

DEVELOPMENTAL TRAJECTORY OF AGGRESSIVE BEHAVIOR IN CLINICALLY
REFERRED BOYS: A RASCH ANALYSIS

by

SCOTT VINCENT LAFOND
B.A. Southern Illinois University Carbondale, 2002

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science
in the Department of Psychology
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Fall Term
2008

© 2008 Scott LaFond

ABSTRACT

A majority of research investigating aggression and its development in children has relied upon the use of rating scales such as the Child Behavior Checklist (CBCL). These scales are typically developed using a conventional factor analytic approach for the selection and retention of scale items, but may not contain sufficient numbers of items to adequately assess the unidimensional construct or developmental trajectory of aggressive behavior in youths. The present study evaluates specific psychometric properties of CBCL Aggressive and Delinquency Problems clinical syndrome scale items to determine the degree to which they reflect the breadth and established developmental trajectory of aggressive behavior in youth.

TABLE OF CONTENTS

LIST OF FIGURES	v
LIST OF TABLES	vi
INTRODUCTION	1
METHODS	6
Participants.....	6
Measures	6
RESULTS	8
Tier 1	9
Tier 2.....	15
DISCUSSION.....	19
REFERENCES	24

LIST OF FIGURES

Figure 1: Probability bubble plot for all CBCL Aggression and Delinquency Scale items	14
Figure 2: Probability bubble plot for Developmental Research items.....	18

LIST OF TABLES

Table 1: Probability measures (logits for all CBCL Aggression and Delinquency Scale items... 13	13
Table 2: Probability measures (logits) for Developmental Research items..... 17	17

INTRODUCTION

An expansive literature exists concerning aggressive behavior and its social ramifications (Broidy et al., 2003; Eklund & af Klinteberg, 2003; Fergusson & Horwood, 1998; Greenbaum et al., 1998; Leblanc, Côté, & Loeber, 1991; Loeber & 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 dropping out of school (Fergusson & Horwood, 1998; Greenbaum et al., 1998; Loeber & Farrington, 2000), being unemployed (Fergusson & Horwood, 1998; Greenbaum et al., 1998; Loeber & Farrington, 2000), substance abuse (Eklund & af Klinteberg, 2003; Fergusson & Horwood, 1998), and participating in criminal behavior (Eklund & af Klinteberg, 2003; Fergusson & Horwood, 1998; Greenbaum et al., 1998; Loeber & Farrington, 2000; Viemerö, 1996).

Approximately 40% of children with aggressive behavior problems during childhood fail to graduate with a high school diploma or GED (Greenbaum et al., 1998), and up to 60% engage in adult criminal behavior (Greenbaum et al., 1998; Loeber & Farrington, 2000). Criminal behavior by repeat offenders costs society an estimated 1.7 to 2.3 million dollars in both direct damages and society 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 scales that measure aggression or more general conduct problems, such as the Child Behavior Checklist (Achenbach, 1991), were developed utilizing conventional factor analytic methods to identify psychometrically appropriate items that reflect underlying constructs. Scale items are typically scored based upon increasing frequency

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 & 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 & Hay, 1997). This shortcoming may be particularly salient based on extant research concerning the developmental nature 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 these behaviors rises steadily through adolescence (Loeber, Keenan, & Zhang, 1997; Loeber, Wei, Stouthamer, Huizanga, & Thornberry, 1999; Loeber et al., 1993; Zhang, Loeber, & Stouthamer-Loeber, 1997). This increase in youths with conduct disorder likely reflects the combination of early-onset lifelong offenders and late-onset offenders, the latter of which begin exhibiting CD problems later during adolescence (Moffitt, 1993; Patterson, Crosby, & Vuchinich, 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 & 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 & Hay, 1997; Loeber, et al.,1999; Loeber et al., 1997; Zhang et al., 1997). Aggression, oppositionality, and property violations typically decrease on the average with age, while status violations increase over time (Bongers, Koot, van der Ende, & Verhulst, 2004; Tolan, Gorman-Smith, & Loeber, 2000). These developmental trajectories of aggressive behaviors tend to be highly specific to particular cultures (Broidy et al., 2003).

