Unidimensionality and Developmental Trajectory of Aggressive Behavior in Clinically-Referred Boys: A Rasch Analysis

Mark D. Rapport · Scott V. LaFond · Stephen A. Sivo

Abstract A majority of research investigating aggression and its development in children relies on rating scales such as the Child Behavior Checklist (CBCL). These scales typically are developed using a conventional factor analytic approach for the selection and retention of scale items, but may contain insufficient items to assess the unidimensionality and developmental trajectory of youths' aggressive behavior. Rasch analysis was employed to determine the extent to which CBCL Aggressive and Delinquent clinical syndrome scale items reflect the unidimensionality and expected developmental trajectory of aggressive behavior based on parent endorsements of 455, 6 to 16 year old boys referred to community mental health centers. The two scales showed considerable promise as a unidimensional aggression scale and mimic the expected developmental pattern of aggressive behavior in extant literature. Future development of an aggressive CBCL dimensional subscale, however, must eliminate redundant and non-contributing items, and include severe aggressive behavior items exhibited by persistently aggressive youths.

Keywords Aggression · Rasch analysis ·
Developmental psychopathology ·
Child behavior checklist (CBCL)

M. D. Rapport (✉) · S. V. LaFond
Department of Psychology, University of Central Florida,
P.O. Box 161390, Orlando, FL 32816, USA
e-mail: mrapport@mail.ucf.edu

S. A. Sivo
Department of Educational Research, Technology,
and Leadership, University of Central Florida,
Orlando, FL 32816, USA

An expansive literature exists concerning aggressive behavior and its social ramifications (Broidy et al. 2003; Eklund and af Klinteberg 2003; Fergusson and Horwood 1998; Greenbaum et al. 1998; LeBlanc et al. 1991; Loeber and Farrington 2000; Viemerö 1996). Longitudinal studies consistently show strong continuity between aggressive behavior during childhood and adverse adult outcomes. Children rated high in aggression are at risk for school dropout (Fergusson and Horwood 1998; Greenbaum et al. 1998; Loeber and Farrington 2000), unemployment (Fergusson and Horwood 1998; Greenbaum et al. 1998; Loeber and Farrington 2000), substance abuse (Eklund and af Klinteberg 2003; Fergusson and Horwood 1998), and criminal behavior (Eklund and af Klinteberg 2003; Fergusson and Horwood 1998; Greenbaum et al. 1998; Loeber and Farrington 2000; Viemerö 1996). Approximately 40% of children with aggressive behavior problems during childhood fail to earn a high school diploma or GED (Greenbaum et al. 1998), and up to 60% engage in adult criminal behavior (Greenbaum et al. 1998; Loeber and Farrington 2000). Criminal behavior by each repeat offender costs society an estimated 1.7 to 2.3 million dollars in direct damages and societal resources (Cohen 1998). Collectively, these adverse outcomes create an imperative for the accurate and early identification of high-risk youths.

Rating scales are currently the most frequently used instrument for measuring aggressive behavior in children and adolescents. Most instruments that include aggression or conduct problems scales, such as the Child Behavior Checklist (CBCL; Achenbach 1991), Eyberg Child Behavior Inventory (ECBI; Eyeberg and Pincus 1999), and Behavior Assessment System for Children (BASC-2; Reynolds and Kamphaus 2004), were developed using conventional factor analytic methods to identify psychometrically appropriate items that reflect underlying constructs. Scale items are typically scored

based upon increasing frequency, intensity, or severity of the behavior (e.g., *not true*, *somewhat true*, and *often true*; Achenbach 1991). A potential shortcoming of the traditional factor analytic approach as applied to the developmental psychopathology field is its inability to reflect the inherent dimensionality of items within a construct (Bond and Fox 2001). For example, scale items such as "argues a lot" and "attacking others" may receive the same endorsement (e.g., a "3" indicating *often true*); however, physically attacking people is clearly a more serious form of aggressive behavior than arguing and nearly always emerges at a later stage of development (Achenbach 1991; Bird et al. 2005; Loeber and Hay 1997). This shortcoming may be particularly salient given the developmental progression of aggression in children.

Accumulating evidence over the past decade reveals a developmental progression of aggressive behavior in children with persistent conduct disturbance. For example, mild physical aggression tends to emerge before age three, and the percentage of boys whose parents endorse aggressive behaviors of increasing frequency and severity rises steadily through adolescence (Loeber et al. 1993, 1997, 1999; Zhang, Loeber and Stouthamer-Loeber 1997). This increase in youths with conduct problems likely reflects the combination of early-onset lifelong offenders and late-onset offenders, the latter of which begin exhibiting conduct disorder problems later during adolescence (Moffitt 1993; Patterson et al. 1992). In contrast, physical fighting emerges between 5 and 7 years of age, typically escalates in frequency between 10 and 14 years of age, and declines markedly between 15 and 16 years of age (Bird et al. 2005; Elliot 1994; Huizinga 1995; Loeber and Hay 1997; Zhang et al. 1997). Other forms of aggressive behavior evince a similar developmental trajectory. More serious forms of aggressive behavior, such as strong-arm tactics, unprovoked attacks on others, and forced sex occur infrequently before 10 years of age, and rapidly escalate in occurrence for a small subset of children between 12 and 14 years of age before peaking during late adolescence (Loeber and Hay 1997; Loeber et al. 1999; Loeber et al. 1997; Zhang 1997). Aggression, oppositionality, and property violations typically decrease with age, while status violations increase over time (Bongers et al. 2004; Tolan et al. 2000). The frequency and developmental trajectories of aggressive behaviors may be specific to particular cultures (Broidy et al. 2003). Loo and Rapport (1998) found significantly lower levels of aggressive behavior for children of Asian compared to children of Caucasian and Hawaiian ancestry, and aggressive behavior is strongly deplored in Chinese culture (Chen et al. 1998).

Collectively, extant research concerning the development of aggressive behavior in children reveals a relatively consistent picture, wherein less severe forms of aggressive behavior emerge then desist over a 10-year span, followed by an increase in more serious forms of aggressive behavior in a minority of youths (Cairns et al. 1989; Hoeve et al. 2008; Lahey et al. 2005; Loeber and Hay 1997; Tolan et al. 2000; Zhang et al. 1997). Similarly, a small percentage of non-violent offenders eventually engage in violent behavior as they transcend adolescence (Kjelsberg 2002). These findings have important implications for the development and measurement of aggression. For example, rating scales commonly employed to measure aggression in children and adolescents typically rely on a total score for determining aggression severity (Hinshaw and Park 1999). This practice largely ignores the accumulating evidence that particular types of aggressive behavior, such as lying and arguing, represent less severe forms of the construct and occur at an earlier age relative to more severe forms that involve attacking others and forcing sex (Loeber et al. 1992; Lahey et al. 2005; Loeber et al. 1999; Loeber et al. 1997; Tolan et al. 2000; Zhang et al. 1997). Currently available rating scales of aggressive behavior neither weight aggressive acts differentially according to their severity, nor reflect the established developmental trajectories associated with the changing topography of aggressive behavior. These scale properties are essential for the accurate and early identification of high risk children and adolescents based on collective knowledge concerning the developmental trajectory of aggression from childhood through adolescence.

