



# Sex Differences in the Classification of Conduct Problems: Implications for Treatment

Areti Smaragdi<sup>1</sup>  · Andrea Blackman<sup>1</sup> · Adam Donato<sup>1,2</sup> · Margaret Walsh<sup>1</sup> · Leena Augimeri<sup>1,2</sup>

Received: 10 September 2019 / Revised: 4 August 2020 / Accepted: 13 August 2020 /  
Published online: 25 August 2020

© The Author(s) 2020

## Abstract

**Purpose** Conduct problem behaviors are highly heterogeneous symptom clusters, creating many challenges in investigating etiology and planning treatment. The aim of this study was to first identify distinct subgroups of males and females with conduct problems using a data driven approach and, secondly, to investigate whether these subgroups differed in treatment outcome after an evidence-based crime prevention program.

**Methods** We used a latent class analysis (LCA) in Mplus<sup>®</sup> to classify 517 males and 354 females (age 6–11) into classes based on the presence of conduct disorder or oppositional defiance disorder items from the Child Behavior Checklist. All children were then enlisted into the 13-week group core component (children and parent groups) of the program Stop Now And Plan (SNAP<sup>®</sup>), a cognitive-behavioral, trauma-informed, and gender-specific program that teaches children (and their caregivers) emotion-regulation, self-control, and problem-solving skills.

**Results** The LCA revealed four classes for males, which separated into (1) “rule-breaking,” (2) “aggressive,” (3) “mild,” and (4) “severe” conduct problems. While all four groups showed a significant improvement following the SNAP program, they differed in the type and magnitude of their improvements. For females, we observed two classes of conduct problems that were largely distinguishable based on severity of conduct problems. Participants in both female groups significantly improved with treatment, but did not differ in the type or magnitude of improvement.

**Conclusion** This study presents novel findings of sex differences in clustering of conduct problems and adds to the discussion of how to target treatment for individuals presenting with a variety of different problem behaviors.

**Keywords** Conduct problems · Treatment outcome · Classification · Sex differences

---

✉ Areti Smaragdi  
snapuk@childdevelop.ca

## Introduction

Conduct problems in childhood, characterized by defiant, rule-breaking, or violent behaviors, present a major burden for families and are associated with a myriad of negative consequences (Colman et al. 2009). Without successful intervention, these children are at a heightened risk for psychiatric problems later in life, leading to greater unemployment and economic difficulties, higher rates of alcohol and drug abuse, and increased risk of suicide (Baker 2013; Moffitt et al. 2001). Children who engage in violent or chronic behavioral problems from a young age are also more likely to engage in criminality later in life (Howell et al. 2014).

While several effective psycho-social interventions for conduct problems exist, unsurprisingly, some children are able to benefit more than others. One likely explanation is that conduct problems encompass a wide range of behaviors that in various combinations can make up a diagnosis of conduct disorder (CD) or oppositional defiant disorder (ODD) or otherwise lead to trouble at home, in school, or with the police. Although DSM-5 distinguishes CD and ODD as two separate disorders (the former is characterized by physical violence and delinquency and the latter by oppositionality and irritability), the two disorders greatly overlap (Lahey and Waldman 2012; Maughan et al. 2004). This overlap, in combination with the heterogeneity of both disorders, increases the challenges that clinicians and researchers face when planning for treatment and evaluating outcomes of treatment.

Several categorizations of conduct problems have been proposed (e.g., age-of-onset (Moffitt and Caspi 2001), callous and unemotional traits (Frick and White 2008), aggressive vs non-aggressive (Barker et al. 2007; Burt 2013)); however, these categorizations have mostly not used data-driven approaches. This group of statistical methods, including latent class analysis (LCA; Muthén and Muthén 2005), are exploratory in nature and are able to identify clusters of behaviors (or symptoms) that have a high probability of occurring together (Hagenaars and McCutcheon 2002). These types of classifications are argued to have increased ecological validity relative to other, hypothesis-driven, approaches.

A number of epidemiological studies have previously used data-driven approaches to classify conduct problems. In their influential study, Nock et al. (2006) used this approach in a population-based sample of adults and demonstrated that conduct disorder could retrospectively be categorized into five distinct classes: (1) rule violations, (2) deceit/theft, (3) aggression, (4) severe covert behaviors, and (5) pervasive conduct problems (Nock et al. 2006). Similar classes were found in a second retrospective epidemiological study including over 20,000 adults (Breslau et al. 2012). In an adolescent sample, Lacourse et al. (2010) found three distinct subgroups of conduct problems that were characterized by (1) rule violation, (2) physical aggression, and (3) severe/mixed conduct problems (Lacourse et al. 2010). Finally, a study investigating categorical versus continuous conceptualizations found that children aged 6 years old were characterized by (1) oppositional behavior, (2) aggression, and (3) irritability, while 10-year-old children were characterized by (1) disobedient behavior, (2) rule-breaking, (3) aggression, and (4) irritability (Bolhuis et al. 2017). While there is some overlap in themes between the different studies, clusters seem to be dependent on age, and replication using child samples is needed—particularly when attempting to apply the classifications to treatment.

There has been much debate as to whether conduct problems are best conceptualized based on severity or as categorized into distinct subgroups. Many have argued that a categorical conceptualization does not capture the complex nature of conduct problems (Bolhuis et al. 2017), and indeed, there is evidence to suggest that continuous models provide better fits for the data (Walton et al. 2011). However, the clinical utility of the dimensional approach has limitations (Coghill and Sonuga-Barke 2012). To the best of our knowledge, no study has investigated whether either of these conceptualizations makes any valuable contributions in terms of aiding treatment selection and planning, or predicting outcomes. Without investigation of the effect of evidence-based treatment on reducing various conduct problems, the clinical practicality of these different constructs remains intangible.

Interventions for conduct behavior problems are most effective when introduced in middle childhood (Piquero et al. 2016), include both child and caregiver components (Epstein et al. 2015), and are cognitive-behaviorally based (Lipsey et al. 2007). One such intervention is Stop Now And Plan (SNAP®), an extensively validated and cost-effective evidence-based program for middle years children, aged 6–11, with conduct problems (Augimeri et al. 2007; Farrington and Koegl 2015). SNAP was established in 1985 and became gender-specific in 1996 (Augimeri et al. 2017). Randomized controlled trials have demonstrated that the program significantly reduced scores on aggression, conduct behaviors, and internalizing problems for children both in the girl's program (Pepler et al. 2010) and the boy's program (Augimeri et al. 2007; Burke and Loeber 2015). In addition, SNAP is associated with improved emotion-regulation and problem-solving skills (Burke and Loeber 2016) and increased self-control (Augimeri et al. 2018) and has been found to increase recruitment of brain areas involved in self-regulation (Lewis et al. 2008; Woltering et al. 2015).

