Categorical and Dimensional Concepts of Externalizing Disorders in Children

Inaugural Dissertation

zur

Erlangung des Doktorgrades
philosophiae doctor (PhD) in Health Sciences
der Medizinischen Fakultät
der Universität zu Köln

vorgelegt von

Ann-Kathrin Thöne

aus Bremen

A&A Digital Print Center | Copyshop Bonn | Bonn

2024

Betreuer*in:	Prof. Dr. Manfred Döpfner
Gutachter*in:	Prof. Dr. Dr. Kai Vogeley
	Prof. Dr. Christian Albus
	Prof. Dr. Charlotte Hanisch

Datum der Mündlichen Prüfung: 28.05.2024

"But, though no man can draw a stroke between the confines of day and night, yet light and darkness are upon the whole tolerably distinguishable"
(Burke, 1770/1963, p. 147).

"What we call a beginning is often the end
And to make an end is to make a beginning.
The end is where we start from. (...)
We shall not cease from exploration
And the end of all our exploring
Will be to arrive where we started
And know the place for the first time"
(Eliot, 1942, section V).

Table of Contents

Acknowledgments 1

List of abbreviations 2

Summary 4

Zusammenfassung 6

1 Introduction 8

1.1 Advantages and disadvantages of categorical and dimensional concepts 9

1.2 Dimensional assessment instruments of child and adolescent psychopathology 12

1.2.1 Achenbach System of Empirically Based Assessment (ASEBA) 12

1.2.2 Diagnostic System for Mental Disorders in Children and Adolescents based on the ICD-10 and DSM-5 (DISYPS-III) 13

1.3 Multi-informant approach in the assessment of child and adolescent mental disorders 13

1.4 Hierarchical systems of categorical and dimensional classification 14

1.5 Latent factor models of psychopathology 16

1.6 A network approach to psychopathology 18

1.7 Structured clinical interviews for child and adolescent mental disorders 21

1.8 Development and description of the clinical parent interviews DISYPS-ILF 23

1.8.1 DISYPS ILF-EXTERNAL 24

1.9 Interim summary 24

2 Thesis objectives 26

2.1 Study description 27

3 Toward a dimensional assessment of externalizing disorders in children: Reliability and validity of a semi-structured parent interview 29

4 Disentangling symptoms of externalizing disorders in children using multiple measures and informants 60

5 Identifying symptoms of ADHD and disruptive behavior disorders most strongly associated with functional impairment in children: A symptom-level approach 86

6 Discussion 113

6.1 Summary and clinical implications 113

6.2 Limitations 117

6.3 Challenges and future directions 119

6.4 Conclusion 123

6.5 References 125

7 Appendix 147

7.1 Supplementary material for: 147

Toward a dimensional assessment of externalizing disorders in children: Reliability and validity of a semi-structured parent interview 147

7.2 Supplementary material for: 151

Disentangling symptoms of externalizing disorders in children using multiple measures and informants 151

7.3 Supplementary material for: 175

Identifying symptoms of ADHD and disruptive behavior disorders most strongly associated with functional impairment in children: A symptom-level approach. 175

7.4 Curriculum Vitae 195

7.5 Erklärung 198

Acknowledgments

First and foremost, I would like to express my deep gratitude to my supervisor Prof. Dr. Manfred Döpfner for his comprehensive and competent guidance. Also, I would like to express my sincere gratitude to Prof. Dr. Anja Görtz-Dorten for her encouraging and competent support. Their mentorship, as well as their great opportunity for me to contribute to the publication of DISYPS-ILF, has been vital to my growth in research.

My sincere thanks also goes to Prof. Dr. Christian Albus and Prof. Dr. Dr. Kai Vogeley for their inspiring discussions during our annual tutor meetings, which encouraged me to widen my research from various perspectives.

I would also like to thank all members of the ESCAschool consortium who were involved in study conception, coordination, organization and management, recruitment, blind rating, diagnostics, and therapy. Without the ESCAschool multicenter trial, this dissertation would not have been possible.