Item Response Theory (IRT; Bond and Fox 2001; Embretson and Reise 2000)—an approach used extensively in intelligence and academic achievement scale construction—is used in the present study to examine two issues relevant to the measurement of aggressive behavior in youths by means of the psychometrically validated and universally used CBCL rating scale. An initial analysis was undertaken to determine whether CBCL Aggressive and Delinquent scale items collectively reflect a unidimensional or multidimensional construct, and the extent to which each scale item contributes in a meaningful manner to the measurement of aggressive behavior (i.e., construct validity). As such, construct validity in this study concurrently examines (a) whether scale items adequately represent and measure the underlying construct and (b) the frequency of parent-endorsed behavior problems.

Complementary analyses were conducted to determine whether select CBCL rating scale items mirror the developmental trajectory of aggressive behavior in youths. The intent of the analysis was to determine whether a more streamlined subset of aggressive items–corresponding with extant literature concerning the development of specific types of aggression in youths rather than factor analysis–would provide a stronger dimensional model of the aggression construct. If so, the item subset could potentially be used as a dimensional CBCL subscale to specifically estimate the severity of aggression in youths. Items reflecting commonly occurring and less severe forms of

aggressive behavior (e.g., arguing) were expected to yield the highest level of endorsement and appear at the low end of the scale metric. Less frequently endorsed, intermediate forms of aggressive behavior were expected to appear toward the middle of the logit scale; and the least commonly occurring, most severe forms of aggressive behavior (e.g., physically attacks people) were expected to appear at the high end of the scale to the extent that scale items reflect the expected developmental trajectory of aggression reviewed earlier.

Method

Participants

CBCL profiles were selected for all boys between 6 and 16 years of age from a total of 735 profiles collected over a period of 3 years (1992–1995) by each of the community mental health centers (CMHCs) in the state of Hawaii. These centers serve children residing in or attending public school within their respective catchment areas on six islands. Profiles for 455 boys age 6 to 16 years of age were retained for analysis. Profiles for all children aged 4 to 5 (n=47) and girls aged 6 to 16 (n=155) were not included in the present study due to insufficient numbers and a lack of extant research concerning the developmental trajectory of aggressive behavior in females. Sixty-nine profiles were rejected due to invalid usage based on age (i.e., younger than 4 or older than 16). Nine profiles were rejected due to elevated numbers of unendorsed items (i.e., greater than 21 items).

The CBCL was selected by the Hawaii Department of Health as part of its statewide assessment survey of all children receiving mental health services. Primary caregivers completed the instrument following their initial appointment at one of the CMHCs. Parents brought their children to the community mental health centers for a wide range of behavioral and emotional difficulties experienced at home and/or at school (i.e., not specifically for aggressive behavior). Assessment procedures varied across centers as is customary, with the exception that all were required to administer the CBCL and Child Behavior Checklist Teacher Report Form (Achenbach and Edelbrock 1986). The Hawaii Department of Health provided the archival data set to the investigative team to facilitate exploration of mental health patterns and ethnocultural variables among the sample after removing all names and identifying information from the protocols. The university IRB approved the study. For the present study's sample, ratings were obtained from 335 mothers (73%), 58 fathers (13%), and 46 other caregivers (10%). Rater was not specified for 16 (4%) of the profiles.

Parents or primary caregivers provided background information pertaining to ethnicity, parent educational level, and occupational status at intake. Mean age for the sample was 11.65 (SD=1.48). Socioeconomic status of each child's family (M=6.10, SD=2.83) was computed using the Duncan Index (Duncan 1961), which ranges from zero to 10. The sample was primarily indigenous to the Islands. Children's ethnicity was based on parental report and included the following groups: Pacific Islander (6%), Hawaiian/Part–Hawaiian (33%), Asian (9%), Mixed (17%), Hispanic (4%), Caucasian (19%), Black (2%), and not reported (10%). Children were considered 'Part-Hawaiian' if their ethnic background included any Hawaiian ancestry, and 'Mixed' if they could not be unambiguously assigned to one of the foregoing groupings.

Measures

The Child Behavior Checklist (CBCL) was designed to obtain caregiver ratings of children's behaviors over a 6-month period. The CBCL contains 118 behavioral descriptions that are rated for intensity/frequency using a 3-point Likert-type format (i.e., 0 = *not true*, 1 = *somewhat or sometimes true*, 2 = *very true or often true*). Standardized ratings permit separate normed comparisons by gender for three age groups: 4-5 years-olds, 6–11 year-olds, and 12–18 year-olds. *T*-scores are provided for a total problems score, two broad-band dimensions (Internalizing, Externalizing), 8 clinical syndrome scales (Social Withdrawal, Somatic Complaints, Anxiety/Depression, Social Problems, Thought Problems, Attention Problems, Delinquent Behavior, Aggressive Behavior), adaptive functioning, and DSM-oriented scales.

The CBCL Aggressive and Delinquent clinical syndrome scales used in this study contain a total of 39 items that describe a wide range of aggressive-type behaviors observed in children and adolescents such as arguing, cruelty towards others, and attacking others. Its psychometric properties are well established (Achenbach 1991). For example, test-retest reliability of the CBCL Aggressive subscale item responses for boys aged 6 to 11 and 12 to 16 is .95 and .87, respectively. The Delinquent subscale item responses also showed high test-retest reliability for boys aged 6 to 11 and 12 to 16 (i.e., .95 and .97, respectively). Factor stability for the Aggressive subscale scores over 6 and 18 months is relatively strong (i.e., .69 and .76, and .76 and .73 for boys aged 6 to 11 and 12 to 16 years, respectively). Factor stability for the Delinquent subscale scores over 6 and 18 months is also relatively strong (i.e., .71 and .80, and .71 and .61 for boys aged 6 to 11 and 12 to 16 years, respectively). Convergent validity for the CBCL's Aggressive clinical syndrome scale was established based on high estimated correlations of .84 and .88 with the Conners' Parent Questionnaire (Goyette et al. 1978) and Quay-Peterson Revised Behavior Problem Conduct Problem (Quay 1983) scale scores, respectively. The Delinquent clinical syn-

drome scale was found to have an estimated correlation of .61 with the Conners' Parent Questionnaire Antisocial Scale (Achenbach and Edelbrock 1978).