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, Cairns, Neckerman, Ferguson, & Gariépy, 1989; Hoeve et al., 2008; Lahey, Loeber, Burke, & Applegate, 2005; Loeber & Hay, 1997; Tolan, Gorman-Smith, & Loeber, 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 & 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, Green, Lahey, Christ, & Frick, 1992; Lahey et al., 2005; Loeber, et al.,1999; Loeber et al., 1997; Tolan, Gorman-Smith, & Loeber, 2000; Zhang et al., 1997). Currently available rating scales of

aggressive behavior neither weight aggressive acts differentially relative to their severity, nor reflect the established developmental trajectories associated with the changing typography 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 & Fox, 2001; Embretson & 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 proven and universally used CBCL rating scale. An initial analysis was undertaken to determine whether CBCL Aggressive and Delinquency 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 focuses on the notion that scale items adequately represent and measure the underlying construct and by youth's demonstration of the behaviors in question. To accomplish this goal, item severity scores for the CBCL Aggressive and Delinquency clinical syndrome scale were converted to an interval logit scale using a conventional Rasch Model approach – 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 theoretically imagined 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. Inspection of individual items was also undertaken to determine whether they evidenced a wide spread rather than clustering at different parts of the scale (reflecting item redundancy), consistent with the dimensional nature of aggressive behavior.

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.

METHODS

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 three years (1992-1995) by each of the community mental health centers (CMHs) 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 474 boys age 6 to 18 years of age were retained for analysis. Profiles for children aged 4 to 5 (n = 47) and girls aged 6 to 16 (n = 155) were not included in the present study owing to insufficient numbers and a lack of extant research concerning the developmental trajectory of aggressive behavior in females. Nineteen 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 CMHs. 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.

Measures

The Child Behavior Checklist (CBCL) was designed to obtain parents' ratings of children's behaviors over a 6-month period. The CBCL describes 118 behaviors 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) and eight clinical syndrome scales (Withdrawn, Somatic Complaints, Anxious/Depressed, Social Problems, Thought Problems, Attention Problems, Delinquent, Aggressive Problems).

The CBCL Aggressive and Delinquency Problems clinical syndrome scales used in this study contain a total of 39 items that describe a wide range of aggressive-type behaviors commonly 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 aggression subscale items for boys aged 6 to 11 and 12 to 16 is .95 and .87, respectively. The delinquency subscale items 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 problems subscale 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 delinquency problems subscale 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 Problems clinical syndrome scale was established based on high correlations of .84 and .88 with the Conners Parent Questionnaire (Goyette, Conners, & Ulrich, 1978) and Quay-Peterson Revised Behavior Problem Conduct Problem (Quay, 1983) scales, respectively. The Delinquency clinical syndrome scale was found to have a correlation of .61 with the Conners Parent Questionnaire Antisocial Scale (Achenbach & Edelbrock, 1978).

RESULTS

The data analysis was completed using a two tier approach. All CBCL Aggressive (26 items) and Delinquency Problems (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). A subset of items used in the first tier – 9 CBCL Aggressive and 8 Delinquency Problems items, and the ‘sets fires’ item – was included in the second tier analysis based on their direct correspondence with the empirically documented trajectory of aggressive behavior in males between 3 and 18 of age (see Table 2) (Loeber et al., 1992; Loeber & Hay, 1997).

Parent 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 aggression and delinquency 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 unidimensional or multidimensional constructs (Reckase, 1979). The TestFact analysis revealed differences greater than two times the eigenvalue of the second factor between the first and second factors for both sets of scale items. This finding indicates that item endorsements for the

two CBCL scales (Aggression, Delinquency Problems) for the sample are best represented as a single underlying rather than multidimensional construct.

