

**Latent Oppositional Defiant Disorder Symptom
Classes: Longitudinal Evidence for Severity-Based
Distinctions in a High-Risk Sample**

by
Maeve Cyr

M.A. (Clinical Psychology), Simon Fraser University, 2018

B.A. (Hons., Psychology), Simon Fraser University, 2014

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Psychology
Faculty of Arts and Social Sciences

© Maeve Cyr 2022
SIMON FRASER UNIVERSITY
Spring 2022

Declaration of Committee

Name: Maeve Cyr
Degree: Doctor of Philosophy (Clinical Psychology)
Title: Latent Oppositional Defiant Disorder Symptom
Classes: Longitudinal Evidence for Severity-
Based Distinctions in a High-Risk Sample

Committee: **Chair: V. Gordon Rose**
Senior Lecturer, Psychology

Robert McMahon
Supervisor
Professor, Psychology

Marlene Moretti
Committee Member
Professor, Psychology

Jodi Viljoen
Committee Member
Professor, Psychology

Raymond Corrado
Examiner
Professor, Criminology

Jeffrey Burke
External Examiner
Associate Professor, Psychological Sciences
University of Connecticut

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

- a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

- b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

- c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

Update Spring 2016

Abstract

Oppositional defiant disorder (ODD) is characterized by disobedience, irritability, and hostility directed toward authority figures. Evidence suggests that there is an important distinction between the behavioural and affective symptoms of the disorder; however, it is currently unclear whether there are distinct subtypes of ODD. Using data from a high-risk, longitudinal sample from the Fast Track Project ($n = 446$), latent class analysis was used to examine latent classes of youth based on parent-reported ODD symptom criteria, separately at four different time points (grades 3, 6, 9, and 12). Three-class solutions were supported in all grades, with latent classes representing youth with *Low* (69.55-78.17% across grades), *Moderate* (14.52-23.24%), and *High* (3.53-9.03%) probabilities of parent-reported ODD symptoms. Tests of measurement invariance revealed some differences in the structure of latent classes across certain time points. Demographic variables, including race and initial levels of risk for conduct problems, significantly predicted latent class membership. The findings do not support the existence of ODD subtypes and suggest that symptom severity may be more important for distinguishing youth with ODD symptoms. More person-centered research is required to understand how the disorder presents across development.

Acknowledgements

This dissertation was supported by the Canadian Institutes of Health Research (CIHR) Frederick Banting and Charles Best Canada Graduate Scholarship – Doctoral as well as the BC Children’s Hospital Research Institute (BCCHRI) Brain, Behaviour, & Development Theme Trainee Boost Award.

I would like to acknowledge the Principal Investigators of the Fast Track Project, the Conduct Problems Prevention Research Group. In alphabetical order, they include Drs. Karen Bierman, John Coie, Max Crowley, Kenneth Dodge, Mark Greenberg, John Lochman, Robert McMahon, and Ellen Pinderhughes.

I especially want to thank my research supervisor, Dr. Robert McMahon, for his guidance and encouragement throughout my Honours, Master’s, and now Doctorate degrees. I could not have asked for a more knowledgeable, supportive, and available mentor.

I would also like to thank my dissertation committee members, Drs. Marlene Moretti and Jodi Viljoen, for their helpful feedback on this research project, starting with the proposal and continuing to the defence. In addition, I want to express my appreciation of the examiners, Drs. Raymond Corrado and Jeffrey Burke, for their thoughtful comments and questions during my dissertation defence.

This dissertation was made possible by consultation with Dr. Sarah Racz, whose own research on oppositional defiant disorder subtypes inspired this longitudinal project. Along with Dr. Racz, Dr. Natalie Goulter also provided invaluable guidance related to statistical analyses. I am very grateful for their time and assistance.

Table of Contents

Declaration of Committee	ii
Ethics Statement	iii
Abstract	iv
Acknowledgements	v
Table of Contents	vi
List of Tables	vii
List of Figures	vii
Introduction	1
Are There Meaningful Dimensions of ODD Symptoms?	2
Do Subtypes of ODD Exist?	5
The Present Study	11
Method	13
Participants	13
Measures	14
Missing Data	15
Analysis Plan	16
Results	19
Step 1: Descriptive Statistics	19
Step 2: Unconditional Latent Class Models of ODD Criteria Across Grades	19
Step 3: Measurement Invariance	28
Grades 3 and 6	28
Grades 6 and 9	29
Grades 9 and 12	29
Step 4: Covariates	29
Grade 3	30
Grade 6	30
Grade 9	30
Grade 12	31
Discussion	33
Can Latent ODD Symptom Classes Be Identified in Grades 3, 6, 9, and 12?	33
Are the Same ODD Symptom Subgroups Identified Across Time?	35
What Predicts Subgroup Membership?	37
Strengths and Limitations	38
Conclusions and Future Directions	40
References	42

List of Tables

Table 1.	Frequencies of ODD symptom criteria across grades	20
Table 2.	Descriptives	22
Table 3.	Bivariate correlations between demographic characteristics, ODD diagnoses, and ODD criterion counts	23
Table 4.	Summary of information for selecting the number of latent classes in grades 3, 6, 9, and 12	24
Table 5.	Prevalence rates and item-response probabilities for unconditional 3-class models	26
Table 6.	Prevalence rates and item-response probabilities for unconditional 3-class models in grade 3 with different scoring procedures for spiteful and vindictive	27
Table 7.	Prevalence rates and item-response probabilities for conditional 3-class models, including race and total screen scores	32

List of Figures

Figure 1.	Unconditional 3-class solutions in grades 3, 6, 9, and 12	27
-----------	---	----

Introduction

Oppositional defiant disorder (ODD) is a form of childhood psychopathology characterized by disobedience, irritability, and hostility directed towards authority figures (American Psychiatric Association [APA], 2013). In the Diagnostic and Statistical Manual of Mental Disorders (DSM), ODD is listed alongside conduct disorder (CD) under Disruptive, Impulse-Control, and Conduct Disorders. It is defined as the presence of at least four out of the following eight symptoms for a period of 6 months or more: often loses temper; often touchy or easily annoyed; often angry and resentful; often argues with authority figures; often actively defies or refuses to comply with requests from authority figures or rules; often deliberately annoys others; often blames others for his or her mistakes or misbehaviour; and spitefulness or vindictiveness (APA, 2013). Affecting approximately 3.3% of children and youth across cultures (Canino et al., 2010), ODD is one of the most common reasons young people are referred for mental health services (Kimonis et al., 2014). Nevertheless, as highlighted by Frick and Nigg (2012), ODD has long been a controversial diagnostic construct, in part due to its high rates of comorbidity with other disorders (e.g., CD, attention-deficit/hyperactivity disorder [ADHD], anxiety disorders, depression, and substance use disorders; APA, 2013).

Despite the controversy, there is compelling evidence that ODD is a meaningful diagnosis that independently predicts a host of adverse functional and mental health outcomes (Burke et al., 2018; Frick & Nigg, 2012). For example, in a clinic-referred sample of 177 boys, Burke, Rowe, and Boylan (2014) found that, controlling for symptoms of CD, ADHD, depression, and anxiety, ODD symptoms in adolescence were associated with more social problems (i.e., with peers, romantic partners, and fathers) and greater difficulty finding an occupational reference at age 24. In terms of mental health, in a retrospective study of 3,199 adults, Nock et al. (2007) found that 92% of individuals with lifetime ODD had also met criteria for another mental disorder (e.g., mood, anxiety, impulse-control, and substance use disorders), with ODD having an earlier age of onset in most cases. Indeed, similar to other disruptive behaviour disorders, ODD predicts externalizing problems (e.g., later ODD, CD, involvement in justice system; e.g., Burke et al., 2005; Pardini & Fite, 2010). However, ODD also has a strong relationship with internalizing psychopathology (see Boylan et al., 2007; and Burke & Loeber, 2010; for reviews). For example, in a population-based sample of 1,420

children, Copeland et al. (2009) found that childhood ODD was the only significant predictor of adult depression when comorbidities (i.e., childhood depression, anxiety disorders, CD, ADHD) were included in the same model. Other studies have since replicated this result, demonstrating that ODD is more predictive of later depression and anxiety than CD (e.g., Pardini & Fite, 2010; Rowe et al., 2010). Research on the dimensionality of ODD symptoms has shed some light on the wide-ranging sequelae of the disorder.

Are There Meaningful Dimensions of ODD Symptoms?

Burke et al. (2005) were among the first to formally point out the qualitative differences between ODD symptoms reflecting affect versus defiant behaviours/attitudes, hypothesizing that these symptom clusters could account for the link between ODD and diverse forms of psychopathology. Since that time, a number of studies have examined the dimensionality of ODD symptoms (e.g., Aebi et al, 2010; Stringaris & Goodman, 2009a; Burke et al., 2010; Rowe et al., 2010). On theoretical grounds, Stringaris and Goodman (2009a) proposed a three-factor model of ODD symptoms, including *Irritable* (loses temper, angry, touchy), *Headstrong* (argues, defies, annoys others, blames others), and *Hurtful* (spiteful and vindictive) dimensions. They found that the Irritable dimension was the only dimension associated with concurrent emotional problems, while the Headstrong and Hurtful dimensions were the most strongly associated with measures of ADHD and cold-bloodedness/callousness, respectively. All three dimensions were associated with concurrent conduct problems and CD diagnoses in this study; however, the Hurtful dimension was most strongly associated with aggressive CD symptoms, while the Headstrong dimension showed the strongest association with non-aggressive CD symptoms. Longitudinally, Stringaris and Goodman (2009b) found the same pattern of associations between Irritable-emotional problems and Headstrong-ADHD. At 3-year follow up, the Headstrong dimension was the most robust predictor of all types of conduct problems, and the Hurtful dimension specifically predicted aggressive symptoms.

In response to this research, the APA (2013) formally acknowledged ODD symptom dimensions by incorporating Stringaris and Goodman's (2009a) model into the DSM-5 definition of ODD. While this certainly reflects the growing consensus in the field about the multidimensional nature of ODD (Burke et al., 2018; Frick & Nigg, 2012),

evidence for slightly different models has emerged from factor analytic studies. Although some studies have found support for the Irritable dimension specified by Stringaris and Goodman (i.e., loses temper, angry, touchy; e.g., Aebi et al., 2010; Burke, Boylan, et al., 2014; Krieger et al., 2013; Rowe et al., 2010), they nevertheless provide inconsistent results regarding the five remaining ODD symptoms. For example, using exploratory factor analysis in a representative community sample of 11- and 13-year-olds, Rowe et al. (2010) reported that the remaining ODD symptoms loaded onto a single behavioural dimension (i.e., combining Headstrong and Hurtful ODD symptoms). These researchers found that the two dimensions did not significantly differ in their prediction of ODD, CD, or depressive disorder diagnoses up to age 16. That said, the *Headstrong+Hurtful* dimension was uniquely associated with later substance disorder, and the Irritable dimension independently predicted anxiety disorders.

Stringaris and Goodman's (2009a) Irritable dimension has been contested as well. Using exploratory factor analyses in a sample of preadolescent girls, Burke et al. (2010) identified a different three-factor model characterized by *Negative Affect* (touchy, angry, spiteful), *Oppositional Behavior* (loses temper, argues, defies), and *Antagonistic Behavior* (annoys others, blames others). They found that Negative Affect predicted later depression, whereas Oppositional Behavior and Antagonistic Behavior predicted later CD overall, with some significant differences by race. These findings are generally consistent with the studies described above, but nevertheless raise questions about the exact operationalization of ODD symptom dimensions.

To help resolve these discrepancies, several researchers have directly compared competing models using confirmatory factor analysis (e.g., Burke, Boylan, et al., 2014; Ezpeleta et al., 2012; Gomez & Stavropoulos, 2018; Krieger et al., 2013; Lavigne et al., 2015). In a community sample of Brazilian children, Krieger et al. (2013) reported that Stringaris and Goodman's (2009a) 3-factor model (i.e., Irritable-Headstrong-Hurtful) best fit the data. In contrast, considering multiple measures and informants in a large community sample of Spanish preschoolers, Ezpeleta et al. (2012) reported that the other 2-factor (Irritable-Headstrong; Rowe et al., 2010) and 3-factor (Negative Affect-Oppositional Behavior-Antagonistic Behavior; Burke et al., 2010) models were both supported by the data. In the most comprehensive study to date, Burke, Boylan, et al. (2014) investigated the structure of ODD in five different datasets. Across all five samples consisting of American and British children, they found that the structure of

ODD was best characterized by a bifactor model, which included a general ODD factor and two correlated symptom dimensions (Irritability and Oppositional Behavior). They also tested the two competing models of affective ODD symptoms (i.e., Burke et al., 2010; Stringaris & Goodman, 2009a), finding that items touchy, angry, and temper best characterized this dimension in four out of the five datasets. That said, Gomez and Stavropoulos (2018) recently reported that this bifactor model had the least empirical support of the six factor analytic models that they examined in a representative Australian community sample.