Results

The data analysis was completed using a two tier approach. All CBCL Aggressive (26 items) and Delinquent (13 items) clinical syndrome scale items were used in the first tier. These items were chosen to examine the logit distribution of all CBCL items intended to reflect aggressiveness in

children (see Table 1). An interval logit scale is based upon the frequency of item endorsement, with items located on a dimensional scale. Conventional IRT metrics were used to estimate the extent to which items fell sufficiently close to a theoretical centered line proceeding from −3 to +3 logit units based on the expected unidimensionality of the construct. Items that significantly diverge from this expected pattern of behavior or have excessive measurement error are considered unproductive and imprecise indices of aggression, respectively.

A subset of items used in the first tier–9 CBCL Aggressive and 8 Delinquent items, and the "cruelty to animals" item—was included in the second tier analysis

Table 1 Probability measures (logits) for all CBCL aggression and delinquency items

CBCL item	Abbreviation	Measure (logit)	OFMS	IFMS	Standard Error
Vandalism (106)	VAN	+2.39	0.76	0.90	0.14
Uses alcohol or drugs (105)	ALC	+2.37	1.65	1.21	0.14
Sets fires (72)	FIR	+2.29	1.28	1.06	0.14
Runs away from home (67)	RUN	+2.15	1.37	1.06	0.13
Truancy, skips school (101)	TRU	+1.57	1.66	1.30	0.12
Steals at home (81)	STH	+1.55	0.91	0.98	0.12
Steals outside the home (82)	STO	+1.50	1.03	1.00	0.12
Suspicious (89)	SUS	+1.27	0.98	0.97	0.12
Threatens people (97)	THR	+0.86	0.75	0.83	0.11
Physically attacks people (57)	ATK	+0.56	0.98	0.92	0.11
Screams a lot (68)	SCR	+0.49	1.08	1.06	0.11
Sulks a lot (88)	SUL	+0.39	1.03	1.02	0.11
Feels others are out to get him/her (34)	PSC	+0.37	0.95	0.99	0.11
Destroys things belonging to his/her family or other children (21)	DOT	+0.35	0.88	0.94	0.11
Destroys his/her own things (20)	DOW	+0.25	1.02	1.03	0.11
Not liked by other children (48)	UNL	+0.22	1.09	0.97	0.11
Nervous, highstrung, or tense (45)	NER	+0.10	1.13	1.05	0.11
Unusually loud (104)	LOU	+0.04	0.89	0.97	0.11
Talks too much (93)	TAL	+0.01	1.20	1.13	0.11
Easily jealous (27)	JEA	0.00	1.10	1.02	0.11
Gets in many fights (37)	FIG	−0.03	1.06	0.94	0.11
Cruelty, bullying, or meanness to others (16)	CRU	−0.08	0.81	0.84	0.11
Hangs around with children who get in trouble (39)	FRI	−0.09	1.10	1.11	0.11
Bragging, boastful (7)	BRA	−0.30	0.99	1.02	0.11
Doesn't get along with other children (25)	PER	−0.38	0.85	0.89	0.11
Teases a lot (94)	TEA	−0.42	0.93	0.91	0.11
Sudden changes in mood or feelings (87)	MOD	−0.73	0.88	0.92	0.11
Swearing or obscene language (90)	SWE	−0.73	1.00	1.03	0.11
Lying or cheating (43)	LIE	−0.78	0.88	0.96	0.11
Showing off or clowning (74)	SHO	−0.95	1.03	0.96	0.12
Temper tantrums or hot temper (95)	TEM	−1.05	0.73	0.91	0.12
Demands a lot of attention (19)	ATT	−1.14	1.02	0.98	0.12
Poor school work (61)	SCH	−1.20	1.37	1.16	0.12
Disobedient at school (23)	DOS	−1.46	0.85	0.96	0.13
Can't sit still, restless, or hyperactive (10)	HYP	−1.57	1.31	1.22	0.13
Disobedient at home (22)	DOH	−1.68	0.64	0.90	0.13
Stubborn, sullen, or irritable (86)	STU	−1.71	0.74	0.94	0.13
Impulsive or acts without thinking (41)	IMP	−1.98	0.61	0.84	0.14
Argues a lot (3)	ARG	−2.46	0.80	0.95	0.16

Numbers in parentheses following CBCL Aggression and Delinquency Scale items correspond with the question number on the CBCL; *IFMS* standardized infit mean square, *OFMS* standardized outfit mean square

based on their direct correspondence with the empirically documented trajectory of aggressive behavior in boys between 3 and 18 years of age (see Table 2; Loeber et al. 1992; Loeber and Hay 1997).

Caregiver scale item endorsements were recoded for all items used in the Tier I and Tier II analyses in the following manner. Endorsements indicating the occurrence of a described behavior on the Aggressive and Delinquent CBCL scales (i.e., 1 = *somewhat or sometimes true*; 2 = *very true or often true*) were coded as present (1 = present); endorsements indicating the nonoccurrence of a described behavior on the two scales were coded as not present (0 = not present) to better approximate the developmental levels in extant research. That is, aggressive behaviors were coded as occurring or not occurring as opposed to how frequently they occurred.

Both sets of CBCL scale items described above were subsequently tested using TestFact software (Bock et al. 2003) to determine whether they met recommended criteria as a unidimensional construct (Reckase 1979). The nonlinear factor analysis revealed differences greater than two eigenvalues between the first and second factor solutions. This finding indicates that item endorsements for the two CBCL scales (Aggressive, Delinquent) for the sample are best represented as a single underlying construct rather than the separate factors depicted by the CBCL.

Tier 1

The combined CBCL Delinquent and Aggressive scale items were analyzed using WinSTEPS version 3.57.2

(Linacre 2005). The Standardized Outfit Mean Square (OFMS) and Standardized Infit Mean Square (IFMS) were calculated for each of the items. Both OFMS and IFMS items with values between 0.5 and 1.5 are considered to be productive for measurement development; items with values between 1.5 and 2.0 are considered unproductive but not degrading to the analysis; and items with values greater than 2.0 are considered degrading to the analysis due to their unpredictability. Higher values indicate that the item characteristics differ from predicted values to a greater extent than expected by chance (Linacre 2005).

High OFMS values occur when caregivers of highly aggressive children fail to endorse expected low aggression items based on the developmental trajectory of aggressive behavior (Lahey et al. 1999; Loeber and Hay 1997). They may also occur when caregivers of children with low levels of aggressive behavior unexpectedly endorse high aggression items in their children. Inspection of the OFMS values revealed that "truancy" (1.66) and "alcohol use" (1.65) were found to be unproductive for measurement construction but not degrading to the overall analysis (i.e., between 1.5 and 2.0 logits), and were thus retained (Fig. 1).

High IFMS values—in contrast to high OFMS values—occur when caregivers fail to endorse severe aggressive behavior items in their children who otherwise exhibit high levels of aggressive behavior. High IFMS scores can also occur when caregivers fail to endorse mildly aggressive items in their children who display low levels of aggression. The IFMS scores for all aggressive items (see Table 1) were within the productive range (i.e., between 0.5 and 1.5 logits) and retained.