Tier 1

Both sets of CBCL scale items were subsequently 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 aggressive scale items. Both OFMS and IFMS are assessed on a scale measuring the t-distribution of the residuals from 0 to 2.0. Values between 0.5 and 1.5 are considered to be productive for measurement development; 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 parents of highly aggressive children fail to endorse expected items of low aggression based on the developmental trajectory of aggressive behavior (Lahey et al., 1999; Loeber & Hay, 1997). They may also occur when parents of children with low levels of aggressive behavior endorse unexpected items of highly aggressive behavior in their children. For the first set of aggressive items (see Table 1), 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 in the analysis.

High Standardized Infit Mean Squares (IFMS) values – in contrast to high OFMS values – occur when parents fail to endorse serious aggressive behavior items in their children who otherwise exhibit high levels of aggressive behavior. High OFMS scores can also occur when

parents fail to endorse mildly aggressive items in their children who display low levels of aggressiveness. 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.

The logit distribution (Figure 1) was evaluated using all aggressive items. A log unit represents the probability of item endorsement. The more positive the logit, the less often the aggressive behavior is endorsed, and is thus considered a more serious form of aggression. The probability of a child being endorsed as exhibiting a particular aggressive behavior is the difference between their reported level of aggression (subject ability logit level) and the item logit value. 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. 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 Figure 1), with the diameter of the bubble representing two times the standard error of the item above the logit level and below the logit level. Thus, an item falling at the 0.0 logit level with a standard error of 0.11 will have a diameter spanning from +0.22 to -0.22 logits. As a general rule, scale items with larger circles (i.e., standard errors that exceed 2 SDs) represent less precise estimates of the aggression construct (i.e., larger relative error in measurement). The diameter of the bubble represents the range within which the true measure of that item is 99% likely to occur.

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, 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). This distribution suggests that, although there are an appropriate number of items representing behaviors falling between the low to moderate level of aggression, several items appear to be redundant indices of aggression owing to their overlap with other scale items. 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., accounts for limited additional discriminative power to the interpretation of the underlying construct). In contrast, there were only 4 items (i.e., suspicious, stealing from home, stealing from others, and truancy) located between +1.0 and +2.0 logits, and 4 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 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.

The Subject Reliability Index (SRI) represents the replicability of subject ordering along a dimensional construct of aggression that one would expect if the current sample of parents were

given another set of items measuring the same construct. The Item Reliability Index (IRI) is a measure of the replicability of item placements along the dimensional pathway if these same items were given to another sample of parents 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 (Box & Fox, 2001).

Table 1: Probability measures (logits for all CBCL Aggression and Delinquency Scale items)

CBCL AGG item ¹	Abbrevia- tion	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

¹ Numbers following CBCL Aggression and Delinquency Scale (AGG) items correspond with the question number on the CBCL.