Thus, there is lingering debate about the particular loadings of ODD symptoms as well as the most appropriate labels for symptom categories. Nevertheless, across samples, and using different measurement instruments, at least two distinct but correlated dimensions of ODD symptoms have been identified in factor analyses (e.g., Aebi et al., 2010; Burke et al., 2010; Burke, Boylan, et al., 2014; Ezpeleta et al., 2012; Krieger et al., 2013; Rowe et al., 2010). There is now considerable evidence that a clinically meaningful distinction exists between the affective and behavioural symptoms of ODD (Burke et al., 2018; Frick & Nigg, 2012). ODD symptoms related to irritable mood/negative affect tend to predict internalizing problems, whereas ODD symptoms related to defiant behaviour tend to be more strongly associated with externalizing problems (e.g., Aebi et al., 2013; Burke et al., 2010; Burke, 2012; Ezpeleta et al., 2012; Herzoff & Tackett, 2016; Stringaris & Goodman, 2009ab; Whelan et al., 2013). In sum, current evidence indicates that these ODD symptom dimensions may signify different prognoses and treatment needs for young people with the disorder (Frick & Nigg, 2012; Stringaris & Goodman, 2009a). The evidence base suggests that there may be meaningful subtypes of ODD, yet the variable-centered methods used in most existing studies have been ill-suited to address this empirical question. Some researchers have attempted to manually assign participants to groups based on a priori operational criteria for possible ODD subtypes, with some promising results (e.g., Drabick & Gadow, 2012). However, this approach involves making assumptions about what ODD subtypes look like (e.g., number of subtypes, characteristic symptoms). To directly test whether distinct subgroups of youth do, in fact, emerge based on the ODD symptoms they present with, researchers have turned to person-centered data analytic methods.

Do Subtypes of ODD Exist?

Variable-centered statistical approaches, such as correlation and regression, are designed to describe associations between variables (e.g., correlations between ODD symptoms; Laursen & Hoff, 2006, p. 379). They operate under the assumption that these associations are similar across individuals within the population, with the aim of identifying common processes (e.g., that the symptoms loses temper, angry, and touchy tend to correlate highly; e.g., Burke, Boylan, et al., 2014; see Laursen & Hoff, 2006). In this sense, variable-centered approaches focus on aggregate data, and thus, are not designed to capture variability across individuals or groups. In contrast, person-centered statistical methods (e.g., latent profile/class/cluster analyses) operate under the assumption that there is heterogeneity within the population (Laursen & Hoff, 2006, p. 379). They are specifically designed to characterize this heterogeneity, typically by utilizing individual-level data to categorize individuals who are similar on a construct of interest (e.g., ODD symptoms; Collins & Lanza, 2010; Laursen & Hoff, 2006). Therefore, while variable-centered methods (e.g., factor analysis) have played an important role in identifying symptom dimensions of ODD, person-centered data analytic procedures are needed to examine whether the disorder has distinct subtypes.

A small number of studies have used group-based trajectory modeling with ODD symptom dimensions to explore subtypes (e.g., Boylan et al., 2017; Ezpeleta et al., 2021). Modelling parallel trajectories of Burke et al.'s (2010) Negative Affect-Oppositional-Antagonistic dimensions from ages 5 to 13 in the Pittsburg Girls Study, Boylan et al. identified three classes defined by low, moderate, and high severity of all symptoms, respectively. They found substantial overlap of symptom dimension trajectories in girls over time, suggesting that there was no evidence for dimension-specific subtypes in this study. Ezpeleta et al. recently reported similar results in a mixed-gender community sample of children in Spain. Modelling trajectories of Rowe et al.'s (2010) Irritability-Headstrong dimensions plus limited prosocial emotions (LPE; e.g., lack of remorse, callousness, lack of concern about performance, shallow/deficient affect; APA, 2013) from ages 3 to 12, they identified four classes: Defiant-Irritability-LPE (7.9%), Defiant-Irritability-without LPE (26.3%), LPE only (20.7%), and All Low (45.1%) groups. That is, classes were differentiated by the severity of symptoms and the relative presence/absence of ODD symptoms and LPE over time, but the two ODD symptom

dimensions tended to co-develop. Although these person-centered studies provide useful information about ODD symptom dimensions as they present in youth, they are limited by their reliance on predefined symptom categories to identify possible subtypes.

In the case of a diagnostic construct like ODD, where there is still some debate about symptom dimensions, latent class analysis (LCA) is a particularly useful person-centered approach for examining subtypes. It involves identifying latent classes (or subgroups) of individuals who show similar patterns of categorical item endorsement (e.g., yes or no responses to the various symptoms of ODD; Collins & Lanza, 2010). With LCA, no a priori assumptions are made about which symptoms will differentiate classes, since class structure is data-driven. To date, at least nine studies have investigated possible ODD subgroups using LCA (Aebi, Barra, et al. 2016; Aebi, van Donkelaar, et al., 2016; Althoff et al., 2014; Burke, 2012; Gomez & Stavropoulos, 2018; Herzoff & Tackett, 2016; Kuny et al., 2013; Roetman et al., 2021; Wesselhoeft et al., 2019). Drawing from different populations, these studies have typically identified three or four latent classes of youth based on ODD symptom endorsement patterns.

Two studies have identified three latent classes of youth that differed in terms of the relative presence of affective and behavioural ODD symptoms (Burke, 2012; Herzoff & Tackett, 2016). In a sample of 177 clinic-referred boys aged 7-12 years from the U.S. (i.e., Pennsylvania and Georgia), Burke (2012) identified three latent classes using parent-reported ODD symptoms on a modified version of the Diagnostic Interview Schedule for Children (DISC). The largest group (47.5%) had a high probability of displaying behavioural (i.e., loses temper, argues, defies, and to a lesser extent, blames), but not affective (i.e., touchy, angry, spiteful, annoys), ODD symptoms, labeled the *Oppositional Behavior* class. The second largest group (36.0%) of boys was distinguished by a high probability of all eight ODD symptoms, labeled the *Irritable* class to reflect the presence of affective ODD symptoms. Finally, a small subset (16.4%) of boys in this clinic-referred sample were unlikely to display any ODD symptoms, labeled the *Low* class. Longitudinally, Burke found that the Irritable group was more likely than the other two groups to display internalizing problems throughout adolescence (i.e., parent-rated depression and anxiety) and at age 18 (i.e., self-reported depression). Boys who had been in the Irritable group were also the most likely to self-report neurotic personality traits at age 18.

Herzoff and Tackett (2016) identified three latent classes using combined parent and child ratings on the DISC in two separate community-based, mixed-sex samples. In a primarily Caucasian sample of 439 children (M age = 9.96 years) recruited in Southern Ontario, Herzoff and Tackett similarly identified *Low* (74.6%), *Oppositional* (10.8%), and *Irritable* (14.7%) subgroups. However, in this sample, the Oppositional group had the highest probability of displaying behavioural symptoms (i.e., loses temper, argues, and defies), whereas the Irritable group had the highest probability of exhibiting affective symptoms (i.e., touchy, angry, and to a lesser extent, spiteful). In a more ethnically diverse replication sample of 291 children (aged 8-10 years) recruited in Texas, the three identified classes more closely resembled those identified by Burke (2012). While most children were unlikely to display any ODD symptoms (the *Low* class; 68.6%), 28.3% of the sample was moderately likely to display mainly behavioural symptoms (i.e., loses temper, argues, defies; the *Oppositional* class), and 3.1% had a high probability of displaying all ODD symptoms (the *Oppositional+Irritable* class).

In other person-centered studies conducted to date, four ODD symptom classes have emerged (Aebi, Barra, et al., 2016; Aebi, van Donkelaar, et al., 2016; Althoff et al., 2014; Kuny et al., 2013; Wesselhoeft et al., 2019). Across these studies, latent classes characterized by low and high probabilities of all ODD symptoms, respectively, have been identified. Two additional classes characterized by either primarily oppositional or primarily irritable ODD symptoms have also been found in three of these studies (Aebi, Barra, et al., 2016; Althoff et al., 2014; Kuny et al., 2013). Using diagnostic interview data, Aebi, Barra, et al. found two subgroups with moderate ODD symptoms in a sample of 158 male adolescent offenders detained in Zurich, Switzerland. In addition to *No ODD* (22.2%) and *Serious ODD* (32.3%) subgroups, 20.9% of the forensic sample were most likely to endorse affective symptoms (i.e., loses temper, touchy, angry, and spiteful; the *Irritable ODD* class), and 24.7% were most likely to display defiant behaviour (i.e., argues and defies; the *Defiant ODD* class). While those in the Serious ODD group showed the highest rates of comorbidity overall, membership in the Irritable group was associated with comorbid CD, anxiety, and suicidality, whereas membership in the Defiant group was only associated with concurrent CD. Moreover, survival analyses revealed that, in the year following release from detention, those in the Serious ODD and Irritable ODD classes had shorter times to reoffending.

In two large-scale, representative community samples, Althoff et al. (2014) also found that two out of four of the identified classes were distinguished by the relative presence of symptoms reflecting either defiant behaviour or negative affect. In the Achenbach et al. (2002) normative sample and the Zuid-Holland sample (Hofstra et al., 2001), the researchers assessed ODD symptoms among youth (i.e., aged 6-18 and 4-16 years, respectively), using parent ratings on five items of the Oppositional Problems subscale of the Child Behavior Checklist (Achenbach & Rescorla, 2001). In both samples, youth in the identified *Irritable* class had a relatively high probability of displaying affective symptoms (i.e., argues, stubborn/sullen/irritable, and, to a lesser extent, temper tantrums), whereas youth in the *Defiant* class were relatively likely to display primarily defiant behaviours (i.e., argues, disobedient at home/school). The prevalence rates of these two classes differed between samples (i.e., Achenbach: 22% Irritable and 7% Defiant; Zuid-Holland: 16% Irritable and 18% Defiant), as did the prevalence rates of the identified *High Symptom* (Achenbach: 36%; Zuid-Holland: 17%) and *No Symptom* (Achenbach: 35%; Zuid-Holland: 50%) classes. Membership in the High Symptom class was associated with lifetime diagnoses of ODD, CD, and ADHD, and membership in the Irritable class was associated with lifetime anxiety disorders (i.e., measured concurrently in a different U.S. sample¹). Longitudinally in the Zuid-Holland sample, those in the High Symptom and Defiant classes were more likely to display delinquent behaviour later as adults (i.e., at 18-32 years old), whereas those in the Irritable class were most likely to meet criteria for any mood disorder.

In the only published study that has used LCA to investigate ODD symptom classes longitudinally to date, Kuny et al. (2013) conducted LCAs in a sample of 14,844 children from the Netherlands, at ages 7, 10, and 12. Using data on six items from the Conners Parent Rating Scales Revised Short Form (Conners, 2001), they found the four-class solutions across the three time points, with only minor variability in the prevalence and item-response probabilities for each class across ages. Like the other studies described above, these researchers identified *No symptom* (69-75% of the sample across ages) and *High symptom* (5-9%) classes. The *Defiant* class (11-12%) displayed moderate probabilities of endorsing the symptoms argues with adults and

¹ Results regarding concurrent diagnostic outcomes refer to analyses conducted with data from the Vermont Family Study (VFS). Specifically, class weights from the Achenbach normative sample were used to fit participants from the VFS into classes before conducting the logistic regressions.

angry/resentful. In contrast, those in the *Irritable* class (9-11%) were relatively likely to endorse the symptoms loses temper, irritable, and angry/resentful. Consistent with previous findings, membership in the Irritable class was associated with higher symptoms of anxious-depression at all ages, and higher symptoms of withdrawn-depression and overall internalizing problems at age 7. Over time, class membership remained relatively stable. Although membership in the Defiant and Irritable classes predicted later membership in the High symptom class (and vice versa) at certain time points, no evidence was found for significant crossover between the Defiant and Irritable classes over time.

It is important to note that a primarily defiant class has not been consistently identified in studies that reported four-class solutions. In a sample of 750 children with ADHD from various European countries (M age = 10.67 years), Aebi, van Donkelaar, et al. (2016) identified four latent classes (i.e., *Low*, *Moderate*, *Irritable*, and *Severe*) using 10 ODD items from the Conners Parent Report (Conners, 1997), but they found a group characterized by moderate probability of endorsing all ODD symptoms instead of a primarily defiant symptom subgroup. Similarly, in a representative sample of 3,435 Nordic children (M age = 8.94 years), Wesselhoelf et al. (2019) identified a *Medium* class in addition to *Low*, *Angry*, and *High* symptom classes using parent ratings on the Development and Well-Being Assessment (DAWBA; Goodman et al., 2000).

Moreover, two recent studies suggest that the severity of ODD symptoms may be more important for distinguishing classes of youth (Gomez & Stravropoulos, 2018; Roetman et al., 2021). In a representative sample of 457 Australian children (M age = 8.60 years), Gomez and Stravropoulos (2018) compared confirmatory factor analytic, latent class, and factor mixture modelling (FMM) approaches using maternal ratings of ODD symptoms on the Disruptive Behavior Rating Scale (Barkley & Murphy, 1998). In the LCA, the researchers found that a two-class model (*High* vs *Low* severity) fit the data best. In this sample, however, they found the most support for a FMM solution that distinguished between three classes of children based on the severity of ODD symptoms (i.e., *Unaffected*, *At-Risk*, and *Affected* classes) as well as the relative prominence of ODD symptom dimensions within each class (i.e., oppositional, antagonistic, and negative affect dimensions; i.e., based on Burke et al.'s [2010] factor model). The *Affected* (4.80%), *At-Risk* (12.69%), and *Unaffected* (82.50%) classes showed high, moderate, and low probabilities of endorsing all the ODD symptoms, respectively.

According to the researchers, the Affected class had relatively higher probabilities of endorsing antagonistic symptoms, while the At-Risk and Unaffected classes had relatively higher probabilities of endorsing oppositional symptoms. The Unaffected class did not endorse any level of the spiteful and vindictive symptom.