The SNAP Lab site, located at the Child Development Institute (CDI), a children's mental health center in Toronto, provides service to more than 120 children and their families per year, creating an ideal opportunity to investigate whether potential subgroups of children with conduct problems differ in terms of treatment outcome. In addition, most of the studies looking at classifications of conduct problems have not separated males and females in their models to allow for comparison. While males and females appear to share the underlying genetic and environmental influences on conduct problems (Burt 2009; Van Hulle et al. 2018), there is growing evidence to suggest that the two sexes partly differ in their neurobiological (González-Madruga et al. 2019; Smaragdi et al. 2017), neuropsychological (Sidlauskaite et al. 2018), and clinical (Ackermann et al. 2019) profiles. Throughout their development, these differences become more salient. Males are more likely to continue on an antisocial trajectory through adolescence and to develop antisocial personality disorder in adulthood, whereas females are more likely to diverge from an antisocial path and steer toward destructive and self-harming behavior in adulthood (Moffitt et al. 2001). The gender-specific program allows for a comparison between males and females who have undergone the equivalent treatment, with a reduced risk of introducing bias in the results.

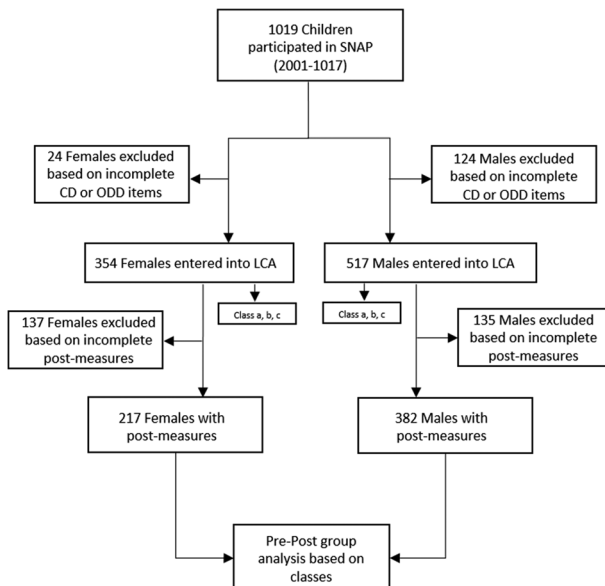
The current study aimed to first identify distinct subgroups of males and females with conduct problems and secondly to investigate whether these subgroups differed in the magnitude or type of treatment outcome following SNAP. We predicted that both males and females would show behavior profiles similar to those previously observed

in the literature (i.e., aggressive, rule-breaking, pervasive/severe). While we were not able to make specific predictions regarding treatment outcome prior to identification of the subgroups, based on the literature on differences between aggressive and non-aggressive conduct problems (Barker et al. 2007; Burt 2013; Burt and Donnellan 2008), we expected to see differences in the type and magnitude of the treatment response between the subgroups.

## Methods

The data were drawn retrospectively from a database of children aged between 6 and 11 years, who were enlisted into the SNAP program at CDI, between 2001 and 2017. A subset of the sample has previously been described in Augimeri et al. (2018); Augimeri et al. (2012); Jiang et al. (2011); and Pepler et al. (2010). During this time period, 1019 participants completed the 13-week group component of SNAP. Only participants with complete CD and ODD subscales of the Child Behavior Checklist (CBCL; Achenbach and Rescorla 2001) were included, leaving 871 participants (354 females and 517 males) for the initial LCA analysis. For the second analysis, the treatment evaluation, only participants with both pre- and post-assessment measures were analyzed, which included 278 females and 382 males (see Fig. 1). The study was approved by the CDI Research Ethics Board, and informed consent was obtained from the primary caregiver at the first assessment.

The CBCL was used to identify emotional and behavioral problems of the participants before and after the core group component of treatment. The checklist is composed of 112 “problem items” that are rated as 0 (not true of the child), 1



**Fig. 1** Participant selection and analysis breakdown. CD = conduct disorder, ODD = oppositional defiance disorder, LCA = latent class analysis, SNAP = Stop Now And Plan

(sometimes or somewhat true), or 2 (very true or often true), based on the child's behavior in the preceding 2 months. The items are grouped into three composite scales (externalizing problems, internalizing problems, and total problems). The externalizing composite scale is made up of the aggressive and rule-breaking behavior subscales. Selected items from these scales also make up the DSM-IV scales of ODD and CD. The overall inter-rater reliability of this tool is .96, and test-retest reliability has been found to be .90 and .92 for the externalizing and internalizing subscales, respectively (Achenbach and Rescorla 2001).

Consent and the CBCL pre-measure were completed by the primary caregiver at admission. The first assessment was conducted between 1 month and a few days prior to the first treatment session. CBCL post-measures were collected after the core SNAP group was completed (median time 3 months and 18 days). For consistency, only one informant per child was included in this analysis.

When there were two or more primary caregivers, priority was given to the caregiver who completed both pre- and post-measures. In cases where several caregivers had completed measures, the mother's information was selected. This decision was based on previous research showing only moderate agreement between mothers and fathers in rating their child's behavior. The differences in agreement were especially pronounced for female children (Davé et al. 2008). In order to reduce gender bias in reporting, which would hinder gender comparisons, mothers in this study were selected over fathers, as they constitute over 80% of respondents. While SNAP is a comprehensive treatment approach, the core component of the program includes 13 concurrent child and parent group sessions. The group component uses well-established strategies such as role-playing, cognitive restructuring, and reinforcement learning to teach children how to improve their emotion regulation and self-control and to "make better choices in the moment." The concurrent parent education group is informed by parent management training strategies. These strategies focus on strengthening the caregiver-child relationship, as well as enhancing emotion-regulation and effective parenting strategies (Forgatch and Patterson 2010; Kazdin et al. 1992).

