = 0.01$).

An additional significant interaction was found in the overall multivariate for the eight clinical subscales, indicating a significant interaction between status and time, Wilks’s $\Lambda = 0.82$, $F(16, 322) = 2.01$, $p = 0.008$, partial $\eta^2 = 0.09$. Follow-up ANOVAs indicated a trend toward significance for the interaction between status and time for the anxious/depressed subscale, $(F(2, 168) = 4.20, p = 0.02)$, given the conservative $\alpha$ level of 0.006 set for follow-up ANOVAs. Significant interactions were not found between rater and time, Wilks’s $\Lambda = 0.93$, $F(16, 322) = 0.79$, $p = 0.70$, partial $\eta^2 = 0.04$, or between all three independent variables combined, Wilks’s $\Lambda = 0.92$, $F(16, 322) = 0.81$, $p = 0.68$, partial $\eta^2 = 0.04$.

Discussion

The results of our analysis generally suggest an overall lack of significant differences between parent and teacher ratings of problem behavior across three time points over 9 years for both ID and NID groups. No significant main effect was found for time in our analyses, suggesting that behavior problems were relatively stable over time for both groups. This finding is similar to previous research suggesting stability of problem behavior over time (Emerson et al., 2011; Gray et al., 2012).

A number of group differences were also identified. As hypothesized, the ID group had higher total problem behavior scores than the NID group overall. This finding is consistent with previous research demonstrating greater behavior problems in children and adolescents with developmental and intellectual disabilities (Baker et al., 2003; de Ruiter et al., 2007). There are a number of potential explanations for why individuals with ID have greater behavior problems, including communication difficulties and poor social skills (Guralnick et al., 1998; Marston et al., 1997). Researchers also assert that children with developmental disabilities and IDs often lack adaptive self-regulation skills, and this leads to greater challenging behaviors (Baker et al., 2007; Wilson, 1999). Although their scores for behavior problems were significantly higher than the NID group, the mean total problem behavior score for the group with ID was in the normal range at all three time points, with group mean scores ranging over time from 53.3 to 58.5 (see Table 1). However, 40% of participants in the ID group had total problem behavior scores above 60 (in the borderline or clinically significant range) compared to only 13% of participants in the NID group.

Further examination of the internalizing and externalizing scores suggested more subtle variations in parent and teacher ratings over this long period of time for the ID and NID groups. More specifically, there was a significant interaction between status and rater for the externalizing scale in that teachers rated the ID group as higher on externalizing behavior than parents did. It is possible that the school environment elicits more externalizing behavior due to greater requirements to sustain attention and follow directions (Frick et al., 2010). It is also possible that teachers are more
knowledgeable about what constitutes externalizing behaviors because they have numerous opportunities to observe and compare the behavior of different children in their classrooms.

Further analysis of the eight clinical factors indicated that the ID group had significantly higher ratings than the NID group for the social problem, attention problem, and thought problem subscales, but the groups did not differ significantly on the anxious/depressed, somatic complaints, withdrawn, aggression, or delinquent behavior subscales. The finding that the ID group was rated higher than the NID group on social problems and attention problems is similar to previous research (Baker et al., 2003; de Ruiter et al., 2007). Parents also rated children’s attention problems significantly higher than teachers across both groups. This finding is generally consistent with previous findings that parents tend to rate more behavior problems than teachers (Loeber et al., 1991). Our analyses examining other potential differences in parent and teacher reports were not conclusive. Although we found significant interactions between status and rater for the delinquent behavior and anxious/depressed subscales, follow-up analyses were not significant although there were trends toward significance in both cases. These findings could be related to our small sample size and our evaluation of results using an increasingly stringent $p$ value. Although not significant in our conservative analysis, future research examining the interactions between diagnosis, rater, and time may provide more insight into the patterns of behavior ratings for a population of developmentally delayed children over time.

This study had a number of strengths as well as several weaknesses. The major strengths of the present study were its use of a control group, multi-informant approach, a standardized behavior rating system, and its longitudinal nature. This study is one of the first to longitudinally examine parent and teacher agreement in a sample of children with ID. This study included a well-defined sample of boys with ID and a chronologically age-matched control group comprised boys with NID. In addition, the groups were similar in their family demographic characteristics.

Limitations of this study included use of a sample defined as ID with unknown etiology. In addition, two separate versions of the measure were used (1991 and 2001) over the course of the 9-year study. Because the correlations between all subscales for the two versions of the form are high, we believe that the forms are compatible for use for our analysis. We also excluded participants with very high ratings of problem behavior, which poses challenges to the external validity of our findings to children with more severe behavior problems. Our study was very limited in terms of race, ethnicity, and gender, limiting the generalizability of our findings. In addition, our study collected data at three time points over the course of 9 years, and, consequently, there were 3- to 5-year gaps in our analysis. It is possible that we missed fluctuations in behavior problems during the years we did not collect data. Another limitation of this study is its small sample size, which as discussed above, may have limited our analysis and interpretation.

Conclusion

The ID group was rated higher than the NID group on the social problems, attention problems, thought problems, and a score of total problem behavior. Analyses indicated that parents and teachers did not differ significantly on behavior ratings with the exception of the externalizing and attention problem subscales. Our findings imply that children with IDs may be more likely to be perceived as having greater externalizing behaviors by teachers than by parents. These findings also imply that parents may perceive children (both delayed and not delayed) as having greater attention problems than teachers.
Although parents and teachers did not differ significantly on the majority of the subscales, correlation analysis between parent and teacher ratings was low overall. One possible explanation for this is that different observers may perceive an individual’s behavior differently. Another possibility for low parent–teacher correlations is that children’s behavior may not be consistent across home and school environments. Finally, time was not a significant variable in our analyses. These findings highlight the relative stability of parent and teacher ratings over a 9-year period from young childhood through adolescence for individuals with and without ID.

**Funding**

This research was supported in part by grants from The Advocates for the Rights of Citizens with Developmental Disabilities (formerly the Association for Retarded Citizens) of Washington and the Department of Psychology at the University of Washington to Beverly J Wilson. Additional funding was supplied by The Gatzert Child Welfare and The W.W. Stout Fellowships.

**References**