In a clinic-referred sample of 2,185 children and adolescents (5 to 18 years old) in the Netherlands, Roetman et al. (2021) conducted a LCA using a combination of parent and teacher ratings of ODD symptoms on the DAWBA (Goodman et al., 2000). They identified three classes of youth based on nine ODD symptoms (where vindictiveness and spitefulness were measured separately), which differed in terms of the relative probability of endorsing all ODD symptoms: a *High* class (25.8% of the sample), a *Moderate* class (39.4%), and a *Low* class (34.7%). Compared to the two lower severity classes, youth in the High class were the most clinically affected, showing the highest rates of mental health/social problems and diagnoses at the time of referral², the highest rates of ODD and CD at the end of the multidisciplinary team-based diagnostic process, and the worst clinician-rated global functioning across the study period. Compared to the Low class, youth in the Moderate class also had more mental health/social problems and diagnoses at referral³, a higher rate of clinically confirmed ODD diagnoses, and greater impairment across the study period. Although youth in the Low class were most likely to have generalized anxiety disorders (and to a lesser extent, fear-based disorders) based on clinician report, they were the least affected in terms of mental health/social problems, diagnoses, and impairment overall. Thus, in this study, clinic-referred youth were distinguished by the relative severity of their ODD symptoms (vs. the relative presence of affective and/or behavioural symptoms), and these severity-based classifications were found to be clinically meaningful.

In sum, existing evidence from person-centered studies suggests that latent subgroups of youth can be identified based on their endorsement of ODD symptoms. In most studies using LCA to date, Low and High symptom classes have been identified

² At referral, the High class displayed higher rates of ODD, CD, ADHD than the Moderate and Low classes, as well as higher rates of autism spectrum disorder and generalized anxiety disorder than the Low class; however, the High class did not differ from the other classes in terms of depressive or fear disorders at intake.

³ The Moderate and Low classes showed comparable levels of internalizing disorders and dimensionally measured emotional problems at referral.

(Aebi, Barra et al., 2016; Aebi, van Donkelaar, et al., 2016; Althoff et al., 2014; Burke, 2012; Gomez & Stavropoulos, 2018; Herzoff & Tackett, 2016 [one sample]; Kuny et al., 2013; Roetman et al., 2021; Wesselhoeft et al., 2019). Some support has also been found for so-called primarily oppositional and/or primarily irritable subgroups, with membership in these groups differentially predicting both concurrent and long-term outcomes (e.g., Aebi, Barra, et al., 2016; Althoff et al., 2014; Kuny et al., 2013). However, there is also some evidence that latent classes are distinguished mainly by the relative severity of ODD (Gomez & Stavropoulos, 2018; Roetman et al., 2021). Moreover, important gaps in the current evidence base leave uncertainty about whether meaningful subtypes of ODD exist. As mentioned above, only one published study has explicitly examined ODD symptom classes over time (Kuny et al., 2013). Those analyses utilized six items from the Oppositional Subscale of the Conners Parent Rating Scales Revised Short Form (Conners, 2001) rather than the eight DSM-5 symptom criteria for ODD (APA, 2013). There is also a relative lack of person-centered research on ODD symptom classes into later adolescence, despite evidence that ODD symptoms can persist into this period of development (e.g., Leadbeater & Homel, 2015; Leadbeater et al., 2012). Consequently, much remains to be learned about how potential ODD symptom subgroups present across time.

The Present Study

Given the risks that ODD poses for later mental health and adaptive functioning (e.g., Burke, Rowe, & Boylan, 2014; Copeland et al., 2009; Nock et al., 2007), a clearer understanding of the possible presentations of the disorder over time is needed. Substantial evidence has accumulated to support the distinction between affective and behavioural symptom dimensions of ODD (Burke et al., 2018; Frick & Nigg, 2012), yet person-centered studies on possible ODD subtypes are currently lacking in both number and scope. To address the limitations of the existing evidence base, the present study utilized longitudinal data on DSM-defined ODD symptom criteria from a high-risk, racially diverse U.S. sample of children to: (1) investigate latent ODD symptom classes at four different time points from middle childhood to late adolescence; (2) examine measurement invariance of identified classes across these time points; and (3) investigate whether relevant demographic variables were associated with latent class

membership. The results may have important implications for the assessment and treatment of ODD at various stages of youth development.

Method

Participants

This study utilized data from the Fast Track Project (Conduct Problems Prevention Research Group [CPPRG], 2019), which began with the selection of 55 elementary schools deemed to be high risk based on neighbourhood crime and poverty in four geographical locations in the United States: Seattle, Washington ($n = 15$); Nashville, Tennessee ($n = 10$); Durham, North Carolina ($n = 13$); and rural Pennsylvania ($n = 17$). Schools within these sites were then matched based on size, ethnic composition, and poverty rates, and randomly assigned to either intervention or control conditions. Within these schools, 9,594 kindergarten children (approximately 6 years of age) across three cohorts (1991–1993) participated in a multiple-gating screening procedure (Lochman & CPPRG, 1995). The procedure began with a screening for classroom conduct problems that utilized teacher ratings on items from the Teacher Observation of Classroom Adaptation – Revised (e.g., stubborn, breaks rules, harms others; Werthamer-Larsson et al., 1991). Children in the top 40% of their cohort and site were then screened for externalizing behaviour problems at home using parent ratings on 24 items drawn from existing behaviour checklists (i.e., Child Behavior Checklist [Achenbach, 1991], Revised Problem Behavior Checklist [Quay & Peterson, 1987]). Next, a total severity-of-risk screen score was created by summing standardized scores on the parent and teacher screens, and children with the highest total risk scores were selected for inclusion in the study until desired sample sizes were reached across sites, schools, and groups. Children were not included in the study if they refused to participate or failed to enroll in first grade at 1 of the 55 core schools. Deviations from the selection procedure were also made to ensure that no child would be the only girl in an intervention group.

Ultimately, 891 children participated in the randomized controlled trial, with 445 and 446 children assigned to the intervention and high-risk control groups, respectively. In addition, a normative comparison sample ($n = 387$) was selected from children in the first cohort from the control schools (i.e., approximately 100 children from each site). This longitudinal comparison group, which was randomly selected after participants had been stratified according to race, sex, and teacher-rated conduct problems scores, was

used to represent the population-normative range of conduct problems within the schools. After initial recruitment in kindergarten, children assigned to the intervention group participated in a multiple-phase prevention program until the end of grade 10. Annual assessments of all children took place post-kindergarten through 2 years post-high school, with the entire sample being assessed again at age 25.

In the current project, only data from the high-risk control group were utilized ($n = 446$). The high-risk sample was 66.37% male, 48.88% White, 48.43% Black 1.57% Hispanic, 0.22% Asian, and .90% Other race.

Measures

At the outset of the Fast Track project (i.e., in kindergarten), parents reported on child sex, child race, and family socioeconomic status (SES). In addition, cohort and geographic site were recorded for each participant, and initial risk scores were calculated in the manner described above. For the purpose of this study, a dichotomous *urban status* variable (0 = rural, 1 = urban) was created to represent whether the child had been recruited from an urban (i.e., Durham, Nashville, Seattle; 74.89%) or rural (i.e., central Pennsylvania; 25.11%) geographic site. In addition, a binary indicator of child race (0 = non-Black, 1 = Black) was used, since the prevalence of non-Black minority races was very low (2.69%) in the sample. While an indicator of Black/non-Black race was included due to the diversity of the sample, it is important to acknowledge that race is a social construct and, thus, is not in itself associated with ODD symptomatology. Rather, the present findings should be interpreted with the understanding that the indicator of Black/non-Black race represents a host of social, economic, political, and other factors (e.g., parenting) that intersect with child race, the complexity of which was not captured in this study.

ODD symptom criteria were assessed in the summers following grades 3, 6, 9, and 12 using the NIMH Computerized Diagnostic Interview Schedule for Children – Parent Version (CDISC; Shaffer & Fisher, 1991; 1997). The CDISC is a highly structured diagnostic interview that is widely used to assess child psychiatric symptoms and has adequate reliability and validity (Schwab-Stone et al., 1996; Shaffer et al., 1996; Shaffer et al., 2000). Primary caregivers (usually mothers) answered a series of questions about whether the target children had experienced ODD symptoms in the past 6 months (for

grade 3) or in the past year (for grades 6, 9, and 12). In grade 3, version 3.0 of the CDSIC was administered in paper and pencil format and approximated proposed DSM-IV criteria for ODD at the time. In grades 6, 9, and 12, version 4 of the CDISC was administered via computer and was in line with published DSM-IV ODD criteria. Symptom questions were scored to determine whether ODD symptom criteria were met, with multiple items sometimes being used to determine the presence or absence of a given criterion. Across time points, criterion counts were calculated by summing the total number of ODD symptom criteria met at each time point. Finally, criterion scores were used to determine whether children were eligible for a diagnosis of ODD. Notably, the procedures used for scoring ODD criteria were similar across the two versions of the CDISC utilized in this study, with the one exception being the spiteful and vindictive criterion. In grade 3 (CDISC 3.0), spiteful and vindictive was scored considering three interview questions related to getting even with people, being mean or nasty, and trying to get others in trouble. In later grades (CDISC 4), only the first two interview questions were considered in scoring this criterion. To ensure scoring accuracy, lay interviewers were trained by Prudence Fisher of Columbia University (i.e., the co-developer of the CDISC) or someone who had been trained directly by her. For the latent class analyses conducted at each time point, eight binary indicators (0 = no, 1 = yes) were used to represent whether a study child had met each of the specific ODD symptom criteria.

Missing Data

Missing data on ODD criteria increased over time, with rates of overall missingness ranging from 14.3% ($n = 64$) in grade 3 to 41.9% ($n = 187$) in grade 12 (See Table 1 for item-level frequencies). Participants with and without missing data on ODD criteria at each of the four time points were compared on demographic variables (i.e., age, sex, race, urban status, family SES), initial screen scores, kindergarten externalizing and internalizing problems (as measured by the raw Externalizing and Internalizing scales of the Child Behavior Checklist [Achenbach, 1991] in the summer following kindergarten), and grade 3 ODD criterion counts (in the case of later grades). In terms of demographic variables, initial child age and family SES were not associated with missingness on ODD criteria at any time point. Missingness was significantly associated with: race, $\chi^2(1) = 7.30, p = .007$ (grade 3 only); sex, $\chi^2(1) = 4.07, p = .044$ (grade 6 only); and urban status, $\chi^2(1) = 12.47, p = .000$ (grade 12 only). In grade 9,

individuals with missing data had significantly lower mean initial screen scores, $t(438) = 2.94, p = .003$. Otherwise, initial screen scores, kindergarten externalizing problems, kindergarten internalizing problems, and grade 3 ODD criterion counts were not significantly associated with missingness on ODD symptom criteria across time. Therefore, there was very minimal evidence of differential attrition based on early behavioural or emotional symptoms in the present study.

Analysis Plan

Data analyses proceeded in four general steps. In step one, descriptive statistics and bivariate correlations were examined in *SPSS 24* (IBM Corp, 2016).

Subsequent analyses were conducted using *Mplus 8.4* (Muthen & Muthen, 1998-2017). To maximize power, all participants who had some data on the outcome variables were included and missing data were accommodated using full information maximum likelihood (FIML) estimation, which assumes that data are missing at random. A maximum likelihood estimator with robust standard errors (MLR) was used.

In step two, LCAs were conducted separately in grades 3, 6, 9, and 12 to determine the number of latent classes required to characterize heterogeneity within the sample at each time point. LCA is a person-centered statistical method that aims to identify a finite number of unobserved homogenous subgroups/classes of individuals in the population who are similar to each other in terms of their patterns of responses on a set of categorical variables (e.g., DSM-defined ODD symptoms; Collins & Lanza, 2010). Two main parameters are estimated in LCA: (1) unconditional latent class membership probabilities (e.g., the prevalence of each latent ODD symptom class) and (2) conditional item-response probabilities (i.e., the probability that an individual in a given class will endorse a particular item, such as often loses temper; Lanza & Cooper, 2016). LCA involves fitting a series of models to determine the number of classes required to sufficiently capture the heterogeneity in the population (Nylund, 2007). Typically, the series of models is examined in a stepwise fashion (i.e., adding a class with each model) and the best-fitting model is selected. In the present project, 1- to 6-class solutions were examined. In line with recommendations presented by Nylund and Choi (2018), latent classes were enumerated using unconditional models (i.e., prior to including covariates) at each time point.

The following indices were considered in examining model fit: (a) information criteria, including the Bayesian information criteria (BIC), the sample size-adjusted BIC (aBIC), and the Akaike information criteria (AIC), with lower values reflecting better model fit (Nylund et al., 2007); and (b) likelihood ratio tests, including the Lo-Mendell-Rubin adjusted likelihood ratio test (LMR-LRT: Lo et al., 2001) and the bootstrapped likelihood ratio test (BLRT). These tests are statistical comparisons of fit between a given model with k classes and a model with one fewer classes (i.e., $k-1$), with significant p values indicating improved model fit with the k -class model. In addition to these statistical criteria, entropy (an indicator of class separation), parsimony, and the interpretability of models were taken into consideration (Collins & Lanza, 2010; Nylund, 2007).

In step three, measurement invariance of the best-fitting unconditional latent class models was examined at adjacent time points (i.e., grades 3 and 6, grades 6 and 9, and grades 9 and 12), to determine whether there were differences in the structure of latent ODD classes over time. In each set of analyses, models assuming non-invariance, full invariance, and partial invariance were compared (Nylund, 2007). For non-invariance, item thresholds within classes were allowed to vary freely across adjacent time points. For full invariance, item thresholds in all classes were constrained to be equal across adjacent time points. For partial invariance, item thresholds for one to two classes were constrained to be equal across adjacent time points, and remaining classes were allowed to freely vary. Using -2 loglikelihood difference tests with scaling correction factors (Muthen & Muthen, no date; see also Satorra & Bentler, 2010), models of full invariance were compared to models of non-invariance and partial invariance, and then models of partial invariance were compared to models of non-invariance. A non-significant -2 log likelihood difference test suggests that the more restrictive (i.e., more invariant) model in the comparison is a better fit for the data.