## Analysis

LCA (Muthén and Muthén 2005) is a data-driven analysis method that provides distinct classes of symptoms (or problem behaviors) that have a high probability of occurring together. In addition, the model identifies the probability of each participant belonging to each class, allowing participants to be sorted into classes to use for group analysis. Models are run sequentially, starting with a 2-class model, and increasing in number until the best-fitting model is found (as assessed by the Bayes and Akaike Information Criteria (BIC and AIC) (Nylund et al. 2007) and Entropy). We ran the LCA in Mplus v.8.1 using binary ratings of the CD and ODD subscales of the CBCL (22 items in total). A score of 0 on an item was entered as 0; a score of either 1 or 2 was converted to a 1. Models were run for males and females separately.

Treatment outcome was assessed using the pre- and post-assessment raw scores on the aggressive, rule-breaking, total externalizing, and total internalizing subscales of the CBCL. These analyses were run in SPSS 24 for each class identified in the LCA. Differences in demographics and treatment outcome were analyzed using independent and mixed-model ANOVAs in SPSS 24.

## Results

### Sample Demographics

Males and females, overall, did not differ in age, number of aggressive, externalizing, or total symptoms scores for the CBCL pre-measure (mean age = 8.61 (SD = 1.61) and 8.55 (SD = 1.72) for males and females, respectively). However, males had higher scores of rule-breaking ( $F(1, 868) = 3.36, p = .01, d = .23$ ), whereas females had higher scores of internalizing problems ( $F(1, 868) = 3.13, p = .002, d = .21$ ).

### Latent Class Analysis

**Males** Based on the BIC, AIC, and Entropy, the best-fitting model for the data was a four-class model (BIC = 10,243.05, AIC = 9856.48, Entropy = .79, LMR-LRT  $p < .01$ ). Prevalence and symptom probabilities for the four classes can be seen in Table 1. Class 1, “rule-breaking,” was characterized by stealing both at home and outside of home, lying, and breaking rules. Class 2, “aggressive,” was characterized by cruelty toward people, fighting, physically attacking, breaking rules, and threatening other people. Class 3, “mild,” showed the lowest levels of conduct problems, mainly presenting with disobedience, rule-breaking, stubbornness, and temper tantrums. Finally, Class 4, “severe,” was characterized by endorsing almost all problem items from the two subscales and had the highest endorsement of fire setting, vandalism, running away, and cruelty to animals. The differences in symptomology between the four groups were confirmed by the analysis at pre-assessment (rule-breaking  $F(513, 3) = 208.65, p < .001$ ; aggressive  $F(513, 3) = 158.76, p < .001$ ; externalizing  $F(513, 3) = 221.69, p < .001$ ; internalizing  $F(513, 3) = 2.40, p < .001$ ). The “mild” group had significantly lower levels of aggression, rule-breaking, externalizing, and internalizing relative to the other three classes (all  $p < .001$ ). Similarly, the “severe” group scored significantly higher than the other three groups on these measures (all  $p < .001$ ). As expected, the “rule-breaking” group scored significantly higher than the “aggressive” group in rule-breaking ( $p < .001$ ), while the reverse was true for levels of aggression ( $p < .001$ ). There were also main effects of age ( $F(3, 513) = 3.83, p < .05$ ), where participants in the “aggressive” group were significantly younger than in the “rule-breaking” group.

**Females** A two-class model was the best-fitting model of the female data (BIC = 715.6, AIC = 7324.7, Entropy = .81, LMR-LRT  $p < .001$ ). Classes 1 “mild” and 2 “severe” were characterized by the same type of behavior, but the “severe” group had significantly higher scores across all measures at pre-assessment (rule-breaking  $t(351) = 13.40, p < .001$ ; aggressive  $t(351) = 14.27, p < .001$ ; externalizing  $t(351) = 18.27, p < .001$ ; internalizing  $t(351) = 3.41, p < .001$ ) and additionally endorsed symptoms of cruelty to animals, running away, and vandalism.

### Treatment Outcome Analysis

As unified groups, both males and females showed significant reductions in rule-breaking, aggression, internalizing, and externalizing scores pre- versus post-SNAP treatment, as measured by the corresponding CBCL subscales (all  $p < .001$ , effect sizes

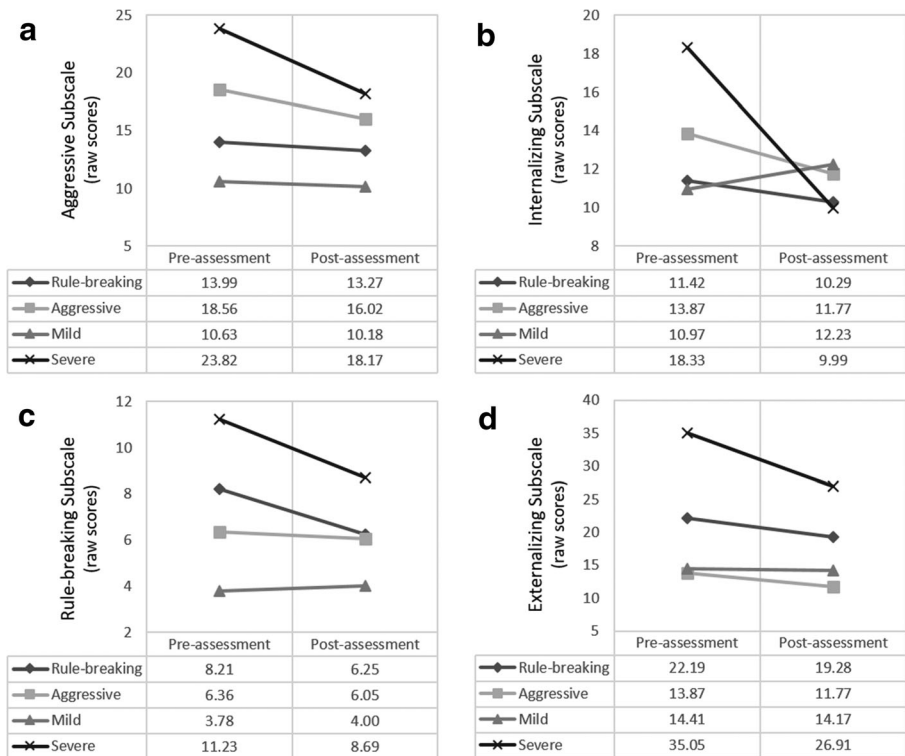
**Table 1** Percentage of symptoms endorsed from the conduct disorder and oppositional defiance disorder items from the Child Behavior Checklist separated by sex and latent classes ( $N = 870$ )