Table 2 Probability measures (logits) for all developmental research items

CBCL AGG item	Abbreviation	Measure (logit)	OFMS	IFMS	Standard Error
Vandalism (106)	VAN	+2.35	0.72	0.85	0.15
Sets fires (72)	FIR	+2.24	1.47	1.10	0.14
Cruelty to Animals (15)	CRA	+2.14	0.96	1.10	0.14
Runs away from home (67)	RUN	+2.08	1.36	1.05	0.14
Truancy, skips school (101)	TRU	+1.47	1.73	1.31	0.12
Steals at home (81)	STH	+1.44	0.89	0.90	0.12
Steals outside the home (82)	STO	+1.39	0.98	0.91	0.12
Physically attacks people (57)	ATK	+0.38	1.12	0.97	0.11
Gets in many fights (37)	FIG	−0.26	1.07	0.95	0.11
Cruelty, bullying, or meanness to others (16)	CRU	−0.31	0.82	0.86	0.11
Teases a lot (94)	TEA	−0.68	1.33	1.03	0.12
Swearing or obscene language (90)	SWE	−1.01	0.94	0.99	0.12
Lying or cheating (43)	LIE	−1.06	0.79	0.92	0.12
Temper tantrums or hot temper (95)	TEM	−1.36	0.84	0.98	0.12
Disobedient at school (23)	DOS	−1.80	0.84	0.96	0.13
Disobedient at home (22)	DOH	−2.04	0.60	0.85	0.14
Stubborn, sullen, or irritable (86)	STU	−2.08	0.92	1.06	0.14
Argues a lot (3)	ARG	−2.90	1.09	1.05	0.17

Numbers in parentheses following CBCL Aggression and Delinquency Scale items correspond with the question number on the CBCL; *IFMS* standardized infit mean square, *OFMS* standardized outfit mean square
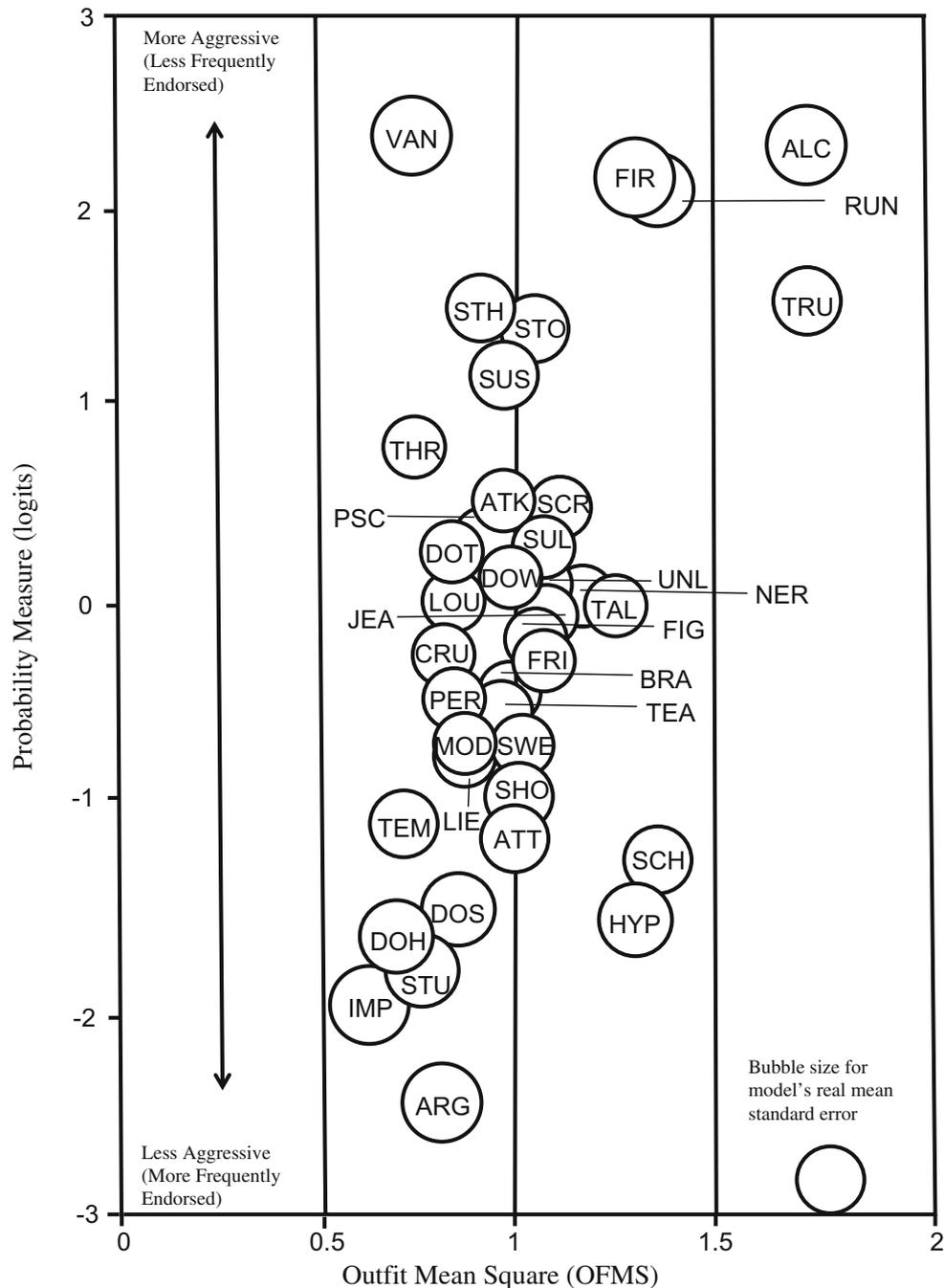
Fig. 1 Bubble Plot of Child Behavior Checklist (CBCL) Aggressive and Delinquent item endorsement probabilities ranging from low aggression/ high frequency item endorsements (*bottom*) to high aggression/low frequency item endorsements (*top*). Items between 0.5 and 1.5 on the abscissa are considered productive for measurement. CBCL behavioral item abbreviations include: ALC = uses alcohol or drugs; ARG = argues a lot; ATK = physically attacks people; ATT = demands a lot of attention; BRA = bragging, boastful; CRU = cruelty, bullying, or meanness to others; DOT = destroys things belonging to his/her family or other children; DOW = destroys his/her own things; DOH = disobeys at home; DOS = disobeys at school; FIG = gets in many fights; FIR = sets fires; FRI = hangs around children who get in trouble; HYP = can't sit still, restless, or hyperactive; IMP = impulsive or acts without thinking; JEA = easily jealous; LIE = lying or cheating; LOU = unusually loud; MOD = sudden changes in mood or feeling; NER = nervous, highstrung, or tense; PER = doesn't get along with other children; PSC = feels others are out to get him/her; RUN = runs away from home; SCH = poor school work; SCR = screams a lot; SHO = showing off or clowning; STH = steals at home; STO = steals outside the home; STU = stubborn; SUL = sulks a lot; SUS = suspicious; SWE = swearing or obscene language; TAL = talks too much; TEA = teases a lot; TEM = temper tantrums or hot temper; THR = threatens people; TRU = truancy or skips school; UNL = not liked by other children; VAN = vandalism