In the fourth step, relevant demographic variables were included as covariates in the latent class model selected at each time point, to determine if they were associated with latent class membership (Nylund & Choi, 2018). Given existing evidence for sex differences in the prevalence of ODD (APA, 2013; Demmer et al., 2017), child sex was selected *a priori* as a potential covariate. Using a data-driven approach, other demographic variables (i.e., initial child age, race, urban status, family SES, initial screen scores) were also considered as possible covariates. A covariate was selected for

inclusion if it was found to significantly correlate with both ODD diagnosis *and* ODD criterion count at any single time point, or with diagnoses or criterion counts across two or more time points. Separately for grades 3, 6, 9, and 12, latent ODD classes were regressed on selected covariates. Odds ratios were examined to determine whether select latent class prevalence significantly differed by levels of the covariate. Given that the addition of covariates can result in different latent class solutions (Nylund & Choi, 2018), final or conditional class membership probabilities and item-response probabilities were inspected at each time point for evidence of substantive changes in latent class structure. In all applicable analyses, alpha was pre-set to 0.05.

Results

Step 1: Descriptive Statistics

Descriptive statistics for demographic variables and ODD diagnoses and criterion counts are presented in Table 2. Frequencies for the eight ODD criteria across grades 3, 6, 9, and 12 are presented in Table 1. The percentage of children meeting diagnostic criteria for ODD was 13.18% in grade 3, 17.81% in grade 6, 15.52% in grade 9, and 10.04% in grade 12. Argues was the most frequently endorsed criterion across grades (17.37-30.96%), whereas the spiteful and vindictive criterion tended to be the least frequently endorsed (1.54-11.40%).⁴ For most ODD symptom criteria, frequencies peaked in grade 6 or grade 9 and then decline thereafter. The only exception was the frequency of the spiteful and vindictive criterion, which steadily declined across time.

Table 3 presents bivariate correlations between demographic variables and ODD diagnoses and criterion counts. ODD diagnoses were significantly correlated across time (significant r 's ranging from 0.16 to 0.38), as were ODD criterion counts (significant r 's ranging from 0.23 to 0.54); however, grade 3 ODD diagnosis was not significantly correlated with grade 12 ODD diagnosis. Initial child age, child sex, urban status, and family SES did not show robust correlations with ODD diagnoses or criterion counts over time. The indicator of Black race was negatively correlated and total screen scores were positively correlated with ODD diagnoses and criterion counts across grades (significant r 's ranging from -0.23 to -0.11 and 0.12 to 0.23, respectively). That said, total screen scores were not significantly correlated with ODD diagnosis or criterion count at grade 6.

Step 2: Unconditional Latent Class Models of ODD Criteria Across Grades

Table 4 presents model fit indices for the 1- to 6-class unconditional latent class models examined in grades 3, 6, 9, and 12. Entropy was satisfactory for all models. A 3-class latent solution was supported across grades.

⁴ In grade 3, the frequency of angry/resentful (11.11%) was slightly lower than the frequency of spiteful and vindictive when original scoring was used (11.40%). Using revised scoring (described in detail below), spiteful and vindictive had the lowest frequency in grade 3 (10.10%).

Table 1. Frequencies of ODD symptom criteria across grades

ODD Criterion	Frequency (valid %)			
	Grade 3	Grade 6	Grade 9	Grade 12
Loses temper				
No	320 (82.69)	282 (77.26)	275 (82.09)	224 (86.49)
Yes	67 (17.31)	83 (22.74)	60 (17.91)	35 (13.51)
Missing	59	81	111	187
Argues				
No	314 (81.14)	252 (69.04)	248 (74.03)	214 (82.63)
Yes	73 (18.86)	113 (30.96)	87 (25.97)	45 (17.37)
Missing	59	81	111	187
Defies/refuses				
No	327 (84.28)	285 (78.08)	259 (77.31)	227 (87.64)
Yes	61 (15.72)	80 (21.92)	76 (22.69)	32 (12.36)
Missing	58	81	111	187
Annoys				
No	342 (88.37)	314 (86.03)	295 (88.06)	237 (91.51)
Yes	45 (11.63)	51 (13.97)	40 (11.94)	22 (8.49)
Missing	59	81	111	187
Blames				
No	336 (86.60)	314 (85.03)	312 (93.13)	245 (94.59)
Yes	52 (13.40)	51 (13.97)	23 (6.87)	14 (5.41)
Missing	58	81	111	187
Touchy				
No	340 (87.63)	297 (81.40)	272 (81.19)	228 (88.03)
Yes	48 (12.37)	68 (18.63)	63 (18.81)	31 (11.97)
Missing	58	81	111	187
Angry/resentful				
No	344 (88.89)	295 (80.82)	285 (85.07)	222 (85.71)
Yes	43 (11.11)	70 (19.18)	50 (14.93)	37 (14.29)
Missing	59	81	111	187

ODD Criterion	Frequency (valid %)			
	Grade 3	Grade 6	Grade 9	Grade 12
Spiteful/vindictive [revised]				
No	342 (88.60) [347 (89.90)]	344 (94.25)	327 (97.61)	255 (98.46)
Yes	44 (11.40) [39 (10.10)]	21 (5.75)	8 (2.39)	4 (1.54)
Missing	60 [60]	81	111	187

Note. Data in brackets reflect frequencies using the revised scoring criteria for spiteful and vindictive in grade 3

Table 2. Descriptives

	Valid <i>N</i>	<i>M (SD)</i> or <i>N (%)</i> of Yes	Range	Skewness	Kurtosis
Age ^a	446	5.89 (0.54)	5.00-7.00	-0.08 (0.12)	0.26 (0.23)
Male ^a	446	296 (66.37)	-	-	-
Race ^a					
Black	446	216 (48.43)	-	-	-
White	446	218 (48.88)	-	-	-
Hispanic	446	7 (1.57)	-	-	-
Asian	446	1 (0.22)	-	-	-
Other race	446	4 (0.90)	-	-	-
Urban site ^a	446	334 (74.89)	-	-	-
Family SES ^a	446	24.17 (12.46)	4.50-66.00	0.62 (0.12)	-.13 (0.23)
Initial screen score ^a	440	125.05 (11.20)	103.05-173.38	0.88 (0.12)	1.23 (0.23)
ODD					
Gr 3 6 mo. ODD dx	387	51 (13.18)	-	-	-
Gr 3 6 mo. criterion count	388	1.12 (2.06)	0.00-8.00	1.94 (0.12)	2.78 (0.25)
Gr 6 1 yr. ODD dx	365	65 (17.81)	-	-	-
Gr 6 1 yr. criterion count	365	1.47 (2.20)	0.00-8.00	1.49 (0.13)	1.14 (0.26)
Gr 9 1 yr. ODD dx	335	52 (15.52)	-	-	-
Gr 9 1 yr. criterion count	335	1.21 (1.94)	0.00-8.00	1.58 (0.13)	1.51 (0.27)
Gr 12 1 yr. ODD dx	259	26 (10.04)	-	-	-
Gr 12 1 yr. criterion count	259	0.85 (1.72)	0.00-7.00	2.15 (0.15)	3.68 (0.30)

Note. a = Assessed in kindergarten, SES = socioeconomic status, ODD dx = oppositional defiant disorder diagnosis.

Table 3. Bivariate correlations between demographic characteristics, ODD diagnoses, and ODD criterion counts

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Male ^a	-												
2. Black ^a	.01	-											
3. Urban ^a	.04	.53**	-										
4. Age ^a	.06	-.08	-.08	-									
5. Family SES ^a	-.00	-.12**	-.05	-.11*	-								
6. Screen score ^a	.09	.04	-.16**	-.04	-.08	-							
7. ODD dx 3	.08	-.06	-.06	-.02	-.02	.23**	-						
8. ODD cc 3	.10*	-.15**	-.08	.02	-.02	.21**	.89**	-					
9. ODD dx 6	.02	-.12*	-.03	.02	.09	.02	.38**	.42**	-				
10. ODD cc 6	.04	-.13*	-.05	.02	.11*	.08	.43**	.48**	.88**	-			
11. ODD dx 9	.06	-.04	.05	.11*	-.02	.12*	.16**	.22**	.31**	.39**	-		
12. ODD cc 9	.02	-.11*	.01	.08	-.00	.16**	.21**	.28**	.44**	.54**	.87**	-	
13. ODD dx 12	-.02	-.21*	.06	-.02	.07	.11	.10	.19**	.32**	.38**	.29**	.35**	-
14. ODD cc 12	.02	-.23**	-.01	-.00	.06	.14*	.16*	.23**	.33**	.43**	.38**	.46**	.88**

Note. ^a = Assessed in kindergarten, SES = socioeconomic status, ODD dx 3 = past 6-month ODD diagnosis in grade 3, ODD cc 3 = past 6-month ODD criterion count in grade 3, ODD dx 6 = past year ODD diagnosis in grade 6, ODD cc 6 = past year ODD criterion count in grade 6, ODD dx 9 = past year ODD diagnosis in grade 9, ODD cc 9 = past year ODD criterion count in grade 9, ODD dx 12 = past year ODD diagnosis in grade 12, ODD cc 12 = past year ODD criterion count in grade 12. * $p < .05$, ** $p < .01$

Table 4. Summary of information for selecting the number of latent classes in grades 3, 6, 9, and 12

Number of latent classes	LL	AIC	BIC	aBIC	LMR-LRT <i>p</i> -value	BLRT <i>p</i> -value	Entropy
Grade 3 – original scoring (<i>n</i> = 388)							
1	-1243.634	2503.269	2534.957	2509.573	-	-	-
2	-870.566	1775.133	1842.470	1788.530	0.000	0.000	0.930
3	-836.323	1724.646	1827.632	1745.136	0.001	0.000	0.880
4	-826.127	1722.254	1860.890	1749.838	0.087	1.000	0.896
5	-817.891	1723.783	1898.067	1758.459	0.263	0.600	0.913
6	-808.273	1722.547	1932.480	1764.316	0.112	0.333	0.932
Grade 3 – revised spiteful and vindictive item (<i>n</i> = 388)							
1	-1233.049	2482.099	2513.787	2488.403	-	-	-
2	-861.812	1757.625	1824.962	1771.022	0.000	0.000	0.932
3	-825.817	1703.633	1806.619	1724.124	0.001	0.000	0.877
4	-814.100	1698.200	1836.835	1725.783	0.057	0.082	0.884
5	-806.845	1701.690	1875.974	1736.366	0.728	1.000	0.906
6	-799.240	1704.479	1914.412	1746.248	0.208	0.500	0.930
Grade 6 (<i>n</i> = 365)							
1	-1342.988	2701.976	2733.175	2707.795	-	-	-
2	-983.337	2000.673	2066.972	2013.038	0.000	0.000	0.917
3	-942.726	1937.452	2038.849	1956.362	0.079	0.000	0.907
4	-926.465	1922.929	2059.426	1948.385	0.430	0.000	0.918
5	-917.146	1922.292	2093.888	1954.294	0.324	0.375	0.873
6	-906.045	1918.089	2124.784	1956.637	0.030	0.087*	0.855
Grade 9 (<i>n</i> = 335)							
1	-1075.924	2167.847	2198.360	2172.983	-	-	-
2	-787.707	1609.414	1674.255	1620.329	0.000	0.000	0.925
3	-773.501	1599.002	1698.169	1615.695	0.010	0.000	0.928
4	-767.844	1605.687	1739.182	1628.158	0.673	1.000	0.940
5	-759.636	1607.271	1775.093	1635.521	0.619	0.333	0.956
6	-752.902	1611.804	1813.953	1645.832	0.118	0.429	0.962
Grade 12 (<i>n</i> = 259)							
1	-670.518	1357.035	1385.490	1360.127	-	-	-
2	-462.826	959.652	1020.118	966.222	0.000	0.000*	0.947
3	-448.750	949.500	1041.978	959.548	0.087	0.000	0.907
4	-440.681	951.363	1075.852	964.889	0.071	0.333	0.953
5	-432.180	952.360	1108.860	969.364	0.132	0.667	0.952
6	-426.439	958.877	1147.389	979.360	0.530	0.375	0.950

Note. LL = log likelihood, AIC = Akaike information criterion, BIC = Bayesian information criterion, aBIC = sample-size adjusted BIC, LMR-LRT = Lo-Mendell-Rubin adjusted likelihood ratio test, BLRT = bootstrapped likelihood ratio test. Best model bolded. * = Model estimation indicated that this value may not be reliable.

In grade 3, the 3-class solution was supported by lower BIC and aBIC statistics as well as significant LMR-LRT and BLRT p -values. In grade 9, the 3-class solution was supported by lower AIC and aBIC values, as well as significant LMR-LRT and BLRT p -values. In grade 12, the AIC, aBIC, and BLRT also supported the 3-class solution. Model selection was more complicated in grade 6. The AIC did not clearly differentiate solutions at this time point (i.e., it continued decreasing across the 1- to 6-class models). While the aBIC and BLRT provided some support for the 4-class solution in grade 6, the BIC provided support for the 3-class solution, and the LMR-LRT indicated that the 2-class solution was preferable. Given the relatively small reductions in AIC and aBIC values from the 3- (AIC = 1937.452, aBIC = 1956.362) to the 4-class (AIC = 1922.929, aBIC = 1948.385) solutions, and in the interest of maximizing parsimony and interpretability both in grade 6 and across time points, the 3-class solution was chosen as the preferred model in grade 6.