Symptoms	Males $n = 517$				Symptom endorsed (%)	Females $n = 354$		Symptom endorsed (%)
	Rule-breaking ( $n = 78$ )	Aggressive ( $n = 191$ )	Mild ( $n = 134$ )	Severe ( $n = 114$ )		Mild ( $n = 176$ )	Severe ( $n = 177$ )	
Argues	.93	.98	.90	1.00	95.6	.90	1.00	95.2
Cruel animals	.07	.07	.03	.31	11.4	.05	.19	12.1
Cruel bullying	.47	.77	.18	1.00	61.9	.34	.86	6.5
Destroy things	.60	.55	.34	.91	58.4	.36	.82	59.9
Disobedient home	.86	.97	.77	.99	9.5	.83	1.00	91.5
Disobedient school	.78	1.00	.68	.92	86.3	.53	.78	65.8
Lacks guilt	.72	.74	.42	.89	68.7	.44	.83	64.4
Breaks rules	.98	.97	.79	.99	92.8	.78	1.00	89.0
Fights	.44	.86	.26	.90	64.2	.22	.72	48.0
Bad friends	.58	.61	.31	.80	57.1	.23	.52	38.1
Lies cheats	1.00	.73	.45	.94	74.7	.54	.87	7.9
Physically attacks	.34	.82	.38	.97	66.0	.39	.84	61.9
Runs away	.07	.03	.07	.30	1.8	.05	.19	12.4
Sets fire	.16	.00	.03	.16	7.2	.00	.03	1.7
Steals at home	.74	.05	.03	.59	28.0	.14	.47	3.8
Steals outside home	.54	.07	.00	.43	21.1	.11	.37	24.3
Stubborn	.86	.89	.74	.99	86.7	.78	.97	88.1
Swears	.50	.61	.40	.95	61.3	.31	.59	45.2
Temper tantrums	.81	.94	.76	1.00	88.6	.72	.99	86.2
Threatens	.26	.54	.08	.85	44.5	.13	.68	41.2
Truancy	.07	.00	.06	.14	5.6	.03	.10	6.5
Vandalism	.16	.08	.04	.50	17.6	.04	.29	16.9
Number of items m (SD)	12 (2.18)	12 (1.50)	8 (2.06)	17 (1.67)		8 (2.43)	14 (2.02)	

ranging from  $d = .25$  to  $.36$ ). There were no significant sex differences or treatment time-by-sex interactions.

**Treatment Outcome Males** Figure 2 depicts a significant treatment time-by-class interactions for aggression ( $F(3, 371) = 8.87, p < .001$ ), rule-breaking ( $F(3, 371) = 14.01, p < .001$ ), externalizing ( $F(3, 371) = 1.05, p < .001$ ), and internalizing ( $F(3, 371) = 4.09, p < .001$ ). The “aggressive” group showed significant reductions in aggression ( $p < .001, d = .36$ ), externalizing ( $p < .001, d = .34$ ), and internalizing ( $p < .001, d = .23$ ), but not in rule-breaking behavior ( $p < .05, d = .05$ ), while the “rule-breaking” groups showed significant reductions in rule-breaking ( $p < .001, d = .66$ ) and externalizing ( $p < .05, d = .39$ ), but not in aggression ( $p > .05, d = .18$ ) or internalizing ( $p > .05, d < .23$ ). The “mild” group did not show a significant reduction on any measures, but a slight increase in internalizing, albeit with a very small effect size ( $p < .05, d = .14$ ). Finally, the “severe” group showed significant reductions with large effect sizes across all measures (all  $p$ 's  $< .001$ ; rule-breaking ( $d = .74$ ), aggression ( $d = .93$ ), externalizing ( $d = .95$ ), internalizing ( $d = 1$ )).

**Treatment Outcome Females** There were no significant treatment time-by-class interactions for either of the measures for the females; both groups showed reductions in all four measures: the “mild” group (aggressive ( $t(102) = 3.13, p = .02, d = .30$ ), rule-breaking ( $t(102) = 2.41, p = .02, d = .24$ ), externalizing ( $t(102) = 3.90, p < .001, d = .36$ ), and internalizing ( $t(102) = 2.45, p = .03, d = .17$ )) and the “severe” group (aggressive ( $t(113) = 3.22,$



**Fig. 2** Treatment outcomes (pre-post SNAP treatment analysis) for males, separated by the four groups identified in the latent class analysis (rule-breaking, aggressive, mild, and severe). **a** Child Behavior Checklist (CBCL) aggressive subscale, **b** CBCL internalizing subscale, **c** CBCL rule-breaking subscale, and **d** externalizing subscale



$p = .002$ ,  $d = .25$ ), rule-breaking ( $t(113) = 2.68$ ,  $p = .008$ ,  $d = .32$ ), externalizing ( $t(113) = 5.71$ ,  $p < .001$ ,  $d = .43$ ), and internalizing ( $t(113) = 2.78$ ,  $p = .001$ ,  $d = .23$ )).

## Discussion

The aim of this study was to identify subgroups of males and females with conduct problems and investigate whether these subgroups differed in levels of change in aggressive, rule-breaking, externalizing, and internalizing problems following SNAP treatment. We identified four distinct classes of males, which were characterized by (1) “rule-breaking,” (2) “aggressive,” (3) “mild,” and (4) “severe” conduct problems. The four groups differed in treatment outcome, such that the severe group improved on all measures, the mild group on none, the aggressive group specifically on aggressive and externalizing behavior, and the rule-breaking group specifically on rule-breaking and externalizing behavior. We did not observe similar groups for females; instead, the females formed two groups that were not distinguishable by type of behavior but instead were characterized by number of behaviors endorsed (i.e., severity). Consequently, both groups of females improved on all measures, but there were no differences in rate of change between groups.