Many thanks also go Dr. Michaela Junghänel for her emotional and practical support during our joint journey to completing our dissertations and for the many hours we spent together discussing statistical results and searching the Mplus forum trying to find solutions to our statistical problems.

Special thanks to Dr. Christina Dose who has provided valuable support in data interpretation and for her comprehensive and constructive revisions of my three articles of this cumulative dissertation.

I also wish to thank my colleagues and friends from AKiP Cologne for providing support, advice, and discussion when needed.

My friends and family members have given me encouragement over the years. I would particularly like to thank my partner Dr. Niclas Braun who has provided huge support in all kinds of ways through all the times we have experienced together. I am very much looking forward to the next adventures together with you.

Finally, I would also like to thank all participants and families who volunteer for research studies at AKiP Cologne. Without the gift of their time and energy, these studies would not be possible.

List of abbreviations

ADHD	Attention-deficit/hyperactivity disorder
AIC	Akaike information criterion
APA	American Psychiatric Association
ASEBA	Achenbach System of Empirically Based Assessment
BIC	Bayesian information criterion
CAPA	Child and Adolescent Psychiatric Assessment
CBCL	Parent's Report Form on Child and Adolescent Behavior
CD	Conduct disorder
CFA	Confirmatory factor analysis
CFI	Comparative fit index
ChIPS	Children's Interview for Psychiatric Symptoms
CU	Callous unemotional
DCL	Diagnostic checklist
DICA	Diagnostic Interview for Children and Adolescents
DISYPS-III	Diagnostic System for Mental Disorders in Children and Adolescents based on the ICD-10 and DSM-5
DISYPS-ILF	Clinical Parent Interviews According to the DSM-5 Diagnostic System of Mental Disorders for Children and Adolescents
DMDD	Disruptive mood dysregulation disorder
DSM	Diagnostic and Statistical Manual of Mental Disorders
EBICglasso	Extended Bayesian information criteria for gaussian graphical models
ECV	Explained common variance
EFA	Exploratory factor analysis
ESCA	Evidence-Based Stepped Care of ADHD
ESEM	Exploratory structural equation modeling
FBB	Proxy-rating form
FI	Functional impairment
g-factor	General factor
GGM	Gaussian graphical model
HI	Hyperactivity-impulsivity
HiTOP	Hierarchical Taxonomy of Psychopathology
ICC	Intraclass correlation
ICD	International Classification of Mental Disorders

ILF-EXTERNAL	Interview for Externalizing Disorders
ILF-INTERNAL	Interview for Internalizing Disorders
ILF-KONTAKT	Interview for Contact Disorders
ILF-SCREEN	Interview for Screening of Mental Disorders
ILF-ZWANG/TIC	Interview for Obsessive-Compulsive and Tic Disorders
IN	Inattention
IRR	Interrater reliability
K-SADS	Schedule for Affective Disorders and Schizophrenia for School Aged Children
Mini-KID	Mini-International Neuropsychiatric Interview for Children and Adolescents
MLR	Maximum likelihood estimation with robust standard errors
NIMH DISC-IV	NIMH Diagnostic Interview Schedule for Children Version IV
ODD	Oppositional defiant disorder
ODD-AD	Oppositionality with chronic irritability/anger
RDoC	Research Domain Criteria
RMSEA	Root mean square error of approximation
s-factor	Specific factor
SBB	Self-rating form
SCICA	Semi-structured Clinical Interview for Children and Adolescents
SRMR	Standardized root mean square residual
TLI	Tucker-Lewis index
TRF	Teacher’s Report Form on Child and Adolescent Behavior
WLSMV	Weighted least squares means and variance adjusted
YSR	Youth Self-Report Form

Summary

One of the most difficult challenges faced by researchers and clinicians is how to best conceptualize the classification of mental disorders. Can we assume that there is a clear distinction in which an individual is either ‘normal’ (i.e., not meeting diagnostic criteria) or ‘abnormal’ (i.e., meeting diagnostic criteria)? And should we continue to endorse this black-and-white mindset, even though reality presents us with various shades of gray? Or should we not rather assume that the frequency and severity of psychological symptoms vary continuously across the full range of each dimension, without any distinct, meaningful binary threshold between 'having' or 'not having' a mental disorder? These questions have sparked an ongoing debate regarding whether it is more appropriate to conceptualize mental disorders using categorical or dimensional approaches– a debate that holds significant implications for scientific research and clinical practice.