In further support of selecting the 3-class solutions, closer inspection of the unconditional 3-class models suggested a similar structure of parent-rated ODD classes across time (See Table 5 and Figure 1). Specifically, classes characterized by low, moderate, and high probabilities of item endorsement were identified across grades. In grade 3, the class prevalence rates were 72.82% Low, 18.16% Moderate, and 9.03% High. In grade 6, the class prevalence rates were 69.55% Low, 22.16% Moderate, and 8.29% High. In grade 9, the class prevalence rates were 73.22% Low, 23.24% Moderate, and 3.53% High. In grade 12, the classes prevalence rates were 78.17% Low, 14.52% Moderate, and 7.31% High. Despite the prevalence of the High class in grade 9 being quite low (i.e., 3.53%), this model was selected given considerable support for the existence of a similar class in other grades and in other LCA studies (Aebi, Barra, et al. 2016; Aebi, van Donkelaar, et al., 2016; Althoff et al., 2014; Burke, 2012; Herzoff & Tackett, 2016; Kuny et al., 2013; Roetman et al., 2021; Wesselhoeft et al., 2019).

In terms of notable qualitative observations, the probability of endorsing the spiteful and vindictive symptom criterion appeared to decline in the High (and, to a lesser extent, Moderate) class across grades, which is consistent with descriptive statistics regarding overall item frequencies described above. Given the slight ODD criterion scoring difference in grade 3 (vs. later grades), the 1- to 6-class solutions were re-run in grade 3 using an alternative scoring procedure for the spiteful and vindictive criterion that more closely matched that used in later grades. Specifically, the spiteful and

vindictive criterion was re-scored in grade 3, so that endorsement of this criterion only referred to interview questions related to getting even with and being mean to others (i.e., excluding the DISC question related to getting others in trouble, as in the later grades). The fit statistics for these latent class solutions are also presented in Table 4. Similar to the results described above, the 3-class solution was best supported by the data, and had very similar latent class probabilities (i.e., 73.02% Low, 17.32% Moderate, and 9.66% High). Upon visual inspection, the effect of this scoring change appeared to modestly decrease the item-response probability for the spiteful and vindictive criterion in the Moderate class in grade 3, while item-response probabilities for other criteria appeared largely the same (See Table 6).

Table 5. Prevalence rates and item-response probabilities for unconditional 3-class models

	Grade 3 (<i>n</i> = 388)			Grade 6 (<i>n</i> = 365)			
	Low (72.82%)	Moderate (18.16%)	High (9.03%)	Low (69.55%)	Moderate (22.16%)	High (8.29%)	
Temper	.020	.437	.873	Temper	.049	.532	.912
Touchy	.009	.239	.819	Touchy	.018	.439	.926
Angry	.010	.160	.832	Angry	.019	.458	.932
Argues	.027	.556	.753	Argues	.088	.772	.936
Defies	.012	.439	.764	Defies	.009	.631	.878
Annoys	.000	.267	.768	Annoys	.007	.332	.736
Blames	.003	.356	.747	Blames	.037	.140	1.00
Spiteful	.005	.167	.878	Spiteful	.011	.027	.531
	Grade 9 (<i>n</i> = 335)			Grade 12 (<i>n</i> = 259)			
	Low (73.22%)	Moderate (23.24%)	High (3.53%)	Low (78.17%)	Moderate (14.52%)	High (7.31%)	
Temper	.017	.566	1.00	Temper	.003	.549	.731
Touchy	.037	.566	.843	Touchy	.001	.397	.842
Angry	.018	.444	.932	Angry	.004	.495	.928
Argues	.061	.785	.929	Argues	.040	.572	.818
Defies	.043	.698	.936	Defies	.024	.255	.929
Annoys	.008	.349	.909	Annoys	.004	.177	.767
Blames	.000	.143	1.00	Blames	.006	.081	.516
Spiteful	.004	.052	.251	Spiteful	.000	.000	.211



Figure 1. Unconditional 3-class solutions in grades 3, 6, 9, and 12

Table 6. Prevalence rates and item-response probabilities for unconditional 3-class models in grade 3 with different scoring procedures for spiteful and vindictive

	Grade 3 Original Scoring (n = 388)			Grade 3 Revised Spiteful and Vindictive Scoring (n = 388)		
	Low (72.82%)	Moderate (18.16%)	High (9.03%)	Low (73.02%)	Moderate (17.32%)	High (9.66%)
Temper	.020	.437	.873	.019	.446	.846
Touchy	.009	.239	.819	.009	.228	.805
Angry	.010	.160	.832	.010	.161	.791
Argues	.027	.556	.753	.024	.562	.758
Defies	.012	.439	.764	.011	.434	.767
Annoys	.000	.267	.768	.000	.256	.761
Blames	.003	.356	.747	.010	.329	.723
Spiteful	.005	.167	.878	.006	.066	.875

Step 3: Measurement Invariance

Given qualitative observations of similar class structure across grades, measurement invariance of unconditional 3-class solutions was examined at adjacent time points (i.e., grades 3 and 6, grades 6 and 9, and grades 9 and 12). In each set of analyses, models of full invariance, partial invariance, and non-invariance were compared.

Grades 3 and 6

Using original scoring for the spiteful and vindictive criterion in grade 3, the model of non-invariance fit significantly better than the model of full invariance, $\chi^2_{\text{invariance vs. noninvariance}}(24) = 58.42, p = .000$. Similarly, a model of partial invariance, in which item thresholds in two classes were constrained to be equal, fit significantly better than the model of full invariance, $\chi^2_{\text{invariance vs. partial-2}}(8) = 30.86, p = .000$. However, compared to the model of partial invariance in which item thresholds in two classes were constrained to be equal, the model of full non-invariance fit significantly better, $\chi^2_{\text{partial-2 vs. noninvariance}}(16) = 26.51, p = .047$. Consequently, a less restrictive model of partial invariance, in which item thresholds in only one class were constrained to be equal, was examined. This less restrictive partial invariance model was supported over the model of non-invariance, $\chi^2_{\text{partial-1 vs. noninvariance}}(8) = 12.28, p = .139$, and full invariance, $\chi^2_{\text{invariance vs. partial-1}}(16) = 43.58, p = .000$. Overall, this suggests that a model of partial invariance, in which item thresholds in the Low class were constrained to be equal (but allowed to vary in the Moderate and High classes) across grades 3 and 6, fit the data best. Therefore, using original scoring procedures, the structure of the latent classes in which ODD symptoms were likely to be present was found to differ across grades 3 and 6.

These analyses of measurement invariance were re-run using revised scoring for the spiteful and vindictive criterion in grade 3. The pattern of results was generally the same. However, compared to non-invariance, the model of partial invariance in which item thresholds in two classes were constrained to be equal was also supported, $\chi^2_{\text{partial-2 vs. noninvariance}}(16) = 19.28, p = .254$. This suggests that, using the revised scoring procedure for the spiteful and vindictive criterion in grade 3, a more restrictive model of partial invariance, in which item thresholds in the Low and Moderate classes were constrained to be equal (but allowed to vary in the High class) across grades 3 and 6,

was preferred. Thus, using the revised scoring procedure, only the structure of the latent class with the highest probabilities of ODD symptom endorsement was found to differ across grade 3 and 6.

Grades 6 and 9

Across grades 6 and 9, models of full invariance and partial invariance both fit better than the model of non-invariance, $\chi^2_{\text{invariance vs noninvariance}}(24) = 32.62, p = .112$, $\chi^2_{\text{partial-2 vs noninvariance}}(16) = 9.63, p = .885$. The model of partial invariance, in which item thresholds in two classes were constrained to be equal, also fit significantly better than the model of full invariance, $\chi^2_{\text{invariance vs. partial-2}}(8) = 21.59, p = .006$. Thus, across grades 6 and 9, a model of partial invariance, in which item thresholds in the High and Moderate classes were constrained to be equal (but allowed to freely vary in the Low class), fit the data best. That is, across grades 6 and 9, the structure of the latent classes in which ODD symptoms were likely to be present was found to be similar.

Grades 9 and 12

Across grades 9 and 12, models of full invariance and partial invariance both fit better than the model of non-invariance, $\chi^2_{\text{invariance vs noninvariance}}(24) = 28.43, p = .243$, $\chi^2_{\text{partial-2 vs noninvariance}}(16) = 16.07, p = .448$. Moreover, the model of full invariance was supported when compared to the model of partial invariance, $\chi^2_{\text{invariance vs. partial-2}}(8) = 12.43, p = .133$. This suggests that, across grades 9 and 12, a model of full invariance fit the data best. In other words, the structure of all three latent classes was found to be comparable across these time points.

Step 4: Covariates

In the final step, covariates were added to the latent class models, separately at grades 3, 6, 9, and 12, using starting values obtained from the respective unconditional 3-class models. As discussed previously, child sex was chosen as a potential covariate *a priori*, on theoretical grounds. In addition, a dichotomous indicator of child race (i.e., Black versus non-Black) and a continuous variable of kindergarten conduct problem risk level (i.e., total screen scores) were examined as covariates, given the significant correlations found between these variables and ODD diagnoses and criterion counts

across time in descriptive analyses. In all comparisons, the Low class served as the reference class. Notably, up to five additional participants were excluded from covariate analyses across grades, due to missing data on total screen scores.

Grade 3⁵

Five children were excluded from covariate analyses in grade 3 because they were missing data on total screen scores ($n = 383$). Child sex was not significantly associated with latent class probabilities in grade 3, so it was removed from the model. Black race was associated with a lower probability of membership in the High class (OR = 0.41, confidence interval [CI] = 0.17-0.99, $p = .007$) and the Moderate class (OR = 0.31, CI = 0.13-0.79, $p = .000$), compared to the Low class. Moreover, children with higher screen scores were significantly more likely to be in the High class (OR = 1.05, CI = 1.02-1.07, $p = .003$) and the Moderate class (OR = 1.04, CI = 1.01-1.07, $p = .024$), compared to the Low class.

Grade 6

Four children were excluded from covariate analyses in grade 6 because they were missing data on total screen scores ($n = 361$). Similar to grade 3, child sex was removed from the model because it was not significantly associated with latent class probabilities. Black race was associated with a lower probability of membership in the Moderate class (OR = 0.47, CI = 0.23-.97, $p = .010$), compared to the Low class. Total screen scores were not significantly associated with latent class probabilities in grade 6.

Grade 9

Five children were excluded from covariate analyses because they were missing data on total screen scores ($n = 330$). Child sex was not significantly associated with latent class probabilities, so it was removed from the model. Black race was associated with a lower probability of membership in the Moderate class (OR = 0.49, CI = 0.30-0.80, $p = .000$), compared to the Low class. Children with higher screen scores were

⁵ Using original scoring for the spiteful and vindictive symptom criterion.

significantly more likely to be in the High class (OR = 1.08, CI = 1.02-1.13, $p = .025$) and the Moderate class (OR = 1.03, CI = 1.01-1.05, $p = .025$), compared to the Low class.

Grade 12

Two children were excluded from covariate analyses in grade 12 because they were missing data on total screen scores ($n = 257$). Child sex was not significantly associated with latent class probabilities, so it was removed from the model. Black race was associated with a lower probability of membership in the High class (OR = 0.17, CI = 0.05-0.56, $p = .000$) and the Moderate class (OR = 0.40, CI = .17-.92, $p = .003$), compared to the Low class. Children with higher screen scores were more likely to be in the High class (OR = 1.04, CI = 1.01-1.08, $p = 0.050$), compared to the Low class.

Information regarding class structure (i.e. class membership probabilities and item-response probabilities) for final/conditional 3-class models is presented in Table 7. As can be observed in the table, the structure of the latent ODD classes across grades remained largely the same with the inclusion of these two covariates (i.e., the Black race indicator and total screen scores) in the models.

Table 7. Prevalence rates and item-response probabilities for conditional 3-class models, including race and total screen scores

	Grade 3 (<i>n</i> = 383)			Grade 6 (<i>n</i> = 361)			
	Low (72.87%)	Moderate (17.97%)	High (9.17%)	Low (69.80%)	Moderate (21.74%)	High (8.47%)	
Temper	.021	.437	.859	Temper	.051	.530	.904
Touchy	.009	.234	.840	Touchy	.020	.433	.920
Angry	.010	.157	.814	Angry	.022	.461	.926
Argues	.024	.558	.742	Argues	.087	.783	.936
Defies	.012	.425	.752	Defies	.010	.634	.873
Annoys	.000	.252	.752	Annoys	.008	.324	.742
Blames	.005	.330	.739	Blames	.040	.132	1.00
Spiteful	.005	.165	.863	Spiteful	.011	.024	.532
	Grade 9 (<i>n</i> = 330)			Grade 12 (<i>n</i> = 257)			
	Low (73.11%)	Moderate (23.58%)	High (3.31%)	Low (77.93%)	Moderate (14.53%)	High (7.55%)	
Temper	.018	.562	1.00	Temper	.005	.529	.734
Touchy	.037	.566	.832	Touchy	.000	.392	.843
Angry	.018	.432	1.00	Angry	.004	.488	.924
Argues	.060	.791	.925	Argues	.033	.580	.809
Defies	.046	.693	.921	Defies	.023	.248	.931
Annoys	.009	.355	.935	Annoys	.005	.172	.755
Blames	.000	.155	1.00	Blames	.006	.084	.503
Spiteful	.004	.052	.272	Spiteful	.000	.000	.206

Discussion

Given the growing evidence supporting a clinically meaningful distinction between the affective and behavioural symptoms of ODD (Burke et al. 2018; Frick & Nigg, 2012), researchers have recently turned to investigating whether there are subtypes of the disorder. While studies are beginning to use LCA to investigate this important question (Aebi, Barra, et al. 2016; Aebi, van Donkelaar, et al., 2016; Althoff et al., 2014; Burke, 2012; Gomez & Stavropoulos, 2018; Herzoff & Tackett, 2016; Kuny et al., 2013; Roetman et al., 2021; Wesselhoeft et al., 2019), the findings to date are inconsistent, and there is currently a lack of longitudinal research. The aim of this dissertation was to contribute to the knowledge base by investigating whether: (a) latent classes of youth could be identified based on their DSM-defined ODD symptoms at four different time points, ranging from middle childhood to late adolescence; (b) the structure of latent classes was equivalent across time; and (c) baseline demographic variables predicted latent class membership. The overarching goal of the project was to contribute to the developmental understanding of ODD and its possible presentations in young people at high risk for the disorder.