The clear separation between aggressive and non-aggressive (rule-breaking) behavior classes identified in the male sample conforms with previous classifications of conduct problems (Barker et al. 2007; Burt 2013; Monuteaux et al. 2009; Niv et al. 2013). The identifications of the two classes characterized by “mild” and “severe” conduct problems also lend support to the argument that conduct problems could be conceptualized on a severity dimension (Bezdjian et al. 2011; Krueger et al. 2005). The severe group identified in this study has a higher probability of endorsing the symptoms that are shared with other groups, such as bullying, as well as endorsing additional symptoms that are not present in other groups, such as cruelty to animals and vandalism. The opposite can be seen in the mild group, where both fewer symptoms and lower probability of symptoms can be seen. This demonstrates a range of severity within our sample. Simultaneously, while there is some overlap in symptoms between aggressive and rule-breaking groups, they display distinct features that clearly distinguish the two groups from each other. On this basis, conceptualizing conduct problems as either categorical or dimensional may be too simplistic. Instead, the idea of a complex, multimodal conceptualization of conduct problems in males that considers the type of conduct problems as well as the severity, as previously demonstrated by Bolhuis et al. (2017), may be most appropriate. In this regard, we want to stress that the design and analysis of the current study are not optimally suited to test this theory, as the objective of the study was to compare subgroups, if any were identified. Future research should be conducted that are specifically designed to investigate how to best conceptualize conduct problems for males and females.

In this regard, the findings presented here have several practical implications. Firstly, they highlight how composition of the problem behaviors can aid treatment selection and planning, such that high levels of internalizing problems may be more apparent in those with high levels of aggression, but not necessarily in those with high levels of non-aggressive conduct problems. Secondly, they demonstrate that the composition of

conduct problems can influence treatment outcome, thus averaging over a large group risks biasing or diluting outcome results. This is not surprising from a life-course perspective. It is well known that children with severe conduct problems are likely to have an early onset of problems as well as a higher risk of a long-lived criminal career, relative to those with less severe problems (Moffitt et al. 2001). Similarly, aggressive and non-aggressive antisocial behaviors have at least partly distinct etiologies and divergent trajectories (Barker et al. 2007).

Thirdly, the observed sex differences in the composition of conduct problems likely reflect underlying differences in risk factors and etiology between males and females and support the need for gender-specific treatment programs. Lastly, they demonstrate the effectiveness of the SNAP program to identify and target key behavior problems, especially for treating males with high levels of aggressive and externalizing behaviors. These findings are in line with a randomized controlled trial study that found that males with the most severe behavioral problems showed the greatest improvement following the SNAP treatment, while males with milder problems improved less (Burke and Loeber 2015). In a similar vein, intensive treatment has been found to be less effective for individuals presenting with low relative to high level of risk, and in some cases the treatment may worsen the behavior and outcome of low-risk individuals (Lowenkamp and Latessa 2004). While data on the level of risk was not available for this study, risk and severity of conduct problems are highly associated (Enebrink et al. 2006), and it is feasible that a relatively intensive program such as SNAP may not address the risk and need of low risk children. While this requires more examination, it may be that SNAP is the most beneficial for high-risk males with moderate to severe conduct problems. Indeed, the risk-need-responsivity principle stipulates that the type and intensity of treatment should match the level of risk and individual needs of the child (Bonta and Andrews 2007). As such, there may be more suitable interventions, specifically tailored to males with mild conduct problems. For example, a less intense version of SNAP, such as SNAP irritability (I-SNAP), has been created specifically aimed at reducing oppositional and irritable behavioral problems (Derella et al. 2020). While the key facets of the original SNAP program remain, the I-SNAP program targets less severe behavioral problems with the specific focus on increasing frustration tolerance by means of emotion-focused coping strategies and relaxation.

There are several probable reasons as to why we did not observe similar categories for males and females. Methodologically, the differences in classifications are unlikely to be due to uneven sample sizes or a lack of variability of symptom presentation at pre-assessment. Both males and females presented with high numbers of aggressive and externalizing symptoms at pre-assessment and the number of participants were sufficiently high to reliably identify a number of classes in both groups (Wurpts and Geiser 2014). Several studies have suggested that emotional closeness and quality of the caregiver-child relationship may play a greater role in females' internalizing and delinquent behavior relative to males (Hart et al. 2007; Lewis et al. 2015; Tisak et al. 2017). In this regard, it may be important to include additional variables such as relationship quality or internalizing symptoms when analyzing female behavior, rather than relying on conduct behaviors alone. Another explanation for the observed differences relate to the use of categorical models of conduct problems. The two female classes identified from the LCA were only distinguishable based on number of symptoms endorsed, suggesting that severity may be more informative than clusters for

females, over and above what might be the case for males. Furthermore, as we did not have risk assessment data available for this study, we were unable to explore if, and to what extent, the level of risk might interact with behavior and influences treatment outcome in different ways for males and females. The study has several important strengths; we used a self-referred community sample, adopted a data-driven approach to categorize our sample rather than pre-defined, inflexible categories, and ran separate models for males and females to avoid averaging out important effects of sex. However, the results need to be interpreted with a number of limitations in mind. Firstly, by focusing on the CBCL subscales as the outcome measure, we excluded a number of important potential treatment outcomes such as increased problem-solving skills, self-control, or the caregiver-child relationship. While these factors are interesting and worthy of investigation, increasing the number of variables under investigation would have decreased power and added complexity to the interpretation of results. Secondly, this study was an analysis of archive data, and a control group was not available. While utilizing archive data allowed us to include a large group of representative participants, whose data were not confined to a brief time period of data collection, we were not able to manipulate or chose any of the variables ahead of the study, such as additional standardized measures of aggression and rule-breaking, which limited our choices of analysis and thus interpretation. In addition, there was no control group available, which curbs interpretation of the intervention outcome in its own right. A control group without any conduct problems would not lend interpretation to how the conduct problem groups differed from each other, which was a key research question in this study. In this regard, using the pre-treatment scores to control for the post-treatment scores within the same groups allowed for more control over extraneous variables and less risk of introducing noise to the data based on differences between individuals. A control group with similar levels of conduct problems who underwent alternative treatment would strengthen the interpretation of SNAP in general. Several such studies, using randomized controlled trials and wait-list designs, have been published previously (see Augimeri et al. 2007, 2018; Burke and Loeber 2015, 2016; Pepler et al. 2010). However, the evaluation of the SNAP program was not the primary aim of the current study.

Thirdly, it is possible that caregiver ratings of child behavior post-treatment were influenced by their own participation in the parent component of SNAP; if parenting skills and awareness of their child's behavior increased as a consequence of participating in the program, it is possible that the increased skills could influence their rating. Since parent data was not available for the current study, the possible effect of participating in the parent group remains speculative, but future prospective studies should keep this issue in mind and consider alternative assessments from teachers or other close adults.