Building upon this debate, the main objective of this doctoral dissertation is to contribute to establishing a rigorous standard for diagnosing mental disorders in childhood and adolescence. Having developed the DISYPS-ILF, an extensive set of clinical parent interviews for diagnosing mental disorders, this dissertation provides a comprehensive psychometric evaluation of the *Interview for Externalizing Disorders* (ILF-EXTERNAL). The ILF-EXTERNAL covers diagnostic criteria for attention-deficit/hyperactivity disorder and disruptive behavior disorders. One major advantage of the ILF-EXTERNAL compared to existing clinical interviews is that it allows both a categorical assessment and a dimensional characterization, thereby integrating the strengths of both diagnostic approaches. Participant data (N = 474) for analyses of this dissertation were obtained from the ESCAschool (*Evidence-based, Stepped Care of ADHD in school-aged children*) multicenter trial.

This doctoral dissertation is designed as a cumulative work and comprises three empirical studies. The first study (Thöne et al., 2020) assesses the psychometric properties of the ILF-EXTERNAL. Overall, the ILF-EXTERNAL displays sound psychometric properties in terms of validity and reliability. Furthermore, diagnostic agreement between clinicians is generally higher at the dimensional level compared to categorical diagnoses. The second study (Thöne et al., 2021) evaluates the factorial structure of the ILF-EXTERNAL and systematically tests the underlying latent dimensions of externalizing symptoms using confirmatory factor analyses and exploratory structural equation modeling. Specifically, results demonstrate that a novel bifactor S-1 factor model displays a statistically sound factor structure and allows for meaningful interpretation. Besides, results from measurement invariance analyses highlight meaningful cross-informant discrepancies, that is, different informants (i.e., clinicians, parents,

teachers) provide unique perspectives how children's behavior varies across settings, such as at school or at home. The third study (Thöne et al., 2023) zooms in on the symptom-level and enhances the understanding of how individual symptoms differentially relate to psychological distress and functional impairments. Overall, results from the multivariate regression analyses demonstrate significant variations in the associations between symptoms and global functional impairment. Furthermore, results from network analyses reveal a more nuanced yet multifaceted perspective on the associations between individual symptom-impairment relations compared to factor-analytic techniques.

In conclusion, findings from this doctoral dissertation, especially the comparison of latent factor and network models, demonstrate that there is no simple or comprehensive solution to the categorical versus dimensional debate from an empirical perspective. Both categorical and dimensional approaches have their merits and combining both approaches may help to compensate for each other's shortcomings. These findings, whilst intriguing, represent only the beginning of the journey and there is still a need to investigate categorical and dimensional concepts as well as recently proposed hybrid systems of psychopathology in greater detail. In the future, there needs to be further focus on the clinical utility of empirically identified dimensions, their clinical implications regarding the selection of appropriate treatments and adaptive interventions, and systematic research on patient outcomes when adopting dimensional instead of categorical measures of psychopathology.