Can Latent ODD Symptom Classes Be Identified in Grades 3, 6, 9, and 12?

In the present sample, 3-class solutions fit the data best in grades 3, 6, 9, and 12. In line with previous studies in which LCA has been used to identify subgroups of youth with ODD symptoms (Aebi, Barra et al., 2016; Aebi, van Donkelaar, et al., 2016; Althoff et al., 2014; Burke, 2012; Gomez & Stavropoulos, 2018; Herzoff & Tackett, 2016 [one sample]; Kuny et al., 2013; Roetman et al., 2021; Wesselhoeft et al., 2019), two latent classes with relatively high and low probabilities of all ODD symptom criteria, respectively, were found at each time point. Although one research group did not identify a High class in a sample recruited in Southern Ontario, they did find evidence for such a subgroup in a more ethnically and economically diverse replication sample recruited in Texas (Herzoff & Tackett, 2016). Thus, the current findings add to the already strong empirical support for the existence of latent classes of youth characterized by high and low severities of ODD symptoms overall.

Across all four time points, a third class characterized by moderate probabilities of most ODD symptom criteria was also identified. This is inconsistent with the proposition that there may be subtypes of ODD, and with results from some other studies that have found latent classes characterized by relatively higher item-response probabilities for primarily irritable and/or primarily oppositional ODD symptoms (e.g., Aebi, Barra, et al., 2016; Althoff et al., 2014; Burke, 2012; Kuny et al., 2013; Herzoff & Tackett, 2016). The current findings may have been influenced by the high-risk nature of the sample, since participants were specifically recruited on the basis of displaying conduct problems in kindergarten. However, this is not the only study to find this pattern of results. Aebi, van Donkelaar, et al. (2016) and Wesselhoeft et al. (2019) identified a similar class (labeled the Moderate class and Medium class in their studies, respectively), in addition to classes with primarily affective ODD symptoms, all ODD symptoms (i.e., a High class), and no ODD symptoms (i.e., a Low class). Moreover, two recent LCA studies have found evidence for *severity-based* classifications, much like the current study (Gomez & Stavropoulos, 2018; Roetman et al., 2021). Given that the affective and behavioural symptoms of ODD have been shown to correlate highly (e.g., Burke, Boylan, et al., 2014) and to codevelop over time (e.g., Ezpeleta et al., 2021), it makes theoretical sense that latent classes distinguished by the presence of affective or behavioural symptoms have not been consistently identified.

It is worth noting that, in the present study, a 3-class solution was selected over a 4-class solution in grade 6. The 4-class model did receive some statistical support in grade 6, in the form of slightly lower AIC and aBIC values, as well as a significant BLRT result. Inspection of the resultant classes in this 4-class model indicated that something like an Irritable class (i.e., with relatively high item-response probabilities for symptoms touchy and angry) and an Oppositional class (i.e., with relatively high item-response probabilities for symptoms argues and defies) emerged, as did High and Low classes. That said, this 4-class model was not supported by the BIC or the LMR-LRT, and while the AIC was technically lower in the 4-class model, the AIC did not clearly differentiate between the 1- to 6-class solutions. Moreover, considering current theoretical, empirical, and clinical knowledge about ODD, there is no reason to expect that a 4-class solution would emerge in grade 6 (i.e., at age 11/12 years) and not in earlier or later grades. That is, in addition to having equivocal statistical support, the 4-class model was not the most parsimonious or interpretable solution in grade 6.

Casting even further doubt on the evidence for possible ODD subtypes, a closer look at some LCA studies reveals variability in how so-called Irritable and Oppositional classes have been labeled. For example, in a clinic-referred sample of boys, Burke (2012) labeled a latent class characterized by high probabilities of all symptoms the Irritability class. Also in that study, Burke's Oppositional Behaviour class had a distinctly high probability of displaying the symptom loses temper, which has been shown in some factor analytic studies to load onto the affective ODD dimension (e.g., Burke, Boylan, et al., 2014; Stringaris & Goodman, 2009a). Similarly, Kuny et al.'s (2013) Defiant class had a relatively higher probability of displaying the angry and resentful symptom (typically considered an affective ODD symptom; Aebi et al., 2010; Burke, 2010; Burke, Boylan, et al., 2014; Krieger et al., 2013; Rowe et al., 2010) than the defies or annoys symptoms. Moreover, in two separate community samples, Althoff et al. (2014) identified Defiant and Irritable classes that displayed a similarly high probability of displaying the argues symptom (typically considered a behavioural ODD symptom; e.g., Burke et al., 2010; Burke, Boylan, et al., 2014; Ezpeleta et al., 2012; Krieger et al., 2013; Rowe et al., 2010). Thus, even though Irritable and/or Oppositional latent classes have been reported, they do not look the same across LCA studies. Some of these apparent inconsistencies are likely due to the different samples and measures of ODD symptoms used. In some cases (e.g., as with the loses temper symptom in Burke, 2012), inconsistencies may also reflect the ongoing debate about item loadings in the factor analytic literature on ODD symptom dimensions (see Burke et al., 2018).

Are the Same ODD Symptom Subgroups Identified Across Time?

A major contribution of the present study was its emphasis on examining latent ODD symptom classes over a lengthy developmental period, from middle childhood (grade 3) to late adolescence (grade 12). Visual inspection of prevalence rates showed that the Low class was the largest group (69.55-78.17%) and the High class was the smallest group (3.53-9.03%) at all four time points. Visual inspection of the item-response probabilities within these classes across grades revealed some interesting patterns; namely, that (a) the probability of the spiteful and vindictive symptom being endorsed for those in the High (and to a lesser extent, Moderate) class appeared to drop after grade 3 (even when the CDISC scoring difference for this item was accounted for)

and that (b) the probability of the blames symptom being endorsed for those in the High class appeared lower in grade 12 than in earlier grades. Otherwise, the basic structure of Low, Moderate, and High classes appeared to be similar from grade 3 to grade 12.

This qualitative observation was assessed quantitatively by examining the measurement invariance of the identified 3-class models at adjacent time points. For the grades 3/6 comparison, two sets of measurement invariance analyses were run, using original and revised CDISC scoring procedures for the spiteful and vindictive symptom in grade 3. Using the original scoring procedure, a less restrictive model of partial invariance was supported, with results suggesting that only the Low class was similar across grades 3 and 6. Using the revised scoring procedure, a more restrictive model of partial invariance was supported, with results suggesting that the Low and Moderate classes were similar across grades 3 and 6. This pattern of results suggests that, when CDISC scoring differences were accounted for, only the structure of the High class changed significantly from grades 3 to 6. This may reflect the fact that frequencies of five out of eight parent-rated ODD symptoms tended to peak in grade 6 within the present sample and/or the observation that the frequency of the spiteful and vindictive symptom dropped steadily in the High class after grade 3 (see Table 1). It suggests that there may be something quantitatively different about the High class in grade 3 (at age 8/9 years) versus grade 6 (at age 11/12 years). Further investigation is needed to determine whether this quantitative difference is replicable and clinically meaningful.

Partial invariance of solutions was also found in the grades 6/9 comparison. Results indicated that the Moderate and High classes were similar across grades 6 and 9, but that there was a significant difference in the structure of the Low classes across these time points. This is a perplexing finding, given that the Low class is characterized by low probabilities of all symptoms in both grades. Similarly puzzling findings emerged in the grades 9/12 comparison. While full measurement invariance was supported across grades 9 and 12, suggesting equivalence of class structure, the estimated full invariance model appeared to substantially alter class prevalences and item-response probabilities for the 3-class solutions in these grades. It may be the case that decreases in available data compromised estimation for these later comparisons – especially since measurement invariance analyses involved running latent transition models, in which more parameters are estimated, and therefore, greater samples sizes are typically recommended (see Collins & Lanza, 2010). The particularly small prevalence of the High

class in grade 9 (3.53%) may have added to this challenge. Measurement invariance was not tested in the one published study that examined latent ODD symptom classes at different time points (Kuny et al., 2013), so the generality of the present results is not yet known. Thus, they should be interpreted with limitations in mind and replicated with larger samples.

What Predicts Subgroup Membership?

Despite evidence that ODD is somewhat more prevalent in males than females (at least in middle childhood and in Western countries; Demmer et al., 2017), child sex did not significantly predict ODD symptom class membership in the current study. These findings may have something to do with the nature of the sample, since male and female participants were recruited in kindergarten on the basis of their heightened risk for conduct problems and, thus, ODD. That said, the relatively modest sex difference found in ODD prevalence is not consistently reported beyond childhood (APA, 2013), and factor analytic studies tend to find no differences in the structure of ODD in males versus females (e.g., Ezpeleta & Penelo, 2015; Herzoff & Tackett, 2016; Lavigne et al., 2015; see Burke et al., 2018, for a summary). Moreover, of the four existing LCA studies on ODD that have directly examined this issue (Althoff et al., 2014; Herzoff & Tackett, 2016; Kuny et al., 2013; Roetman et al., 2021), three did not find child sex (or gender) to be a robust covariate in latent class models (Herzoff & Tackett, 2016; Kuny et al., 2013; Roetman et al., 2021). Therefore, the current findings regarding child sex are not aberrant or surprising.

Total screen scores and the Black/non-Black race indicator were associated with ODD symptom class membership. For grades 3 and 9, children with higher initial screen scores were more likely to be in the High and Moderate classes compared to the Low class. In grade 12, children with higher screen scores were more likely to be in the High class (vs. the Low class). This pattern of findings is generally consistent with the severity-based interpretation of classes, and suggests that the multiple-gating screening procedure designed for the Fast Track project (Lochman & CPPRG, 1995) was effective at predicting children's later development of ODD symptoms – at least for grades 3, 9, and to a somewhat lesser extent, 12. Total screen scores were not significantly associated with latent class membership in grades 6. The result for this time point is in line with initial descriptives for this sample reported above (i.e., that total screen scores

were not significantly correlated with grade 6 ODD diagnoses or criterion counts). Of note, grade 12 was the time point with the lowest frequencies of ODD diagnoses and of most symptoms (except for angry/resentful), and total screen scores were not significantly correlated with grade 12 ODD diagnoses (only criterion counts). This may partially explain why class differences in initial conduct problem risk level were less robust in grade 12.

Black race (vs. non-Black race, which mostly represented White participants) was associated with a lower probability of being in higher severity ODD symptom classes (i.e., Moderate or Moderate and High classes), compared to the Low class, across time. Although comparisons between High and Low classes with respect to child race did not reach statistical significance in grades 6 or 9, the direction of the difference was the same. These findings conflict with evidence that Black youth are more likely to be identified as having disruptive behaviour problems, particularly in school and real-world clinical practice settings (e.g., Ballentine, 2019; Bates & Glick, 2013; Fadus et al., 2020; Jones et al., 2015). The present latent classes were measured using parent report only, raising the possibility that sociocultural factors related to race (e.g., parenting style, cultural norms about child behaviour, attitudes/expectations regarding service-utilization) may have differentially influenced parent ratings of ODD symptoms (e.g., Eiraldi et al., 2006; Jones et al., 2015; McNeil et al. 2002; Washington et al., 2015). If the findings are replicated in different samples using multi-method, multi-informant assessment methods, however, they underscore the need to address potential racial bias in the diagnosis of ODD and other disruptive behaviour disorders (see Atkins-Loria et al. 2015; Ballentine, 2019; Fadus et al., 2020; Grimm et al., 2016; Jones et al., 2015; McNeil et al., 2002, for discussions relevant to this issue).

Strengths and Limitations

The current study has several notable strengths. Unlike other published LCA studies on ODD that focused on primarily White samples (e.g., Aebi, van Donkelaar, et al., 2016; Gomez & Stavropoulos, 2018) or that did not explicitly mention race (e.g., Kuny et al., 2013; Roetman et al., 2021; Wesselhoeft et al., 2019), this project utilized a racially and geographically diverse sample of children and examined the role of race in predicting latent class membership. Moreover, this study adds to the developmental understanding of ODD by examining latent classes of youth at different ages (or

developmental stages), ranging from middle childhood to late adolescence. To date, most studies have looked at a specific developmental stage (usually middle childhood) or otherwise combined data from children and adolescents of a broad range of ages into a single set of analyses. Only one published study has looked at ODD symptom subgroups over time (i.e., in a sample of children at 7, 10, and 12 years old), but this study utilized six items from a rating scale rather than the eight diagnostic criteria for ODD (Kuny et al., 2013). Relatedly, the present study utilized structured clinical interview data (i.e., from the CDISC) to measure ODD symptom criteria. This helped ensure that LCA findings were directly relevant to the primary diagnostic question at issue (i.e., whether there are so-called subtypes of ODD, distinguished by the relative likelihood of displaying specific ODD symptoms).