Lastly, the authors want to stress the point that the classes identified in the current study, by definition, were dependent on the sample of children with conduct problems whose caregivers sought SNAP treatment, and classifications may not apply to other populations of children with conduct problems. Equally, the effects of treatment may be specific to the SNAP program and do not necessarily reflect outcome of other CBT-based treatment programs. Because of the heterogeneity and complexity of conduct problems, it would benefit future studies to conduct a data-driven analysis of their sample prior to investigating treatment outcomes for groups to account for the heterogeneity of the sample.

This lack of synergy between the sexes in our results suggests that while males and females with conduct problems may display similar types and frequency of problem

behaviors, the compositions of these behaviors are different and, furthermore, may need to be conceptualized separately. Large-scale studies with the power to include numerous additional variables would be valuable in order to fully investigate the conceptualization of conduct problems in females, and how it relates to treatment selection, planning, and outcome.

**Conclusion** Despite the fact that the males and females displayed similar types of conduct problems, the composition of these behaviors differed. The distinct groups of conduct problems seen in males could not be found in females. However, a dimensional aspect to conduct problems, based on number and types of symptoms, could be seen in both sexes. This clearly warrants further consideration. Future studies must be mindful of sex differences when investigating classifications of conduct problems and particular investigations of the conceptualization of female conduct problems are warranted. Furthermore, the composition, as well as the severity, of conduct problems should be taken into account when selecting and planning for treatment to ensure that the treatment is targeted specifically for the child's needs. Thus, the targeted treatment is not only necessary based on the gender of the child but also based on the unique combination of problem behaviors shown by individuals.

**Acknowledgments** The authors are grateful to all the families who participated in the SNAP program and made this research possible.

**Authors' Contributions** All authors contributed to the study conception and design. Material preparation and data collection were overseen by Andrea Blackman, Margaret Walsh, and Leena Augimeri, and analysis were performed by Areti Smaragdi, Adam Donato, and Andrea Blackman. The first draft of the manuscript was written by Areti Smaragdi, and all authors commented on the previous versions of the manuscript. All authors read and approved the final manuscript.

## Compliance with Ethical Standards

**Conflict of Interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles: child behavior checklist for ages 6-18, teacher's report form, youth self-report: an integrated system of multi-informant assessment*. University of Vermont, research center for children youth & families.
- Ackermann, K., Kirchner, M., Bernhard, A. et al. (2019) Relational aggression in adolescents with conduct disorder: Sex differences and behavioral correlates. *Journal of Abnormal Child Psychology*, 47, 1625–1637. <https://doi.org/10.1007/s10802-019-00541-6>.

- Augimeri, L. K., Farrington, D. P., Koegl, C. J., & Day, D. M. (2007). The SNAP™ under 12 outreach project: effects of a community based program for children with conduct problems. *Journal of Child and Family Studies*, 16(6), 799–807. <https://doi.org/10.1007/s10826-006-9126-x>.
- Augimeri, L. K., Walsh, M., Donato, A., Blackman, A., & Piquero, A. R. (2018). SNAP (stop now and plan): helping children improve their self-control and externalizing behavior problems. *Journal of Criminal Justice*, 56, 43–49. <https://doi.org/10.1016/j.jcrimjus.2017.08.010>.
- Augimeri, L. K., Walsh, M., Kivlenieks, M., & Pepler, D. (2017). Addressing children's disruptive behavior problems: a 30-year journey with Stop Now And Plan (SNAP). In P. Sturmey (Ed.), *The Wiley Handbook of Violence and Aggression* (pp. 1–13). <https://doi.org/10.1002/9781119057574.whbva028>.
- Augimeri, L., Walsh, M., Woods, S., & Jiang, D. (2012). Risk assessment and clinical risk management for young antisocial children: the forgotten group. *Universitas Psychologica*, 11(4), 1147–1156.
- Baker, K. (2013). Conduct disorders in children and adolescents. *Paediatrics and Child Health*, 23(1), 24–29. <https://doi.org/10.1016/j.paed.2012.09.007>.
- Barker, E. D., Séguin, J. R., White, H. R., Bates, M. E., Lacourse, É., Carbonneau, R., & Tremblay, R. E. (2007). Developmental trajectories of male physical violence and theft: relations to neurocognitive performance. *Archives of General Psychiatry*, 64(5), 592–599. <https://doi.org/10.1001/archpsyc.64.5.592>.
- Bezdzjian, S., Krueger, R. F., Derringer, J., Malone, S., McGue, M., & Iacono, W. G. (2011). The structure of DSM-IV ADHD, ODD, and CD criteria in adolescent boys: a hierarchical approach. *Psychiatry Research*, 188(3), 411–421. <https://doi.org/10.1016/j.psychres.2011.02.027>.
- Bolhuis, K., Lubke, G. H., van der Ende, J., Bartels, M., van Beijsterveldt, C. E. M., Lichtenstein, P., Larsson, H., Jaddoe, V. W. V., Kushner, S. A., Verhulst, F. C., Boomsma, D. I., & Tiemeier, H. (2017). Disentangling heterogeneity of childhood disruptive behavior problems into dimensions and subgroups. *Journal of the American Academy of Child & Adolescent Psychiatry*, 56(8), 678–686. <https://doi.org/10.1016/j.jaac.2017.05.019>.
- Bonta, J., & Andrews, D. A. (2007). Risk-need-responsivity model for offender assessment and rehabilitation. *Rehabilitation*, 6(1), 1–22.
- Breslau, J., Saito, N., Tancredi, D. J., Nock, M., & Gilman, S. E. (2012). Classes of conduct disorder symptoms and their life course correlates in a US national sample. *Psychological Medicine*, 42(5), 1081–1089. <https://doi.org/10.1017/S003329171100198X>.
- Burke, J. D., & Loeber, R. (2015). The effectiveness of the Stop Now and Plan (SNAP) Program for boys at risk for violence and delinquency. *Prevention Science*, 16(2), 242–253. <https://doi.org/10.1007/s11121-014-0490-2>.
- Burke, J. D., & Loeber, R. (2016). Mechanisms of behavioral and affective treatment outcomes in a cognitive behavioral intervention for boys. *Journal of Abnormal Child Psychology*, 44(1), 179–189. <https://doi.org/10.1007/s10802-015-9975-0>.
- Burt, S. A. (2013). Do etiological influences on aggression overlap with those on rule breaking? A meta-analysis. *Psychological Medicine*, 43(9), 1801–1812. <https://doi.org/10.1017/S0033291712001894>.
- Burt, S. A. (2009). Are there meaningful etiological differences within antisocial behavior? Results of a meta-analysis. *Clinical Psychology Review*, 29(2), 163–178. <https://doi.org/10.1016/j.cpr.2008.12.004>.
- Burt, S. A., & Donnellan, M. B. (2008). Personality correlates of aggressive and non-aggressive antisocial behavior. *Personality and Individual Differences*, 44(1), 53–63. <https://doi.org/10.1016/j.paid.2007.07.022>.
- Coghill, D., & Sonuga-Barke, E. J. S. (2012). Annual research review: categories versus dimensions in the classification and conceptualisation of child and adolescent mental disorders—implications of recent empirical study. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 53(5), 469–489. <https://doi.org/10.1111/j.1469-7610.2011.02511.x>.
- Colman, I., Murray, J., Abbott, R. A., Maughan, B., Kuh, D., Croudace, T. J., & Jones, P. B. (2009). Outcomes of conduct problems in adolescence: 40 year follow-up of national cohort. *BMJ*, 338, a2981. <https://doi.org/10.1136/bmj.a2981>.
- Davé, S., Nazareth, I., Senior, R., & Sherr, L. (2008). A comparison of father and mother report of child behaviour on the strengths and difficulties questionnaire. *Child Psychiatry and Human Development*, 39(4), 399–413. <https://doi.org/10.1007/s10578-008-0097-6>.
- Derella, O. J., Burke, J. D., Romano-Verthelyi, A. M., Butler, E. J., & Johnston, O. G. (2020). Feasibility and acceptability of a brief cognitive-behavioral group intervention for chronic irritability in youth. *Clinical Child Psychology and Psychiatry*, 1359104520918331.
- Enebrink, P., Långström, N., Hultén, A., & Gumpert, C. H. (2006). Swedish validation of the Early Assessment Risk List for Boys (EARL-20B), a decision aid for use with children presenting with conduct-disordered behaviour. *Nordic Journal of Psychiatry*, 60(6), 438–446.