Zusammenfassung

Eine der schwierigsten Herausforderungen, mit denen Forscher*innen und Kliniker*innen konfrontiert werden, ist die Frage, wie sich die Klassifikation psychischer Störungen am besten konzeptualisieren lässt. Können wir davon ausgehen, dass es eine klare Diskontinuität gibt, bei der eine Person entweder ‘normal’ ist (d. h. die Diagnosekriterien sind nicht erfüllt) oder ‘abnormal’ (d. h. die Diagnosekriterien sind erfüllt)? Und können wir dieses Schwarz-Weiß-Denken noch vertreten, obwohl es in der Realität nichts als verschiedene Grautöne gibt? Oder sollten wir nicht vielmehr davon ausgehen, dass die Häufigkeit und der Schweregrad psychischer Symptome über die gesamte Bandbreite der einzelnen Dimensionen hinweg kontinuierlich variieren, ohne dass es eine natürliche oder sinnvolle binäre Grenze zwischen dem Vorliegen einer psychischen Störung und dem Nichtvorliegen einer solchen gibt? Diese Fragen haben zu einer anhaltenden Debatte darüber geführt, ob psychische Störungen am besten anhand kategorischer oder dimensionaler Ansätze konzeptualisiert werden sollten – eine Debatte, die tiefgreifende Auswirkungen auf die klinische Praxis und die wissenschaftliche Forschung hat.

Aufbauend auf dieser Debatte besteht das Hauptziel der vorliegenden Dissertation darin, einen qualitativ hochwertigen Beitrag für die Diagnose psychischer Störungen bei Kindern und Jugendlichen zu leisten. Nach der Entwicklung des DISYPS-ILF, eines umfassenden Sets klinischer Elterninterviews zur Diagnostik psychischer Störungen, liefert diese Dissertation eine umfassende psychometrische Evaluation des *Interview-Leitfadens für externale Störungen* (ILF-EXTERNAL). Der ILF-EXTERNAL deckt die diagnostischen Kriterien für Aufmerksamkeitsdefizit-/Hyperaktivitätsstörungen und Störungen des Sozialverhaltens ab. Ein großer Vorteil des ILF-EXTERNAL im Vergleich zu bestehenden klinischen Interviews ist, dass er sowohl eine kategoriale Beurteilung als auch eine dimensionale Charakterisierung ermöglicht und damit die Stärken beider diagnostischer Ansätze vereint. Die Daten (N = 474) für die vorliegenden Analysen stammen aus der multizentrischen Studie ESCAschool (*Evidenz-basierte, stufenweise Versorgung von ADHS bei Schulkindern*).

Die vorliegende Dissertation ist als kumulative Arbeit konzipiert und umfasst drei empirische Studien. In der ersten Studie (Thöne et al., 2020) werden die psychometrischen Eigenschaften des ILF-EXTERNAL untersucht. Insgesamt zeigt der ILF-EXTERNAL solide psychometrische Eigenschaften in Bezug auf Reliabilität und Validität. Darüber hinaus ist die diagnostische Übereinstimmung zwischen Kliniker*innen im Allgemeinen auf dimensionaler Ebene höher als bei kategorialen Diagnosen. Die zweite Studie (Thöne et al., 2021) evaluiert die faktorielle Struktur des ILF-EXTERNAL und testet systematisch die zugrundeliegenden

latenten Dimensionen externaler Symptome mittels konfirmatorischer Faktorenanalysen und explorativen Strukturgleichungsmodellen. Die Ergebnisse zeigen insbesondere, dass ein neuartiges Bifaktor S-1-Faktorenmodell eine statistisch solide Faktorenstruktur aufweist und eine sinnvolle Interpretation ermöglicht. Darüber hinaus zeigen die Ergebnisse der Messinvarianzanalysen bedeutsame Diskrepanzen zwischen den Informant*innen auf, d.h. verschiedene Informant*innen (d. h. Kliniker*innen, Eltern, Lehrer*innen) liefern einzigartige Perspektiven, wie sich das Verhalten von Kindern in verschiedenen Umgebungen unterscheidet, so wie in der Schule oder zu Hause. Die dritte Studie (Thöne et al., 2023) fokussiert sich auf die Symptomebene und verbessert das Verständnis dafür, wie einzelne Symptome in unterschiedlicher Weise mit psychischem Leidensdruck und funktionellen Beeinträchtigungen in Zusammenhang stehen. Insgesamt zeigen die Ergebnisse der multivariaten Regressionsanalysen, dass die Symptome in ihrem Zusammenhang mit der globalen Funktionsbeeinträchtigung sehr unterschiedlich sind. Darüber hinaus ergeben die Ergebnisse der Netzwerkanalysen im Vergleich zu faktorenanalytischen Verfahren eine nuanciertere und komplexere Perspektive auf die Zusammenhänge zwischen einzelnen Symptomen und Funktionsbeeinträchtigungen.