* = Items on the Delinquency Subscale of the CBCL

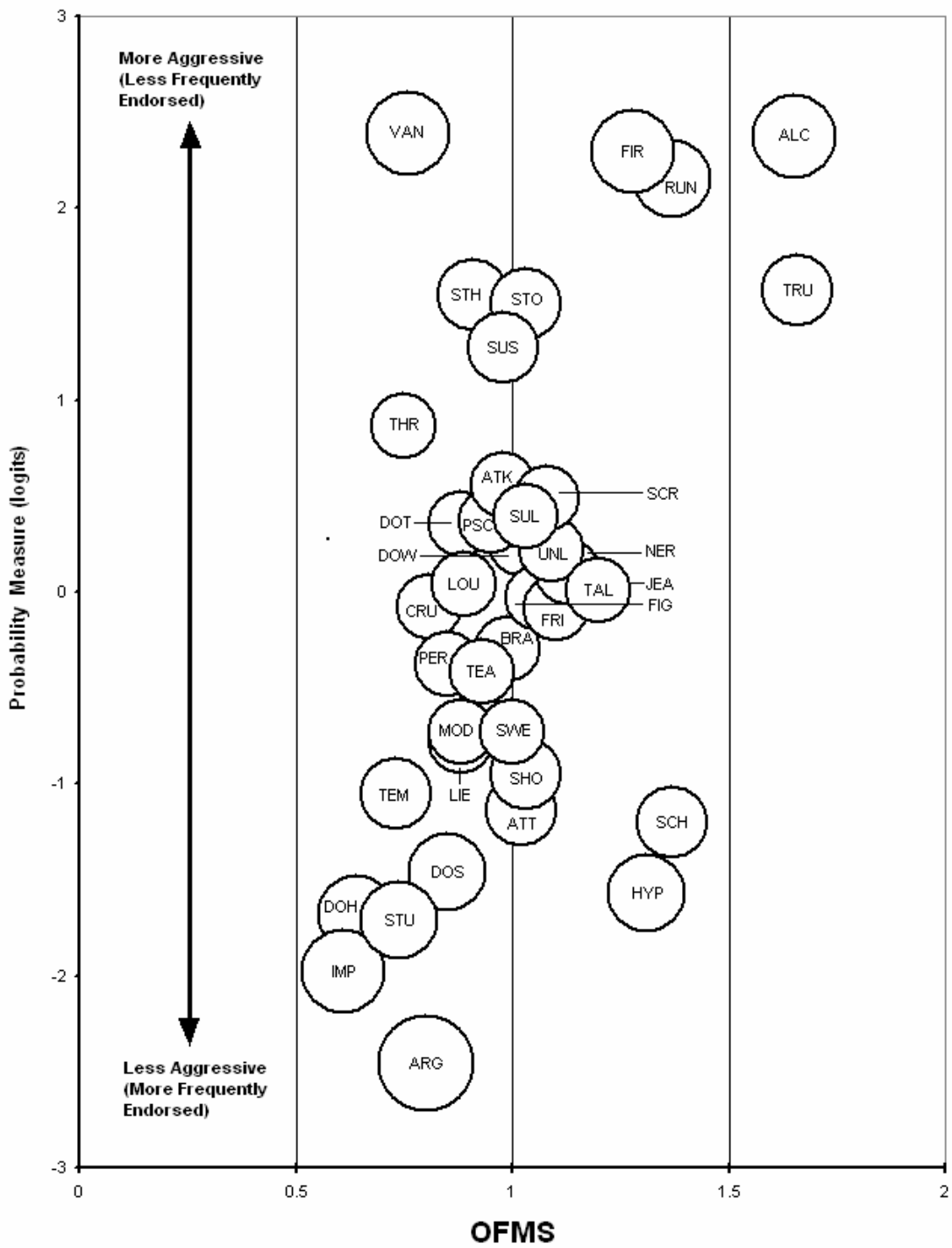


Figure 1: Probability bubble plot for all CBCL Aggression and Delinquency Scale items

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) follows the same reasoning as 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 OFMS and IFMS t-score) and retained (see Table 2).

The overall standard error (SE) for the Tier 2 analysis was 0.13 (SD = 0.01). Inspection of the bubble diameters in Figure 2 suggest that the standard error for all items fell within an acceptable range (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 (i.e., 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), 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). Three (e.g.,

steals from others, truancy) and four items (e.g., runs away from home, cruelty to animals) were located between +1.0 and +2.0 logits and +2.0 and +2.35 logits, respectively. Only one item (i.e., physically attacks people), however, emerged between 0.0 and +1.0 logits (see Figure 2), and indicates a significant gap in the item distribution for this level of aggressive behavior. 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 Person Reliability and Item Reliability Indices were calculated for the Tier 2 data and found to be 0.79 and 0.99, respectively. The lower PRI for the Tier 2 data is likely due to having significantly fewer items measuring the underlying construct.

Table 2: Probability measures (logits) for Developmental Research items

CBCL AGG item ¹	Abbrevia- tion	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

¹ Numbers following CBCL Aggression and Delinquency Scale (AGG) items correspond with the question number on the CBCL.