- Epstein, R. A., Fonnesebeck, C., Potter, S., Rizzone, K. H., & McPheeters, M. (2015). Psychosocial interventions for child disruptive behaviors: a meta-analysis. *Pediatrics*, *136*(5), 947–960. <https://doi.org/10.1542/peds.2015-2577>.
- Farrington, D. P., & Koegl, C. J. (2015). Monetary benefits and costs of the stop now and plan program for boys aged 6–11, based on the prevention of later offending. *Journal of Quantitative Criminology*, *31*(2), 263–287. <https://doi.org/10.1007/s10940-014-9240-7>.
- Forgatch, M. S., & Patterson, G. R. (2010). Parent Management Training—Oregon Model: an intervention for antisocial behavior in children and adolescents. In J. R. Weisz & A. E. Kazdin (Eds.), *Evidence-based psychotherapies for children and adolescents* (2nd ed., pp. 159–177). New York: The Guilford Press.
- Frick, P. J., & White, S. F. (2008). Research review: the importance of callous-unemotional traits for developmental models of aggressive and antisocial behavior. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *49*(4), 359–375. <https://doi.org/10.1111/j.1469-7610.2007.01862.x>.
- González-Madruga, K., Rogers, J., Toschi, N., Riccelli, R., Smaragdi, A., Puzzo, I., et al. (2019). White matter microstructure of the extended limbic system in male and female youth with conduct disorder. *Psychological Medicine*, 1–10. <https://doi.org/10.1017/S0033291718003951>.
- Hagenaars, J. A., & McCutcheon, A. L. (2002). *Applied latent class analysis*. Cambridge, United Kingdom: Cambridge University Press.
- Hart, J. L., O'Toole, S. K., Price-Sharps, J. L., & Shaffer, T. W. (2007). The risk and protective factors of violent juvenile offending: an examination of gender differences. *Youth Violence and Juvenile Justice*, *5*(4), 367–384. <https://doi.org/10.1177/1541204006297367>.
- Howell, J. C., Lipsey, M. W., & Wilson, J. J. (2014). *A handbook for evidence-based juvenile justice systems*. London, United Kingdom: Lexington Books.
- Jiang, D., Walsh, M., & Augimeri, L. K. (2011). The linkage between childhood bullying behaviour and future offending. *Criminal behaviour and mental health: CBMH*, *21*(2), 128–135. <https://doi.org/10.1002/cbm.803>.
- Kazdin, A. E., Siegel, T. C., & Bass, D. (1992). Cognitive problem-solving skills training and parent management training in the treatment of antisocial behavior in children. *Journal of Consulting and Clinical Psychology*, *60*(5), 733–747.
- Krueger, R. F., Markon, K. E., Patrick, C. J., & Iacono, W. G. (2005). Externalizing psychopathology in adulthood: a dimensional-spectrum conceptualization and its implications for DSM-V. *Journal of Abnormal Psychology*, *114*(4), 537–550. <https://doi.org/10.1037/0021-843X.114.4.537>.
- Lacourse, E., Baillargeon, R., Dupéré, V., Vitaro, F., Romano, E., & Tremblay, R. (2010). Two-year predictive validity of conduct disorder subtypes in early adolescence: a latent class analysis of a Canadian longitudinal sample. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *51*(12), 1386–1394. <https://doi.org/10.1111/j.1469-7610.2010.02291.x>.
- Lahey, B. B., & Waldman, I. D. (2012). Phenotypic and causal structure of conduct disorder in the broader context of prevalent forms of psychopathology. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *53*(5), 536–557. <https://doi.org/10.1111/j.1469-7610.2011.02509.x>.
- Lewis, A. J., Kremer, P., Douglas, K., Toumborou, J. W., Hameed, M. A., Patton, G. C., & Williams, J. (2015). Gender differences in adolescent depression: differential female susceptibility to stressors affecting family functioning. *Australian Journal of Psychology*, *67*(3), 131–139. <https://doi.org/10.1111/ajpy.12086>.
- Lewis, M. D., Granic, I., Lamm, C., Zelazo, P. D., Stieben, J., Todd, R. M., Moadab, I., & Pepler, D. (2008). Changes in the neural bases of emotion regulation associated with clinical improvement in children with behavior problems. *Development and Psychopathology*, *20*(3), 913–939. <https://doi.org/10.1017/S0954579408000448>.
- Lipsey, M. W., Landenberger, N. A., & Wilson, S. J. (2007). Effects of cognitive-behavioral programs for criminal offenders. *Campbell Systematic Reviews*, *6*(1), 27.
- Lowenkamp, C. T., & Latessa, E. J. (2004). Understanding the risk principle: how and why correctional interventions can harm low-risk offenders. *Topics in community corrections*, *2004*, 3–8.
- Maughan, B., Rowe, R., Messer, J., Goodman, R., & Meltzer, H. (2004). Conduct disorder and oppositional defiant disorder in a national sample: developmental epidemiology. *Journal of Child Psychology and Psychiatry*, *45*(3), 609–621. <https://doi.org/10.1111/j.1469-7610.2004.00250.x>.
- Moffitt, T. E., & Caspi, A. (2001). Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. *Development and Psychopathology*, *13*(2), 355–375.
- Moffitt, T. E., Caspi, A., Rutter, M., & Silva, P. A. (2001). *Sex differences in antisocial behaviour: conduct disorder, delinquency, and violence in the Dunedin Longitudinal Study*. New York: Cambridge University Press. <https://doi.org/10.1017/CBO9780511490057>.