Zusammenfassend zeigen die Ergebnisse dieser Dissertation, insbesondere der Vergleich von latenten Faktoren- und Netzwerkmodellen, dass es aus empirischer Sicht keine einfache oder umfassende Lösung für die kategoriale vs. dimensionale Debatte gibt. Sowohl der kategoriale als auch der dimensionale Ansatz haben je ihre Vorzüge, und die Kombination beider Ansätze kann dazu beitragen, die Schwächen des jeweils anderen zu kompensieren. Auch wenn diese Ergebnisse faszinierend sind, so stellen sie nur den Anfang der Reise dar, und es besteht weiterhin die Notwendigkeit, kategoriale und dimensionale Konzepte sowie die kürzlich vorgeschlagenen hybriden Systeme der Psychopathologie eingehender zu untersuchen. In Zukunft sollte der klinische Nutzen empirisch ermittelter Dimensionen, ihre klinischen Implikationen für die Auswahl geeigneter Behandlungen und adaptiver Interventionen sowie die systematische Erforschung der Ergebnisse für die Patient*innen bei der Anwendung dimensionaler statt kategorialer Maße der Psychopathologie weiter in den Mittelpunkt gerückt werden.

1 Introduction

The classification of mental disorders distinguishes between categorical and dimensional concepts as two different approaches to understanding and describing psychopathology (Achenbach, 1966, 2020; Brown & Barlow, 2005; Clark et al., 2017; Coghill & Sonuga-Barke, 2012; Döpfner, 2022; Döpfner & Lehmkuhl, 1997; Döpfner & Petermann, 2012; Kotov et al., 2017; Krueger et al., 2018; Lahey et al., 2022). The two major clinical classification systems, the World Health Organization's International Classification of Mental Disorders (ICD) in its 10th and 11th versions (World Health Organization, 1993, 2019) and the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (DSM) in its 5th and 5th text revised versions (American Psychiatric Association, 2013, 2022), are committed to the categorical approach. According to the categorical approach, mental disorders can be regarded as distinct, independent, and categorical phenomena (Clark et al., 2017; Krueger et al., 2018; Krueger & Eaton, 2015). That is, with few exceptions, *"the classification systems imply that a patient either meets the diagnostic threshold for a particular mental disorder or does not (categorical), the disorder does not overlap with other disorders (distinct), and therefore presence of the disorder should not necessarily be associated with a higher probability of having another disorder (independent)"* (Krueger & Eaton, 2015, p. 27). Traditionally, the ICD and DSM were developed based on expert consensus among clinicians appointed to specific workgroups, while the more recent versions increasingly consider empirical research on the organization and nature of psychopathology (Clark et al., 2017; Coghill & Sonuga-Barke, 2012). In categorical systems, the classification criteria for all mental disorders refer to the presentations of observable behaviors (signs) and internally experienced feelings and thoughts (symptoms), often additionally requiring that a certain threshold of a minimum of signs and symptoms from a list of several must be met (American Psychiatric Association, 2013; World Health Organization, 1993). Besides, a mental disorder is typically expected to be accompanied by clinically significant psychological distress or functional impairment (FI) in domains such as social, educational, or occupational functioning (Palermo et al., 2008; Spitzer & Wakefield, 1999; Wakefield, 2007). Further diagnostic criteria regarding age of onset, duration and course of symptomatology, or exclusion criteria in the context of differential diagnostic assessment are usually specified as well (American Psychiatric Association, 2013; World Health Organization, 1993). According to the dimensional approach, in comparison, mental disorders are not all-or-none phenomena. Rather, mental disorders are continuously distributed along several dimensions which are graded in severity, ranging from absence to severe psychopathology (Clark et al., 2017; Coghill & Sonuga-Barke, 2012; Döpfner, 2022; Döpfner

& Petermann, 2012; Lahey et al., 2022). Indeed, meta-analyses of taxonomic research provide empirical support that most mental disorders are probably better described as continuously distributed dimensions rather than discrete disorder categories (Haslam et al., 2012, 2020).