The logit distribution (Fig. 1) of all items was evaluated, wherein a logit represents the probability of item endorsement. Greater logit values are associated with less frequent behavioral endorsement and more serious forms of aggressive behavior. The probability of a child being endorsed as exhibiting a particular aggressive behavior is the difference between the subject ability logit level (their reported level of aggression) and the item logit value (the probability of item endorsement). For example, a moderately aggressive child has a 50% chance of being endorsed as exhibiting a particular type of aggressive behavior at the 0.0 logit level. The same child has a 27%, 5%, and 1% probability of engaging in rarer forms of aggressive behavior at the +1.0, +2.0, and +3.0 logit levels, respectively based on the normal t-distribution. Conversely the same moderately aggressive child has a 73%, 95% and 99% probability of being endorsed as exhibiting more common forms of aggressive behavior at the −1.0, −2.0, and −3.0 logit level, respectively.

Each CBCL rating scale item is represented by a bubble (see Fig. 1), with the diameter of the bubble representing two times the standard error of the item both above and below the logit level. Thus, an item falling at the 0.0 logit level with a standard error of 0.11 logits will have a diameter spanning from +0.22 to −0.22 logits. The diameter of the bubble represents the range within which the true measure of that item is 95% likely to occur. As a general rule, scale items with larger circles (i.e., standard errors that exceed 2 SDs of the overall model's real mean standard error) represent less precise estimates of the aggression construct (i.e., larger relative error in measurement).

Inspection of the item distribution revealed several items covering the lower logit levels including 8 items between the −1.0 and the −2.0 logit level (e.g., "disobeys at home", "disobeys at school", "being impulsive"), 10 items between −1.0 and 0.0 logits (e.g., "bragging, boastful", "swearing or obscene language", "lying or cheating"), and 11 items between the 0.0 and 1.0 logit level (e.g., "not liked by other children", "destroys his/her own things"). The distribution of overlapping scale items falling between the low to moderate level of aggression suggests that several items are redundant indices of aggression. For example, "being jealous", "excessive talking", and "being loud" all fall at 0.0, +0.01, and +0.04 logits, respectively. An excessive number of items in close proximity on the logit scale is conventionally considered undesirable when estimating a dimensional construct such as aggression (i.e., they provide limited to no improvement in the measurement of the underlying construct). In contrast, there were only four items (i.e., "suspicious", "stealing from home", "stealing from others", and "truancy") located between +1.0 and +2.0 logits, and four items (i.e., "runs away", "vandalism", "sets fires", and "alcohol use") reflecting aggressive behaviors between the +2.0 and +2.5 logit level. As noted previously,

two of these items (i.e., "alcohol use", "truancy") were unproductive, leaving only six items representing more severe aggressive behavior between +1.0 and +2.5 logits and no items above +2.5 logits. The overall model has a real mean standard error of 0.12 (SD=0.01), and all items except "argues a lot" fell within 2 SDs of this value. The item "argues a lot" had a standard error of 0.16, which is 4 SDs above the model mean.

Two additional metrics were calculated to estimate the replicability of the current findings. The Subject Reliability Index (SRI) estimates the replicability of subject ordering along a dimensional construct of aggression that one would expect if the current sample of caregivers were given another set of items measuring the same construct. The Item Reliability Index (IRI) estimates the replicability of item placements along the dimensional pathway if these same items were given to another sample of caregivers with similarly aggressive children. The analysis revealed that the obtained SRI (0.90) and IRI (0.99) values were within an acceptable range based on best practice parameters (Bond and Fox 2001).

Tier 2

The Standardized Outfit Mean Square (OFMS) and Standardized Infit Mean Square (IFMS) were calculated for each of the items in the item pool subset specifically representing the developmental trajectory of aggressive behavior (see Table 2). Interpretation of OFMS and IFMS values for the Tier 2 analysis (e.g., productive, degrading) is identical to that described for the Tier 1 analysis. Obtained results revealed only one item ("truancy" = 1.73) to be unproductive for measurement construction but not degrading to the overall analysis (i.e., OFMS value between 1.5 and 2.0 t-score units). This item was consequently retained for the analysis and likely reflects high rates of truancy endorsements by parents who otherwise report low overall aggression in their children. The IFMS scores for all developmental items were within the productive range (i.e., between 0.5 and 1.5 OMFS and IFMS t-score) and retained (see Table 2).

The overall standard error for the Tier 2 analysis was 0.13 (SD=0.01). Inspection of the bubble diameters in Fig. 2 suggest that the standard error for all items was appropriate (i.e., between 0.11 and 0.15) with the exception of "argues a lot." This item's high standard error of 0.17 (4 SDs above the model mean) suggests that it is an imprecise estimate of the aggression construct even when considered among a smaller and more specific subset of aggressive items.

The Tier 2 item distribution based on the developmental trajectory of aggression revealed three items between −3.0 and −2.0 logits (e.g., "argues a lot", "disobeys at home"),