- Monuteaux, M. C., Biederman, J., Doyle, A. E., Mick, E., & Faraone, S. V. (2009). Genetic risk for conduct disorder symptom subtypes in an ADHD sample: specificity to aggressive symptoms. *Journal of the American Academy of Child & Adolescent Psychiatry*, 48(7), 757–764. <https://doi.org/10.1097/CHI.0b013e3181a5661b>.
- Muthén, L. K., & Muthén, B. O. (2005). *Mplus: statistical analysis with latent variables: User's guide*. Los Angeles: Muthén & Muthén.
- Niv, S., Tuvblad, C., Raine, A., & Baker, L. A. (2013). Aggression and rule-breaking: heritability and stability of antisocial behavior problems in childhood and adolescence. *Journal of Criminal Justice*, 41(5), 285–291. <https://doi.org/10.1016/j.jcrimjus.2013.06.014>.
- Nock, M. K., Kazdin, A. E., Hiripi, E., & Kessler, R. C. (2006). Prevalence, subtypes, and correlates of DSM-IV conduct disorder in the national comorbidity survey replication. *Psychological Medicine*, 36(5), 699–710. <https://doi.org/10.1017/S0033291706007082>.
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(4), 535–569. <https://doi.org/10.1080/10705510701575396>.
- Pepler, D., Walsh, M., Yuile, A., Levene, K., Jiang, D., Vaughan, A., & Webber, J. (2010). Bridging the gender gap: interventions with aggressive girls and their parents. *Prevention Science*, 11(3), 229–238. <https://doi.org/10.1007/s11121-009-0167-4>.
- Piquero, A. R., Jennings, W. G., Diamond, B., Farrington, D. P., Tremblay, R. E., Welsh, B. C., & Gonzalez, J. M. R. (2016). A meta-analysis update on the effects of early family/parent training programs on antisocial behavior and delinquency. *Journal of Experimental Criminology*, 12(2), 229–248. <https://doi.org/10.1007/s11292-016-9256-0>.
- Sidlauskaite, J., González-Madruga, K., Smaragdi, A., Riccelli, R., Puzzo, I., Batchelor, M., Cornwell, H., Clark, L., Sonuga-Barke, E. J. S., & Fairchild, G. (2018). Sex differences in risk-based decision making in adolescents with conduct disorder. *European Child & Adolescent Psychiatry*, 27(9), 1133–1142. <https://doi.org/10.1007/s00787-017-1024-9>.
- Smaragdi, A., Cornwell, H., Toschi, N., Riccelli, R., Gonzalez-Madruga, K., Wells, A., Clanton, R., Baker, R., Rogers, J., Martin-Key, N., Puzzo, I., Batchelor, M., Sidlauskaite, J., Bernhard, A., Martinelli, A., Kohls, G., Konrad, K., Baumann, S., Raschle, N., Stadler, C., Freitag, C., Sonuga-Barke, E. J. S., de Brito, S., & Fairchild, G. (2017). Sex differences in the relationship between conduct disorder and cortical structure in adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, 56(8), 703–712. <https://doi.org/10.1016/j.jaac.2017.05.015>.
- Tisak, J., Tisak, M. S., Baker, E. R., Amrhein, K. E., & Jensen, C. (2017). The association among parental bonding, depression, social aggression, and criminal assault: are there gender differences between male and female youth offenders? *Journal of Interpersonal Violence*, 886260517744192, 088626051774419. <https://doi.org/10.1177/0886260517744192>.
- Van Hulle, C. A., Waldman, I., & Lahey, B. B. (2018). Sex differences in the genetic and environmental influences on self-reported non-aggressive and aggressive conduct disorder symptoms in early and middle adolescence. *Behavior Genetics*, 48(4), 271–282. <https://doi.org/10.1007/s10519-018-9907-1>.
- Walton, K. E., Ormel, J., & Krueger, R. F. (2011). The dimensional nature of externalizing behaviors in adolescence: evidence from a direct comparison of categorical, dimensional, and hybrid models. *Journal of Abnormal Child Psychology*, 39(4), 553–561. <https://doi.org/10.1007/s10802-010-9478-y>.
- Woltering, S., Liao, V., Liu, Z.-X., & Granic, I. (2015). Neural rhythms of change: long-term improvement after successful treatment in children with disruptive behavior problems. *Neural Plasticity*. <https://doi.org/10.1155/2015/873197>.
- Wurpts, I. C., & Geiser, C. (2014). Is adding more indicators to a latent class analysis beneficial or detrimental? Results of a Monte-Carlo study. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.00920>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Affiliations

**Areti Smaragdi<sup>1</sup> · Andrea Blackman<sup>1</sup> · Adam Donato<sup>1,2</sup> · Margaret Walsh<sup>1</sup> ·  
Leena Augimeri<sup>1,2</sup>**

<sup>1</sup> Child Development Institute, 46 St. Clair Gardens, Toronto, ON, Canada

<sup>2</sup> Ontario Institute for Studies in Education, University of Toronto, Toronto, ON, Canada