That said, the present study also had some limitations. First, while most existing studies have examined latent ODD classes in community or clinical samples, the current study focused on children who were recruited in kindergarten for being at high risk for developing clinically significant conduct problems. The use of a high-risk sample increased the likelihood of identifying children with ODD symptoms and, thus, of finding possible ODD subtypes; however, it also reduced the generalizability of the abovementioned findings. Second, only parent-reported ODD symptoms were used in the LCA analyses. This choice was made to (1) maximize measurement consistency across time (child report was not available in grade 3) and (2) minimize the assessment challenges that commonly arise with cross-informant disagreement (see De Los Reyes et al., 2015; Kaurin et al., 2016). However, a multi-informant measurement approach (e.g., one that also included teacher ratings and/or youth self-reports) would have provided a more comprehensive understanding of potential ODD subtypes across perspectives and contexts, so future research should explicitly examine informant issues (see Gadow & Drabick, 2012). Third, as mentioned above, the use of a binary variable to represent child race did not sufficiently capture the complex social, economic, political, and other factors that likely gave rise to the observed association between Black/non-Black race and latent ODD symptom class membership, leaving many important yet unanswered questions about those findings. Fourth, despite the benefit of examining latent ODD symptom classes across a broad developmental period, the high rate of attrition in this study reduced power to detect significant effects at all steps of the

analyses. FIML was used to attenuate this limitation, but the present findings require replication before they can be incorporated into clinical decision-making.

Conclusions and Future Directions

In conclusion, caution is warranted in assuming that evidence for ODD symptom dimensions means that there are subtypes of the disorder. Contrary to the idea that youth might be differentiated by their ODD symptom endorsement patterns, the current findings, along with those from other person-centered studies (e.g., Boylan et al., 2017; Ezpeleta et al., 2021; Gomez & Stavropoulos; Roetman et al., 2021), indicate that ODD severity may be more important. This study was the first to use LCA to investigate latent classes of youth based on their ODD symptoms at different time points ranging from middle childhood to late adolescence, and three severity-based latent classes emerged across time. When interpreted in light of the broader literature, this suggests that the symptom groupings and severity-based specifiers listed in the DSM-5 may sufficiently capture current knowledge about the presentation of ODD (APA, 2013). Clinicians should undoubtedly consider research on ODD symptom dimensions when assessing and treating youth. However, they should also understand that affective and behavioural ODD symptoms are highly correlated, because this suggests that symptoms in *both* categories are all likely part of the constellation of interrelated problems that distinguish ODD and make it an impairing disorder to greater or lesser degrees (Burke, Boylan, et al., 2014; Burke & Romano-Verthelyi, 2018). It is important to note that DSM criteria for ODD currently require at least four symptoms to be present, a specification that could not be met with affective ODD symptoms alone (APA, 2013). A severity-based conceptualization of ODD is also more consistent with the push to incorporate dimensional approaches into diagnostic classification systems (e.g., Drabick, 2009; Widiger & Samuel, 2005).

That said, much remains to be learned about how ODD presents in individuals. Although the current study did not find evidence for ODD subtypes, future research should continue to investigate this possibility in different samples, using the eight symptoms listed under diagnostic criteria for ODD (vs. available items from behaviour checklists, which have limited relevance to the diagnosis of ODD). More research with clinical samples (e.g., youth with ODD diagnoses), and at different developmental stages, is particularly needed, to establish whether latent classes suggestive of subtypes

– if identified – are (1) clinically meaningful and (2) reliably identified across time. Related to this second point, future studies should examine stability and change in ODD symptom subgroup membership (e.g., using latent transition analysis; Collins & Lanza, 2010), and researchers should start considering what threshold of longitudinal stability may be required to designate a diagnostic subtype. Given the growing recognition that ODD symptoms are not limited to early/middle childhood (e.g., Johnston et al., 2018; Johnston et al., 2020; Leadbeater & Ames, 2017; see Burke et al., 2018), it will be important to investigate these empirical questions into adolescence and possibly even adulthood (see Johnston et al., 2018). Variable-centered research has provided some pivotal insights into the structure and sequelae of ODD, but additional person-centered studies will be critical to understanding the heterogeneity of the disorder across development.

References

- Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. University of Vermont.
- Achenbach, T. M., Dumenci, L., & Rescorla, L. A. (2002). Ten-year comparisons of problems and competencies for national samples of youth: Self, parent, and teacher reports. *Journal of Emotional and Behavioral Disorders, 10*(4), 194-203. <https://doi.org/10.1177/10634266020100040101>
- Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms & Profiles*. University of Vermont Research Centre for Children, Youth, & Families.
- Aebi, M., Barra, S., Bessler, C., Steinhausen, H., Walitza, S., & Plattner, B. (2016). Oppositional defiant disorder dimensions and subtypes among detained male adolescent offenders. *Journal of Child Psychology and Psychiatry, 57*(6), 729-736. <https://doi.org/10.1111/jcpp.12473>
- Aebi, M., van Donkelaar, M. M. J., Poelmans, G., Buitelaar, J. K., Sonuga-Barke, E. J. S., Stringaris, A., Image Consortium, Faraone, S. V., Franke, B., Steinhausen, H., & van Hulzen, K. J. E. (2016). Gene-set and multivariate genome-wide association analysis of oppositional defiant behavior subtypes in attention deficit/hyperactivity disorder. *American Journal of Medical Genetics: Neuropsychiatric Genetics, 171*(5), 573-588. <https://doi.org/10.1002/ajmq.b.32346>
- Aebi, M., Muller, U. C., Asherson, P., Banachewski, T., Buitelaar, J., Ebstein, R. P., Eisenberg, J., Gill, M., Manor, I., Miranda, A., Oades, R. D., Roeyers, H., Rothenberger, A., Sergeant, J., Sonuga-Barke, E., Thompson, M., Taylor, E., Faraone, S. V., & Steinhausen, H. C. (2010). Predictability of oppositional defiant disorder and symptom dimensions in children and adolescents with ADHD combined type. *Psychological Medicine, 40*(12), 2089-2100. <https://doi.org/10.1017/S0033291710000590>
- Aebi, M., Plattner, B., Metzke, C. W., Bessler, C., & Steinhausen, H. (2013). Parent- and self-reported dimensions of oppositionality in youth: Construct validity, concurrent validity, and the prediction of criminal outcomes in adulthood. *Journal of Child Psychiatry and Psychology, 54*(9), 941-949. <https://doi.org/10.1111/jcpp.12039>
- Althoff, R. R., Kony-Slock, A. V., Verhulst, F. C., Hudziak, J. J., & van der Ende, J. (2014). Classes of oppositional defiant behavior: Concurrent and predictive validity. *Journal of Child Psychology and Psychiatry, 55*(10), 1162-1171. <https://doi.org/10.1111/jcpp.12233>
- American Psychiatric Association. (2013). *The diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Publishing.

- Atkins-Loria, S., MacDonald, H., & Mitterling, C. (2015). Young African American men and the diagnosis of conduct disorder: The neo-colonialization of suffering. *Clinical Social Work Journal*, 43(4), 431-441. <https://doi.org/10.1007/s10615-015-0531-8>
- Ballentine, K. L. (2019). Understanding racial differences in diagnosing ODD versus ADHD using critical race theory. *Families in Society: The Journal of Contemporary Social Sciences*, 100(3), 282-292. <https://doi.org/10.1177/1044389419842765>
- Bates, L. A., & Glick, J. E. (2013). Does it matter if teachers and schools match the student? Racial and ethnic disparities in problem behaviors. *Social Science Research*, 42(5), 1180-1190. <https://doi.org/10.1016/j.ssresearch.2013.04.005>
- Barkley, R. A., & Murphy, K. R. (1998). *A clinical workbook: Attention-deficit hyperactivity disorder*. Guilford.
- Boylan, K., Rowe, R., Duku, E., Waldman, I., Stepp, S., Hipwell, A., & Burke, J. (2017). Longitudinal profiles of girls' irritable, defiant, and antagonistic oppositional symptoms: Evidence for group based differences in symptom severity. *Journal of Abnormal Child Psychology*, 45, 1133-1145. <https://doi.org/10.1007/s10802-016-0231-z>
- Boylan, K., Vaillancourt, T., Boyle, M., & Szatmari, P. (2007). Comorbidity of internalizing disorders in children with oppositional defiant disorder. *European Journal of Child and Adolescent Psychiatry*, 16, 484-494. <https://doi.org/10.1007/s00787-007-0624-1>
- Burke, J. D. (2012). An affective dimension within oppositional defiant disorder symptoms among boys: Personality and psychopathology outcomes in young adulthood. *Journal of Child Psychology and Psychiatry*, 53(11), 1176-1183. <https://doi.org/10.1111/j.1469-7610.2012.02598.x>
- Burke, J. D., Boylan, K., Rowe, R., Duku, E., Stepp, S. D., Hipwell, A. E., & Waldman, I. D. (2014). Identifying the irritability dimension of ODD: Application of a modified bifactor model across five large community samples of children. *Journal of Abnormal Psychology*, 123(4), 841-851. <https://doi.org/10.1037/a0037898>
- Burke, J. D., Derella, O. J., & Johnston, O. J. (2018). Diagnostic issues in oppositional defiant disorder. In J. E. Lochman & W. Matthys, *The Wiley handbook of disruptive and impulse-control disorders* (pp. 21-36). Wiley. <https://doi.org/10.1002/9781119092254>
- Burke, J. D., Hipwell, A. E., & Loeber, R. (2010). Dimensions of oppositional defiant disorder as predictors of depression and conduct disorder in preadolescent girls. *Journal of the American Academy of Child & Adolescent Psychiatry*, 49(5), 484-492. DOI: 10.1097/00004583-201005000-00009

- Burke, J., & Loeber, L. (2010). Oppositional defiant disorder and the explanation of the comorbidity between behavioral disorders and depression. *Clinical Psychology: Science and Practice*, 17(4), 319-326. <https://doi.org/10.1111/j.1468-2850.2010.01223.x>
- Burke, J. D., Loeber, L., Lahey, B., B., & Rathouz, P. J. (2005). Developmental transitions among affective and behavioral disorders in adolescent boys. *Journal of Child Psychology and Psychiatry*, 46(11), 1200-1210. <https://doi.org/10.1111/j.1469-7610.2005.00422.x>
- Burke, J. D., & Romano-Verthelyi (2018). Oppositional defiant disorder. In M.M. Martel (Ed.), *Developmental pathways to disruptive, impulse-control, and conduct disorders* (pp. 21-52). Elsevier Science & Technology.
- Burke, J. D., Rowe, R., & Boylan, K. (2014). Functional outcomes of child and adolescent oppositional defiant disorder symptoms in young adult men. *Journal of Child Psychology and Psychiatry*, 55(3), 264-372. <https://doi.org/10.1111/jcpp.12150>
- Canino, G., Polanczyk, G., Bauermeister, J. J., Rohde, L. A., & Frick, P. J. (2010). Does the prevalence of CD and ODD vary across cultures? *Social Psychiatry and Psychiatric Epidemiology*, 45, 695-704. <https://doi.org/10.1007/s00127-010-0242-y>
- Collins, L. M., & Lanza, S. T., (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. Wiley.
- Conduct Problems Prevention Research Group (2019). *The Fast Track program for children at risk: Preventing antisocial behavior*. Guilford Press.
- Conners, C. K. (1997). *Conners Rating Scales-Revised: Technical manual*. Multi-Health Systems.
- Conners, C. K. (2001). *Conners Rating Scales-Revised*. Multi-Health Systems.
- Copeland, W. E., Shanahan, L., Costello, E. J., & Angold, A. (2009). Children and adolescent psychiatric disorders as predictors of adult disorders. *Archives of General Psychiatry*, 66(7), 764-772. <https://doi.org/10.1001/archgenpsychiatry.2009.85>
- De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, 141(4), 858-900. <https://doi.org/10.1037/a0038498>
- Demmer, D. H., Hooley, M., Sheen, J., McGillivray, J. A., & Lum, J. A. G. (2017). Sex differences in the prevalence of oppositional defiant disorder during middle childhood: A meta-analysis. *Journal of Abnormal Child Psychology*, 45(2), 313-325. <https://doi.org/10.1007/s10802-016-0170-8>

- Drabick, D. A. G. (2009). Can a developmental psychopathology perspective facilitate a paradigm shift toward a mixed categorical-dimensional classification system? *Clinical Psychology: Science and Practice*, 16(1), 41-49. <https://doi.org/10.1111/j.1468-2850.2009.01141.x>
- Drabick, D. A. G., & Gadow, K. D. (2012). Deconstructing oppositional defiant disorder: Clinic-based evidence for an anger/irritability phenotype. *Journal of the American Academy of Child & Adolescent Psychiatry*, 51(4), 384-393. <https://doi.org/10.1016/j.jaac.2012.01.010>
- Eiraldi, R. B., Mazzuca, L. B., Clarke, A. T. & Power, T. J. (2006). Service utilization among ethnic minority children with ADHD: A model of help-seeking behavior. *Administration and Policy in Mental Health and Mental Health Service Research*, 33(5), 607-622. <https://doi.org/10.1007/s10488-006-0063-1>
- Ezpeleta, L., Ganero, R., de la Osa, N., Penelo, E., & Domenech, J. M. (2012). Dimensions of oppositional defiant disorder in 3-year-old preschoolers. *Journal of Child Psychology and Psychiatry*, 53(11), 1128-1138. <https://doi.org/10.1111/j.1469-7610.2012.02545.x>
- Ezpeleta, L., & Penelo, E. (2015). Measurement invariance of oppositional defiant disorder dimensions in 3-year-old preschoolers. *European Journal of Psychological Assessment*, 31(1), 45-53. <https://doi.org/10.1027/1015-5759/a000205>
- Ezpeleta, L., Penelo, E., Navarro, J. B., de la Osa, N., Trepato, E. (2021). Co-developmental trajectories of defiant/headstrong, irritability, and prosocial emotions from preschool age to early adolescence. *Child Psychiatry & Human Development*. <https://doi.org/10.1007/s10578-021-01180-z>
- Fadus, M. C., Ginsburg, K. R., Sobowale, K., Halliday-Boykins, C. A., Bryant, B. E., Gray, K. M., & Squeglia, L. M. (2020). Unconscious bias and the diagnosis of disruptive behavior disorders and ADHD in African American and Hispanic Youth. *Academic Psychiatry*, 44(1), 95-102. <https://doi.org/10.1007/s40596-019-01127-6>
- Frick, P. J., & Nigg, J. T. (2012). Current issues in the diagnosis of attention deficit hyperactivity disorder, oppositional defiant disorder, and conduct disorder. *Annual Review of Clinical Psychology*, 8, 77-107. <https://doi.org/10.1146/annurev-clinpsy-032511-143150>
- Gadow, K. D., & Drabick, D. A. G. (2012). Anger and irritability symptoms among youth with ODD: Cross-informant versus source-exclusive syndromes. *Journal of Abnormal Child Psychology*, 40(7), 1073-1085. <https://doi.org/10.1007/s10802-012-9637-4>
- Gomez, R., & Stavropoulos, V. (2018). Maternal ratings of ODD symptoms: Subtypes versus severity in a general community sample of children. *Comprehensive Psychiatry*, 81, 81-90. <https://doi.org/10.1016/j.comppsy.2017.12.001>