Fig. 2 Bubble Plot of a subset of Child Behavior Checklist (CBCL) Aggression and Delinquent Scale that reflect the developmental trajectory of aggression in youths. Item endorsement probabilities range from low aggression/high frequency item endorsements (*bottom*) to high aggression/low frequency item endorsements (*top*). Items between 0.5 and 1.5 on the abscissa are considered productive for measurement. CBCL behavioral item abbreviations include: ARG = argues a lot; ATK = physically attacks people; CRA = cruelty to animals; CRU = cruelty, bullying, or meanness to others; DOH = disobeys at home; DOS = disobeys at school; FIG = gets in many fights; FIR = sets fires; LIE = lying or cheating; RUN = runs away from home; STH = steals at home; STO = steals outside the home; STU = stubborn; SWE = swearing or obscene language; TEA = teases a lot; TAN = temper tantrums or hot temper; TRU = truancy or skips school; and VAN = vandalism
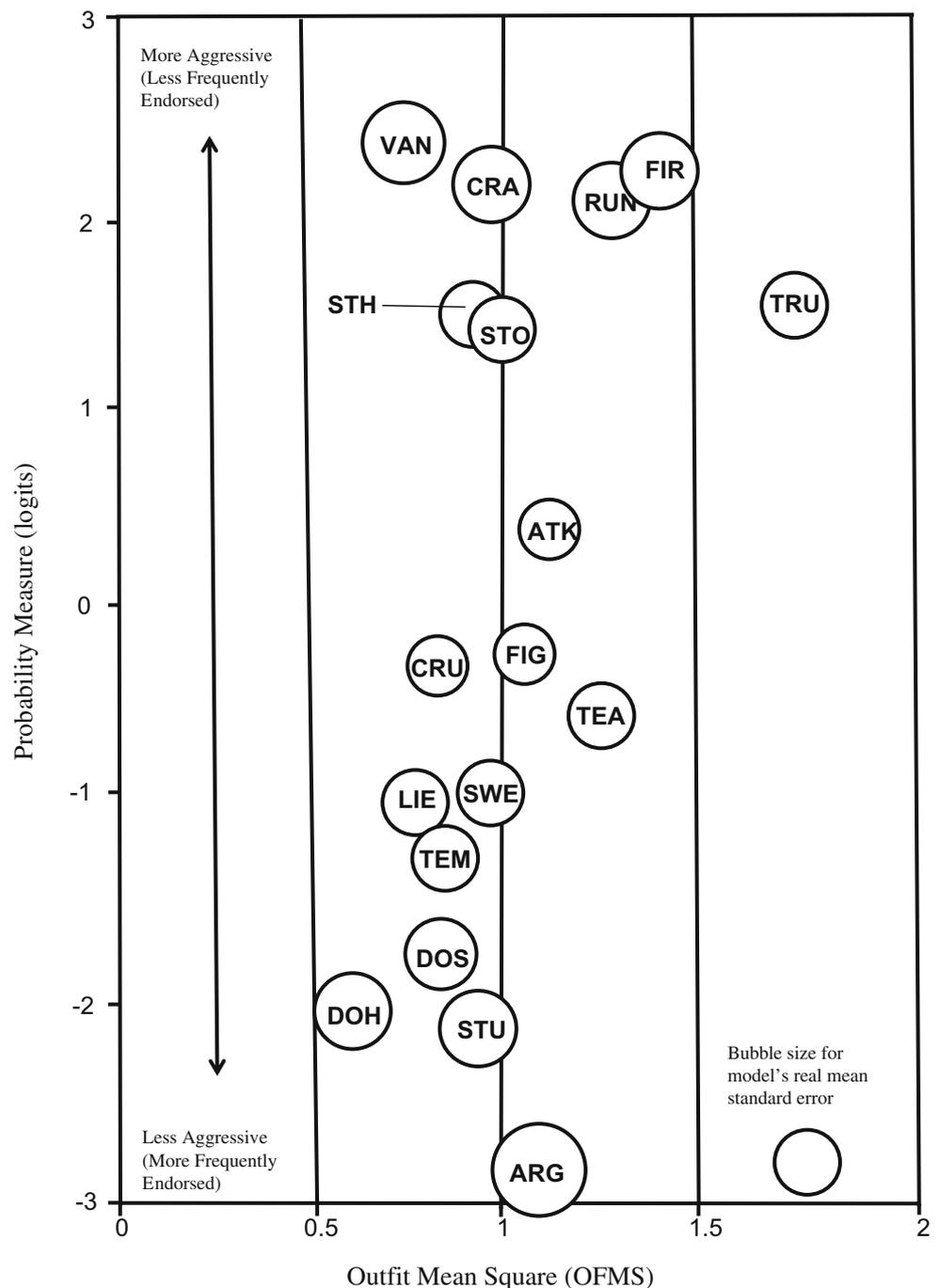
four items between −2.0 and −1.0 logits (e.g., "lying or cheating", "swearing or obscene language"), and three items between −1.0 and 0.0 logits (e.g., "cruelty/bullying others", "gets in many fights"). Only one item (i.e., "physically attacks people"), however, emerged between 0.0 and +1.0 logits (see Fig. 2), which indicates a significant gap in the item distribution for this level of aggressive behavior. Three (e.g., "steals from others", "truancy") and four (e.g., "runs away from home", "cruelty to animals") items were located between +1.0 and +2.0

logits and +2.0 and +2.35 logits, respectively. As in the previous analysis, "truancy" failed to contribute productively to the measurement of aggression due to its high endorsement in children who otherwise exhibit low levels of aggressive behavior. The Subject Reliability and Item Reliability Indices for the Tier 2 data were 0.79 and 0.99, respectively. The IRI was within the acceptable range, whereas the moderately lower SRI was likely due to having significantly fewer items measuring the underlying construct.

Discussion

The present study used Item Response Theory (IRT) to examine two complementary issues concerning the measurement of aggressive behavior in children. An initial set of analyses were conducted to determine whether CBCL Aggressive and Delinquent scale items reflect an underlying unidimensional aggression construct, and the relative contribution of unique scale items as measures of aggressive behavior (i.e., construct validity). Nonlinear factor analysis indicated that item endorsements for the two CBCL scales represent a single, higher-order construct of aggression in youths. This finding is highlighted by the similarity in endorsement frequency and placement for highly similar items such as disobedience at home (from the Aggressive scale) and at school (from the Delinquent scale) on the logit scale. Both items were endorsed as high frequency behaviors and within 0.22 logits of one another.

The contribution of CBCL Aggressive and Delinquent scale items as construct valid measures of aggressive behavior was evaluated by examining the extent to which items fell sufficiently close to a theoretical centered line proceeding from −3 to +3 logit units based on the demonstrated unidimensionality of the construct. Items that significantly diverge from this expected pattern of behavior or have excessive measurement error are considered unproductive and imprecise indices of aggression, respectively. Overall, the distribution of Aggressive and Delinquent scale items across the logit ordinate–ranging from −2.46 logits to +2.39 logits and falling within acceptable IFMS, OFMS and SE limits–indicates that the overwhelming majority of items contributed productively to the measurement of aggressive behavior in youths based on recommended best-practice parameters (Bond and Fox 2001; Linacre 2005). The few exceptions involved alcohol use and truancy–which neither contributed meaningfully nor degraded the measurement scale–and the item "argues a lot," whose large standard error rendered it an imprecise estimate of the aggression construct. The higher values for the former two items suggests a high rate of endorsement of truancy and alcohol use in children who otherwise exhibit low levels of aggressive behavior. This finding may reflect a behavior pattern unique to the sample studied (i.e., youths referred to community mental health centers in Hawaii). Alternatively, this pattern may suggest that truancy and alcohol use in youths are influenced by factors unrelated to aggressive behavior (Lahey et al. 1999; Maggs et al. 2008).

The overall pattern obtained using all CBCL Aggressive and Delinquent items highlights two potential shortcomings of using these scale items as a measure of unidimensional aggressive behavior in youths. There are insufficient items available to accurately measure aggressive behavior at the low (no items below −2.0 logits discounting the "argues" item) and high ends (only three items above +2.0 logits) of the scale, and a disproportionate number of items clustered at the middle of the scale (i.e., between −1.0 and +1.0 logits). Collectively, these findings suggest that it would be difficult to accurately assess aggressive behavior at the extreme ends of the scale when using the combined CBCL Aggressive and Delinquent items, but that one would have considerable confidence (and probable redundancy in measurement based on the item clusters) identifying youths within the middle range (between −1.0 and +1.0 logits).