1.1 Advantages and disadvantages of categorical and dimensional concepts

Few topics generate as much division among researchers and clinicians in our field as the debate over whether mental disorders should be conceptualized categorically or dimensionally (Achenbach, 2020; Brown & Barlow, 2005; Clark et al., 2017; Coghill & Sonuga-Barke, 2012). Both diagnostics concepts are associated with several advantages and disadvantages, which will now be briefly reviewed (Döpfner, 2022; Döpfner & Lehmkuhl, 1997; Döpfner & Petermann, 2012; Kotov et al., 2017; Lahey et al., 2022).

Thresholds: Categorical classification systems require determining cut-off values that allow individuals to be assigned to discrete diagnostic classes (Döpfner & Lehmkuhl, 1997; Döpfner & Petermann, 2012; Lahey et al., 2022). Accordingly, the categorical classification postulates a clear threshold, sometimes also a qualitative jump, between 'health' and 'disorder' of mental phenomena (Clark et al., 2017; Coghill & Sonuga-Barke, 2012; Döpfner & Petermann, 2012). For polythetically defined categories (e.g., attention-deficit/hyperactivity disorder [ADHD]), categorical classification systems specify a certain number of symptoms that must be met for a categorical diagnosis. This number was ultimately defined by expert consensus based on symptom prevalence rates in the population (e.g., Lahey et al., 1990). Recent research, however, highlights that additional external criteria, such as the degree of FI and psychological distress might be helpful in empirically determining thresholds (Arildskov et al., 2022; DuPaul, 2022). While the level of mental illness severity is a fundamental aspect, dimensionality itself does not seem to present conceptual difficulties (Clark et al., 2017; Döpfner, 2022; Döpfner & Petermann, 2012). For example, the intellectual ability continuum can be categorized into semi-arbitrary yet beneficial groupings encompassing severe intellectual disability to genius (Clark et al., 2017). Considering their multidimensional complexity, categorization of mental disorders is necessary for human understanding and communication, as well as for clinical decision making (Clark et al., 2017). Until further research proves otherwise, all thresholds for mental disorders should be considered arbitrary and discrete categories can thus be understood as simplifications of the underlying dimensions (Clark et al., 2017; Döpfner, 2022; Döpfner & Petermann, 2012).

Comorbidity: Co-occurring mental disorders, often referred to as comorbidity, is the rule rather than an exception (Angold et al., 1999; Caron & Rutter, 1991; Caspi et al., 2014;

Kessler et al., 2005). As shown by the National Comorbidity Survey, nearly half of individuals who received mental disorder diagnosis in the past 12 months received two (22%) or three and more (23%) diagnoses (Kessler et al., 2005). In terms of lifetime prevalence, around 50% of individuals were diagnosed with at least one mental disorder. Additionally, more than half of all lifetime disorders were diagnosed among the 14% of the population who have three or more comorbid disorders (Kessler et al., 1994). These rates of comorbidity are far above what we would expect by chance, that is, if the disorders were independent and distinct from each other (Boyd et al., 1984). From a nosological perspective, these high rates of comorbidity may be an indicator of an imperfect classification system and not exclusively, but at least partly, due to overlapping diagnostic criteria, arbitrary subdivisions of syndromes, one disorder serving as an early manifestation another, or one disorder being part of the other (Angold et al., 1999; Beauchaine & McNulty, 2013; Caron & Rutter, 1991; Clark et al., 2017; Krueger & Markon, 2006; Rouquette et al., 2018). Dimensional systems, in comparison, emphasize that all dimensions of psychopathology are positively associated with each other to varying degrees, and that the patterns of associations are as meaningful as the dimensions themselves (Krueger & Markon, 2006; Lahey et al., 2017, 2021).