* = Items on the Delinquency Subscale of the CBCL

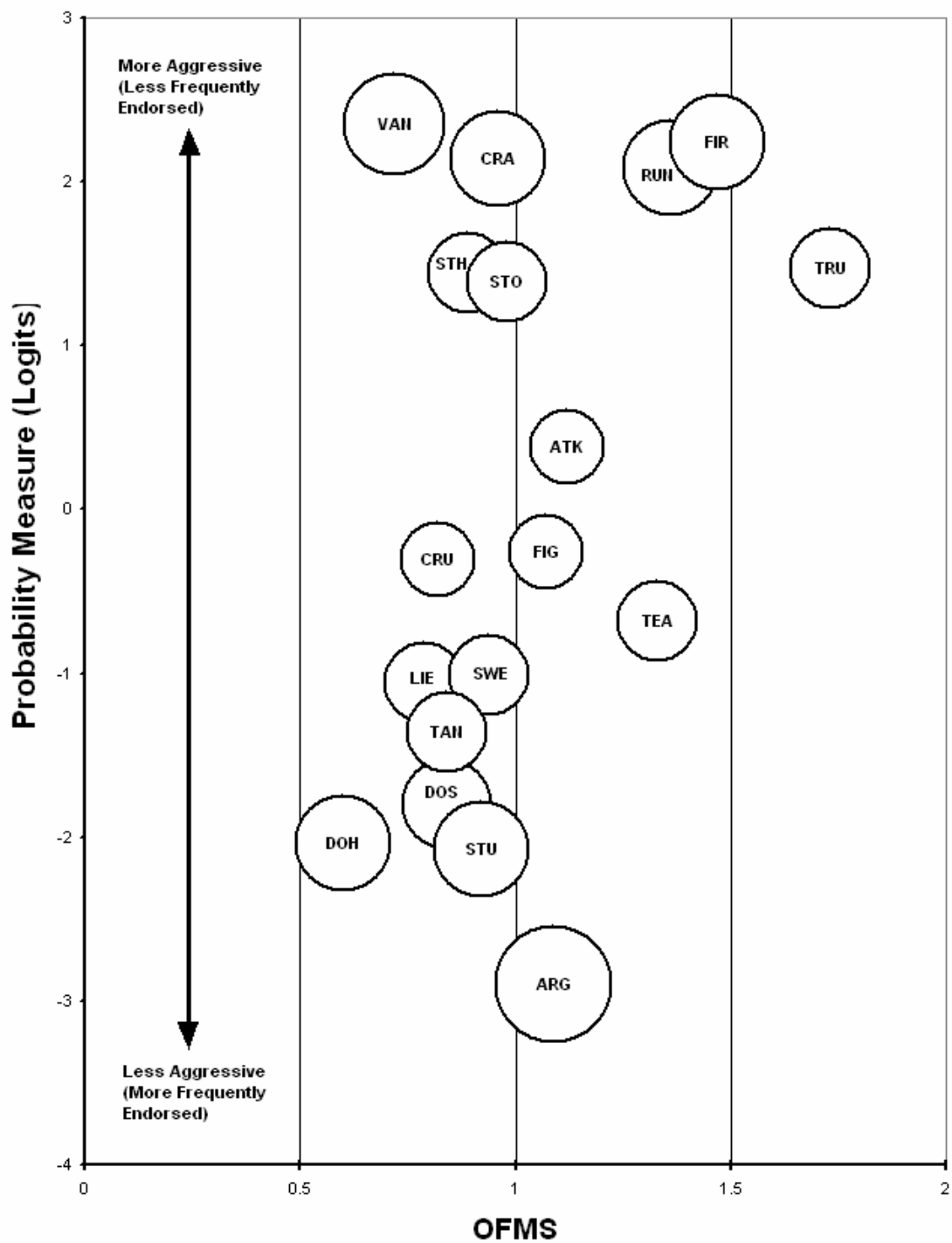


Figure 2: Probability bubble plot for Developmental Research items

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 Aggression and Delinquency clinical syndrome endorsed 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). Our findings, based on the TestFact analysis, indicated that item endorsements for the two CBCL scales represent a higher-order, unidimensional 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 Aggression scale) and at school (from the Delinquency 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 Aggression and Delinquency clinical syndrome scale items as construct valid measures of aggressive behavior was evaluated by examining the extent to which items fell sufficiently close to a theoretically imagined 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 Delinquency 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 youth based on recommended best-practice parameters (Bond & 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 larger relative standard error renders it an imprecise estimate of the aggression construct. The higher values for the former two items suggests a higher rate of endorsement of truancy and alcohol use in children who otherwise tend to 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), or that truancy and alcohol use in youths may be influenced by factors unrelated to aggressive behavior (Lahey et al., 1999; Maggs, Patrick, & Feinstein, 2008).