An ensuing set of analyses examined a subset of item endorsement frequencies derived from the CBCL Aggressive and Delinquent scales that mapped onto the most commonly occurring (least severe) to the least commonly occurring (most severe) types of aggressive behavior reported in the literature (Loeber et al. 1992, 1997; Loeber and Hay 1997). These child by behavior estimates were expected to follow an orderly arrangement across the logit scale from negative (high frequency, common types of aggressive behavior) to positive (low frequency, rarer types of aggressive behavior) logits to the extent to which they mirror the expected developmental trajectory of aggressive behavior. Similar to the initial analysis of CBCL scale items, most of the subset scale items–with the exception of "truancy" and "argues a lot"—appear to be good indicators of the hypothesized path from the bottom to the top of the aggression scale based on recommended fit indices. The item "argues a lot" was again unproductive due to excessive measurement error, and "truancy" was unproductive due to its endorsement in youths with both high and low levels of aggressive behavior. This finding suggests that large numbers of youth in the current sample who exhibit primarily mild forms of aggression also skip classes at school.

The probability profile for the CBCL item subset based upon their logit distribution supports the dimensional progression of aggressive behavior described in the literature. Items representing authority conflict (i.e., argues a lot, disobeys at home, temper tantrums) appear to be endorsed more often in the entire sample, and items representing the high end of the overt and covert behavior problems pathways (i.e., steals in home, sets fires, and vandalism) are endorsed less often (Loeber et al. 1993). The results also suggest that many youths continue to display earlier, milder to intermediate forms of aggressive behavior even as they develop the more serious types of aggressive behavior. This finding is contradictory to several developmental studies that suggest that youths typically discard developmentally earlier types of aggressive behavior as they engage in more advanced types of aggressive behavior (Cairns et al. 1989; Loeber et al. 1997; Loeber and Hay 1997; Zhang et al. 1997).

The overall distribution of the developmental item subset was similar to the total item distribution, and indicated that

there were insufficient CBCL subset items for assessing average to high average levels of aggressive behavior in children. This finding reflects the exclusion of more advanced types of aggressive behavior reported in the literature for youths–running away from home, robbing, breaking and entering, forcing sex–on the CBCL. These types of aggressive/conduct disordered behaviors are typically first observed between 12 and 13 years of age, whereas the most severe types of aggressive behavior represented on the CBCL are first observed between 9 and 10 years of age. Inclusion of these and other later developing, severe forms of aggressive behavior may well push the currently high logit CBCL subset items to the average to high average level and fill the current void.

Collectively, the results indicate that many of the CBCL subset items may contribute meaningfully to the development of a specialized aggression rating scale; however, additional intermediate to severe range items will need to be developed and evaluated to ensure they possess adequate discriminative power for estimating the construct. The abundance of items in the higher part of the low range also suggests that several items are redundant and can be discarded.

A number of potential limitations must be recognized in the current investigation. Despite the relatively large sample size, the results reported herein may not generalize to mainland or other populations with different ethnic compositions and cultural practices as evidenced by previous cross-cultural studies of child psychopathology (Loo and Rapport 1998; Shore and Rapport 1998; Weisz et al. 1993). The sample was also limited to boys referred to community mental health clinics. Samples including younger children, girls, and non-referred children will need to be included in the development of a unidimensional aggression scale for youths. Finally, the Subject Reliability Index (SRI) for the developmental trajectory items subset (0.79) suggests that additional items may be necessary to ensure generalization of the results to other measurements of aggressive behavior in youths.

The Aggressive and Delinquent scales of the CBCL reflect an underlying unidimensional construct of aggression when evaluated concurrently. A subset of these items, selected to match descriptions of aggressive behavior that emerge over time in persistently aggressive youths, show considerable promise as a unidimensional aggression scale and mimic the expected developmental pattern in extant literature. Future development of an aggressive CBCL dimensional subscale will need to eliminate the identified redundant, imprecise, and non-contributing items, and add items that describe the more severe types of aggressive behavior typically exhibited by persistently aggressive youths older than 10 years of age. Information obtained using conventional measures of aggressive behavior in youth must be supplemented by direct interview and historical information to obtain information regarding the occurrence of more severe forms of aggressive behavior until dimensional scales are available for clinical use.

References

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 Child Profile*. Burlington, VT: University of Vermont.

Achenbach, T. M., & Edelbrock, C. S. (1978). The classification of child psychopathology: a review and analysis of empirical efforts. *Psychological Bulletin*, *85*, 1275–1301.

Achenbach, T. M., & Edelbrock, C. S. (1986). *Manual for the Teacher Report Form and teacher version of the Child Behavior Profile*. Burlington, VT: University of Vermont Department of Psychiatry.

Bird, H. R., Davies, M., Canino, G., Loeber, R., Rubio-Stipec, M., & Shen, S. (2005). Classification of antisocial behaviors along severity and frequency parameters. *Journal of Child and Family Studies*, *14*, 325–341.

Bock, R. D., Gibbons, R., Schilling, S. G., Muraki, R., Wilson, D. T., & Wood, D. T. (2003). *TestFact 4*. Lincolnwood, IL: Scientific Software International.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bongers, I. L., Koot, H. M., van der Ende, J., & Verhulst, F. C. (2004). Developmental trajectories of externalizing behaviors in childhood and adolescence. *Child Development*, *75*, 1523–1537.

Broidy, L. M., Nagin, D. S., Tremblay, R. E., Bates, J. E., Brame, B., Dodge, K. A., et al. (2003). Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: A six-site cross-national sample. *Developmental Psychology*, *39*, 222–245.

Cairns, R. B., Cairns, B. D., Neckerman, H. J., Ferguson, L. L., & Gariépy, J. (1989). Growth and aggression 1: childhood to early adolescence. *Developmental Psychology*, *25*, 320–330.

Chen, X., Hastings, P. D., Rubin, K. H., Chen, H., Cen, G., & Stewart, S. L. (1998). Child-rearing attitudes and behavioral inhibition in Chinese and Canadian toddlers: a cross-cultural study. *Developmental Psychology*, *34*, 677–686.

Cohen, M. A. (1998). The monetary value of saving a high-risk youth. *Journal of Quantitative Criminology*, *14*, 5–33.

Duncan, O. D. (1961). A socioeconomic index for all occupations. In A. J. Reiss (Ed.), *Occupations and social strata* (pp. 109–138). New York: Free Press of Glenco.

Eklund, J. M., & af Klinteberg, B. A. (2003). Childhood behaviour as related to subsequent drinking offences and violent offending: a prospective study of 11– to 14 year-old youths into their fourth decade. *Criminal Behaviour and Mental Health*, *13*, 294–309.