- Goodman, R., Ford, T., Richards, H., Gatward, R., & Meltzer, H. (2000). The Development and Well-Being Assessment: Description and initial validation of an integrated assessment of child and adolescent psychopathology. *Journal of Child Psychology and Psychiatry*, 41(5), 645-655. <https://doi.org/10.1111/j.1469-7610.2000.tb02345.x>
- Grimmett, M. A., Dunbar, A. S., Williams, T., Clark, C., Prioleau, B., & Miller, J. S. (2016). The process and implications of diagnosing oppositional defiant disorder in African American males. *The Professional Counselor*, 2, 147-160. <https://doi.org/10.15241/MG.6.2.147>
- Herzoff, K., & Tackett, J. L. (2016). Subfactors of oppositional defiant disorder: Converging evidence from structural and latent class analysis. *Journal of Child Psychology and Psychiatry*, 57(1), 18-29. <https://doi.org/10.1111/jcpp.12423>
- Hofstra, M. B., Van Der Ende, J. A. N., & Verhulst, F. C. (2001). Adolescents' self-reported problems as predictors of psychopathology in adulthood: 10-year follow up study. *The British Journal of Psychiatry*, 179, 203-209. <https://doi.org/10.1192/bjp.179.3.203>
- IBM Corp. (2016). *IBM Statistics for Windows, Version 24.0*. Armonk, NY: Author.
- Johnston, O. G., Cruess, D. G., & Burke, J. D. (2020). Irritability and behavioral symptom dimensions of oppositional defiant disorder in young adults: Associations with DSM-5 pathological personality traits. *Journal of Psychopathology and Behavioral Assessment*, 42(9), 424-435. <https://doi.org/10.1007/s10862-020-09786-8>
- Johnston, O. G., Derella, O. J., & Burke, J. D. (2018). Identification of oppositional defiant disorder in young adult college students. *Journal of Psychopathology and Behavioral Assessment*, 40(4), 563-572. <https://doi.org/10.1007/s10862-018-9696-0>
- Jones, S. C. T., Neblett, E. W., Gaskin, A. L. & Lee, D. B. (2015). Assessing the African American child and adolescent: Special considerations and assessment of behavioural disorders. In L. T. Benuto & B. D. Leany (Eds.), *Guide to psychological assessment with African Americans* (pp. 105-120). Springer. https://doi.org/10.1007/978-1-4939-1004-5_8
- Kaurin, A., Egloff, B., Stringaris, A., & Wessa, M. (2016). Only complementary voices tell the truth: A reevaluation of validity of multi-informant approaches of child and adolescent clinical assessments. *Journal of Neural Transmission*, 123(8), 981-990. <https://doi.org/10.1007/s00702-016-1543-4>
- Kimonis, E. R., Frick, P. J., & McMahon, R. J. (2014). Conduct and oppositional defiant disorders. In E. J. Mash & R. A. Barkley (Eds.), *Child psychopathology* (3rd ed., pp. 145-179). Guilford Press.

- Krieger, F. V., Polanczyk, V. G., Goodman, R., Rohde, L. A., Graeff-Martins, A. S., Salum, G., Gadelha, A., Pan, P., Stahl, D., & Stringaris, A. (2013). Dimensions of oppositionality in a Brazilian community sample: Testing the DSM-5 proposal and etiological links. *Journal of the American Academy of Child & Adolescent Psychiatry*, 52(4), 389-400. <http://doi.org/10.1016/j.jaac.2013.01.004>
- Kuny, A. V., Althoff, R. R., Copeland, W., Bartels, M., Bejsterveldt, V., Baer, J., & Hudziak, J. J. (2013). Separating the domains of oppositional defiant behavior: Comparing latent models of the Conners' Oppositional Scale. *Journal of the American Academy of Child & Adolescent Psychiatry*, 52(2), 172-183. <https://doi.org/10.1016/j.jaac.2012.10.005>
- Lanza, S. T., & Cooper, B. R. (2016). Latent class analysis for developmental research. *Child Development Perspectives*, 10(1), 59-64. <https://doi.org/10.1111/cdep.12163>
- Laursen, B., & Hoff, E. (2006). Person-centered and variable-centered approaches to longitudinal data. *Merrill-Palmer Quarterly*, 52(3), 377-389. <https://doi.org/10.1353/mpq.2006.0029>
- Lavigne, J. V., Bryant, F. B., Hopkins, J., & Gouze, K. R. (2015). Dimensions of oppositional defiant disorder in young children: Model comparisons, gender and longitudinal invariance. *Journal of Abnormal Child Psychology*, 43(3), 423-439. <https://doi.org/10.1007/s10802-014-9919-0>
- Leadbeater, B. J., & Ames, M. E. (2017). The longitudinal effects of oppositional defiant disorder symptoms on academic and occupational functioning in the transition to young adulthood. *Journal of Abnormal Child Psychology*, 45(4), 749-763. <https://doi.org/10.1007/s10802-016-0190-4>
- Leadbeater, B. J., & Homel, J. (2015). Irritable and defiant sub-dimensions of ODD: Their stability and prediction of internalizing symptoms and conduct problems from adolescence to young adulthood. *Journal of Abnormal Child Psychology*, 43(3), 407-421. <https://doi.org/10.1007/s10802-014-9908-3>
- Leadbeater, B., Thompson, K., & Gruposso, V. (2012). Co-occurring trajectories of symptoms of anxiety, depression, and oppositional defiance from adolescence to young adulthood. *Journal of Clinical Child & Adolescent Psychology*, 41(6), 719-730. <https://doi.org/10.1080/15374416.2012.694608>
- Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88(3), 767-778. <https://doi.org/10.1093/biomet/88.3.767>
- Lochman, J. E., & Conduct Problems Prevention Research Group. (1995). Screening of child behavior problems for prevention programs at school entry. *Journal of Consulting and Clinical Psychology*, 63(4), 549-559. <https://doi.org/10.1037//0022-006x.63.4.549>

- McNeil, C. B., Capage, L. C., & Bennett, G. M. (2002). Cultural issues in the treatment of young African American children diagnosed with disruptive behavior disorders. *Journal of Pediatric Psychology, 27*(4), 339-350. <https://doi.org/10.1093/jpepsy/27.4.339>
- Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus users guide* (8th ed.). Authors.
- Muthén, L. K., & Muthén, B. O. (no date). Chi-squared difference testing using the Satorra-Bentler Scales Chi-Square. Retrieved from: <http://www.statmodel.com/chidiff.shtml>
- Nock, M. K., Kazdin, A. E., Hiripi, E., & Kessler, R. C. (2007). Lifetime prevalence, correlates, and persistence of oppositional defiant disorder: Results from the National Comorbidity Survey Replication. *Journal of Child Psychology and Psychiatry, 48*(7), 703-713. <https://doi.org/10.1111/j.1469-7610.2007.01733.x>
- Nylund, K. L. (2007). *Latent transition analysis: Modeling extensions and an application to peer victimization*. (Unpublished doctoral dissertation). University of California, Los Angeles.
- 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, 14*(4), 535-569. <https://doi.org/10.1080/10705510701575396>
- Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science, 4*(4), 440-461. <https://doi.org/10.1037/tps0000176>
- Pardini, D. A., & Fite, P. J. (2010). Symptoms of conduct disorder, oppositional defiant disorder, attention-deficit/hyperactivity disorder and callous-unemotional traits as unique predictors of psychosocial adjustment in boys: Advancing an evidence base for DSM-V. *Journal of the American Academy of Child & Adolescent Psychiatry, 49*(11), 1134-1144. <https://doi.org/10.1016/j.jaac.2010.07.010>
- Quay, H. C., & Peterson, D. R. (1987). *Manual for the Revised Behavior Problem Checklist Report*. University of Miami.
- Roetman, P. J., Siebelink, B. M., Vermeiren, R. R. J. M., & Colins, O. F. (2021). Classes of oppositional defiant disorder behavior in clinic-referred children and adolescents: Concurrent features and outcomes. *The Canadian Journal of Psychiatry, 66*(7), 657-666. <https://doi.org/10.1177/0706743720974840>
- Rowe, R., Costello, J., Angold, A., Copeland, W. E., & Maughan, B. (2010). Developmental pathways in oppositional defiant disorder and conduct disorder. *Journal of Abnormal Psychology, 119*(4), 726-738. <https://doi.org/10.1037/a0020798>

- Satorra, A., & Bentler, P. M (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Theory and Methods*, 75(2), 243-248.
<https://doi.org/10.1007/s11336-009-9135-y>
- Schwab-Stone, M. E., Shaffer, D., Dulcan, M. K., Jensen, P S., Fisher, P., Bird, H. R., Goodman, S. H., Lahey, B. B., Lichtman, J. H., Canino, G., Rubio-Stipec, M., & Rae, D. S. (1996). Criterion validity for the NIMH Diagnostic Interview Schedule for Children Version 2.3 (DISC 2.3). *Journal of the Academy of Child & Adolescent Psychiatry*, 35(7), 878-888. <https://doi.org/10.1097/00004583-199607000-00013>
- Shaffer, D., & Fisher, P. (1991). *NIMH - Diagnostic Interview Schedule for Children: Parent Informant (Interview about Child)*. New York State Psychiatric Institute.
- Shaffer, D., & Fisher, P. (1997). *NIMH - Diagnostic Interview Schedule for Children: Parent Informant*. New York State Psychiatric Institute.
- Shaffer, D., Fisher, P., Dulcan, M. K., Davies, M., Piacentini, J., Schwab-Stone, M. E., Lahey, B. E., Bourdon, K., Jensen, P. S., Bird, H. R., Canino, G., & Regier, D. A. (1996). The NIMH Diagnostic Interview Schedule for Children Version 2.3 (DISC 2.3): Description, acceptability, prevalence rates, and performance in the MECA study. *Journal of the Academy of Child & Adolescent Psychiatry*, 35(7), 865-877.
<https://doi.org/10.1097/00004583-199607000-00012>
- Shaffer, D., Fisher, P., Lucas, C., Dulcan, M. K., & Schwab-Stone, M. E. (2000). NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): Description, differences from previous version, and reliability of some common diagnoses. *Journal of the American Academy of Child & Adolescent Psychiatry*, 39(1), 28-38. <https://doi.org/10.1097/00004583-200001000-00014>
- Stringaris, A., & Goodman, R. (2009a). Three dimensions of oppositionality in youth. *Journal of Child Psychology and Psychiatry*, 50(3), 216-223.
<https://doi.org/10.1111/j.1469-7610.2008.01989.x>
- Stringaris, A., & Goodman, R. (2009b). Longitudinal outcome of youth oppositionality: Irritable, headstrong, and hurtful behaviors have distinctive predictions. *Journal of the American Academy of Child & Adolescent Psychiatry*, 48(4), 404-412.
<https://doi.org/10.1097/CHI.0b013e3181984f30>
- Washington, T., Rose, T., Colombo, G., Sung Hong, J., & Irby Coard, S. (2015). Family-level factors and African American children's behavioral health outcomes: A systematic review. *Child Youth Care Forum*, 44, 819-834.
<https://doi.org/10.1007/s10566-015-9308-z>
- Werthamer-Larsson, L., Kellam, S. G., & Wheeler, L. (1991). Effects of first-grade classroom environment on shy behavior, aggressive behavior, and concentration problems. *American Journal of Community Psychology*, 19(4), 585-602.
<https://doi.org/10.1007/BF00937993>

- Wesselhoeft, R., Stringaris, A., Sibbersen, C., Kristensen, R. V., Bojesen, A. B., & Talati, A. (2019). Dimensions and subtypes of oppositionality and their relation to comorbidity and psychosocial characteristics. *European Child & Adolescent Psychiatry*, 28(3), 351-365. <https://doi.org/10.1007/s00787-018-1199-8>
- Whelan, Y. M., Stringaris, A., Maughan, B., & Barker, E. D. (2013). Developmental continuity of oppositional defiant disorder subdimensions at ages 8, 10, and 13 years and their distinct psychiatric outcomes at age 16 years. *Journal of the American Academy of Child & Adolescent Psychiatry*, 52(9), 961-969. <https://doi.org/10.1016/j.jaac.2013.06.013>
- Widiger, T. A., & Samuel, D. B. (2005). Diagnostic categories or dimensions? A question for the Diagnostic and Statistical Manual of Mental Disorders – Fifth Edition. *Journal of Abnormal Psychology*, 114(4), 494-504. <https://doi.org/10.1037/0021-843X.114.4.494>