The overall pattern obtained using all CBCL Aggressive and Delinquency items highlights two potential shortcomings of the scale as a unidimensional measure of 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 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 Delinquency items, but that we 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 Delinquency clinical syndrome 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. 1997; Loeber et al., 1992; Loeber & 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 logits (low frequency, rarer types of aggressive behavior) to the extent to which they mirror the hypothesized 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 hypothetical path from the bottom to the top of the aggression scale based on recommended fit indices. The item ‘argues a lot’ was again hampered by excessive measurement error, and ‘truancy’ continued to pose a mismatch for the dimensional scale. Large numbers of youth who exhibit primarily mild forms of aggressive behavior apparently 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 & Hay, 1997; Zhang et al., 1997).

The overall distribution of the developmental item subset indicates a sufficient number of items representing low (below 0.0 logits) and high (above +2 logits) levels of aggressive behavior. The paucity of items between the 0.0 and +2.0 logits, however, suggests that there are an insufficient number of CBCL subset items available for assessing average to high average levels of aggressive behavior in children. It is interesting to note, however, that the more advanced types of aggressive behavior reported in the literature for youths are not reflected on the CBCL (e.g., runs away from home, mugs, breaking and entering, forces sex). These types of aggressive/conduct disordered behaviors are typically observed between 12 and 13 years of age, whereas the most severe types of aggressive behavior represented on the CBCL are 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 can be discarded owing to their redundancy.

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 & Rapport 1998; Shore &

Rapport, 1998; Weisz et al., 1993). Our sample was also limited to males referred to community mental health clinics. Samples including younger children, females, and non-referred children will need to be included in any broad-scale 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.

Two of the CBCL clinical syndrome scales – viz., the Aggression and Delinquency scales – 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 ten years of age.

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.
- 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(3), 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, Publishers.
- 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(5), 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(2), 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(2), 320-330.

Cohen, M.A. (1998). The monetary value of saving a high-risk youth. *Journal of Quantitative Criminology*, 14(1), 5-33

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(4), 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). Dordrecht, The Netherlands: Kluwer.

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

Fergusson, D.M., & Horwood, L.J. (1998). Early conduct problems and later life opportunities. *Journal of Child Psychology and Psychiatry*, 39(8), 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(2), pp. 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, Inc.
- 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 Publishers.
- Hobart, J. (2002). Measuring disease impact in disabling neurological conditions: Are patients' perspectives and scientific rigor compatible? *Current Opinion in Neurology*, *15*, 721-724.
- 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, New Jersey: Erlbaum.
- Kjelsberg, E. (2002). Pathways to violent and non-violent criminality in an adolescent psychiatric population. *Child Psychiatry and Human Development*, *33*(1), 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*(4), 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*(3), 389-399.

- Le Blanc, 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: Linacre.
- Loeber, R., & Farrington, D. (2000). Young children who commit crime: Epidemiology, developmental origins, risk factors, early interventions, and policy implications. *Development and Psychopathology*, *12*(4), 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*(1), 21-41.
- 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., Keenan, K., & Zhang, Q. (1997). Boys' experimentation and persistence in developmental pathways towards serious delinquency. *Journal of Child and Family Studies*, *6*(3), 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*(2), 245-263.
- 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.

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(1), 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(3), pp. 244-249.

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

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(2), 203-220.

Viemerö, V. (1996). Factors in childhood that predict later criminal behavior. *Aggressive*

Behavior, 22(2), 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(2), pp. 181-215.