Elliot, D. S. (1994). Longitudinal research in criminology: Promise and practice. In E. G. M. Weitekamp, & H. J. Kerner (Eds.), *Crossnational longitudinal research on human development and criminal behavior* (pp. 189–201). Dodrecht, The Netherlands: Kluwer.

Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.

Eyeberg, S. M., & Pincus, D. (1999). *Eyeberg Child Behavior Inventory and Sutter-Eyeberg Student Behavior Inventory: Professional Manual*. Odessa, FL: Psychological Assessment Resources.

Fergusson, D. M., & Horwood, L. J. (1998). Early conduct problems and later life opportunities. *Journal of Child Psychology and Psychiatry*, *39*, 1097–1108.

Goyette, C. H., Conners, C., & Ulrich, R. F. (1978). Normative data on revised Conners parent and teacher rating scales. *Journal of Abnormal Child Psychology*, *6*, 221–236.

Greenbaum, P. E., Dedrick, R. F., Friedman, R. M., Kutash, K., Brown, E. C., Lardieri, S. P., et al. (1998). National adolescent and

child treatment study (NACTS): Outcomes for children with serious emotional and behavioral disturbance. In M. H. Epstein, K. Kutash, & A. Duchnowski (Eds.), *Outcomes for children and youth with emotional and behavioral disorders and their families: Programs and evaluation best practices* (pp. 21–54). Austin, Texas: Pro-Ed.

Hinshaw, S. P., & Park, T. (1999). Research problems and issues: Toward a more definitive science of disruptive behavior disorders. In H. C. Quay, & A. E. Hogan (Eds.), *Handbook of disruptive behavior disorders* (pp. 593–620). Dordrecht, Netherlands: Kluwer Academic.

Hoeve, M., Blokland, A., Dubas, J. S., Loeber, R., Gerris, J. R., & Laan, P. H. (2008). Trajectories of delinquency and parenting styles. *Journal of Abnormal Child Psychology, 36,* 223–235.

Huizinga, D. (1995). Developmental sequences in delinquency: Dynamic typologies. In L. J. Crockett, & A. C. Crouter (Eds.), *Pathways through adolescence* (pp. 15–34). Mahwah, NJ: Erlbaum.

Kjelsberg, E. (2002). Pathways to violent and non-violent criminality in an adolescent psychiatric population. *Child Psychiatry and Human Development, 33,* 29–42.

Lahey, B. B., Goodman, S. H., Waldman, I. D., Bird, H., Canino, G., Jensen, P., et al. (1999). Relation of the age of onset to the type and severity of child and adolescent conduct problems. *Journal of Abnormal Child Psychology, 27,* 247–260.

Lahey, B. B., Loeber, R., Burke, J. D., & Applegate, B. (2005). Predicting future antisocial personality disorder in males from a clinical assessment in childhood. *Journal of Consulting and Clinical Psychology, 73,* 389–399.

LeBlanc, M., Côté, G., & Loeber, R. (1991). Temporal paths in delinquency: stability, regression, and progression analyzed with panel data from an adolescent and a delinquent sample. *Canadian Journal of Criminology, 33,* 23–44.

Linacre, J. M. (2005). *WinSteps: Rasch measurement (Version 3.57.2).* [*computer program*]. Chicago, IL: Winsteps.com.

Loeber, R., & Hay, D. (1997). Key issues in the development of aggression and violence from childhood to early adulthood. *Annual Review of Psychology, 48,* 371–410.

Loeber, R., & Farrington, D. (2000). Young children who commit crime: epidemiology, developmental origins, risk factors, early interventions, and policy implications. *Development and Psychopathology, 12,* 737–762.

Loeber, R., Green, S. M., Lahey, B. B., Christ, M. A., & Frick, P. J. (1992). Developmental sequences in the age of onset of disruptive child behaviors. *Journal of Child and Family Studies, 1,* 21–41.

Loeber, R., Wung, P., Keenan, K., Giroux, B., Stouthamer-Loeber, M., & Van Kammen, W. B. (1993). Developmental pathways in disruptive child behavior. *Developmental Psychopathology, 5,* 101–132.

Loeber, R., Keenan, K., & Zhang, Q. (1997). Boys' experimentation and persistence in developmental pathways towards serious delinquency. *Journal of Child and Family Studies, 6,* 321–357.

Loeber, R., Wei, E., Stouthamer, M., Huizanga, D., & Thornberry, T. P. (1999). Behavioral antecedents to serious and violent offending: joint analyses from the Denver youth survey, Pittsburgh youth study and the Rochester youth development study. *Studies on Crime & Crime Prevention, 8,* 245–263.

Loo, S. K., & Rapport, M. D. (1998). Ethnic variations in the manifestation of child psychopathology: A cross-sectional, developmental study of Hawaii school children. *Journal of Child Psychology and Psychiatry, 39,* 567–575.

Maggs, J. L., Patrick, M. E., & Feinstein, L. (2008). Childhood and adolescent predictors of alcohol use and problems in adolescence and adulthood in the national child development study. *Addiction, 103,* 7–22.

Moffitt, T. E. (1993). Adolescence-limited and life-cycle-persistent antisocial behavior: a developmental taxonomy. *Psychological Review, 100,* 674–701.

Patterson, G. R., Crosby, L., & Vuchinich, S. (1992). Predicting risk for early police arrest. *Journal of Quantitative Criminology, 8,* 335–355.

Quay, H. C. (1983). A dimensional approach to behavior disorder: the revised behavior problem checklist. *School Psychology Review, 12,* 244–249.

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: results and implications. *Journal of Educational Statistics, 4,* 207–230.

Reynolds, C. R., & Kamphaus, R. W. (2004). *BASC-2: Behavioral Assessment System for Children* (2ndnd ed.). Circle Pines, MN: AGS Publishing.

Shore, G., & Rapport, M. D. (1998). The fear survey schedule for children-revised (FSSC-HI): ethnocultural variations in children's fearfulness. *Journal of Anxiety Disorders, 12,* 437–461.

Tolan, P. H., Gorman-Smith, D., & Loeber, R. (2000). Developmental timing of onsets of disruptive behaviors and later delinquency of inner-city youth. *Journal of Child and Family Studies, 9,* 203–220.

Viemerö, V. (1996). Factors in childhood that predict later criminal behavior. *Aggressive Behavior, 22,* 87–97.

Weisz, J. R., Suwanlert, S., Chaiyasit, W., Weiss, B., Achenbach, T. M., & Eastman, K. L. (1993). Behavioral and emotional problems among Thai and American adolescents: parent reports for ages 12 to 16. *Journal of Abnormal Psychology, 102,* 395–403.

Zhang, Q., Loeber, R., & Stouthamer-Loeber, M. (1997). Developmental trends of delinquent attitudes and behaviors: replications and synthesis across domains, time, and samples. *Journal of Quantitative Criminology, 13,* 181–215.