Lack of diagnostic agreement: Another concern lies in the low to moderate agreement between clinicians on categorical diagnoses (Chmielewski et al., 2015; Markon et al., 2011; Regier et al., 2013). In the DSM-5 field trials, it was found that approximately 40% of the adult DSM-5 diagnoses analyzed did not meet the conventional cut-off criteria for acceptable interrater agreement (Chmielewski et al., 2015; Regier et al., 2013), although these disorders often demonstrated good to excellent reliability when considered dimensionally (Markon et al., 2011). One possible explanation for these findings is that applying a categorical classification system to phenomena that are continuously distributed can result in a significant loss of information and diagnostic instability because symptoms may fall just above or below the clinical threshold, thereby affecting the consistency of the diagnosis (Lahey et al., 2022; Markon et al., 2011).

Information from categorical vs. dimensional systems: One major advantage of dimensional classification systems is that they allow to identify individuals who experience mental health problems that fall just below the diagnostic threshold, but who have significant psychological distress or FI and are at a higher risk for meeting full diagnostic criteria for the disorder as time unfolds (Balázs et al., 2013; Balázs & Keresztény, 2014; see also Lahey et al., 2022). Interestingly, many patients seek help primarily due to their psychological distress or FI and not because of the mere presence of symptoms (Epstein & Weiss, 2012). Therefore, a

dimensional approach may be more comprehensive and facilitate the selection of stepwise-intensifying treatment interventions based on individual symptom severity (Lahey et al., 2021, 2022).

Number and kind of dimensions: Although dimensional classification systems have gained considerable empirical support (Achenbach, 2020; Kotov et al., 2017, 2021, 2022; Krueger et al., 2018), there is yet no universal agreement regarding the number and kind of dimensions. Lahey and colleagues (2021) concede that *"we are hampered by the current absence of a comprehensive dimensional measure of psychopathology that includes all symptoms"* (p. 57) and suggests that *"we cannot comprehensively define the dimensions of psychopathology until we can study all of the symptoms that define the universe of psychopathology at the same time and in the same way"* (p. 57). This desire for a universal dimensional classification system seems challenging at best, as it presupposes that the same measure of 'all symptoms' would be generalizable across ages, genders, raters, cultures and so forth (Achenbach, 2021). While the DSM, with its extensive list of over 200 diagnostic categories, may be excessively specific, the question remains as to which dimensions can be validly distinguished in psychopathology (Hartman, 2021).

Course features and etiology: Compared to categorical classification, dimensional systems do not incorporate course features such as age of onset, number of episodes, duration, and the progression of symptomatology, but instead their focus is on capturing trajectory features such as the average level of symptoms, variability over time, patterns of symptom cascaded, sensitivity to triggers, and response to treatments (Kotov et al., 2022). Also, dimensional systems do not contain assumptions about etiology. According to proponents of the dimensional approach, this was *"a deliberate decision, given limited understanding of mental disorders' etiology and difficulties in linking patient's symptoms to specific causes, such as dysphoria to trauma or psychosis to substance use"* (Kotov et al., 2022, p. 1674). When the underlying causes of symptoms are evident, including descriptions of these contributing factors may serve as a valuable addition to dimensional symptom profiles (Kotov et al., 2022).

Overall, while the debate over whether a mental disorder represents the endpoint of a continuous dimension of psychological problems rather than a discrete category distinct from 'normality' and other mental disorders remains open, so too does the question of under what conditions psychopathology can be more usefully be considered a category or a dimension (Clark et al., 2017; Coghill & Sonuga-Barke, 2012; Döpfner & Petermann, 2012; Andrew Pickles & Angold, 2003). Combining both systems may help to compensate for the shortcomings of the other approach (Clark et al., 2017; Döpfner, 2022; Döpfner & Petermann,

2012; Kraemer et al., 2004). For this purpose, various dimensional assessment instruments are already available.

1.2 Dimensional assessment instruments of child and adolescent psychopathology

Several comprehensive dimensional classification systems have been developed based on patients' self-reports and proxy-reports by caregivers (e.g., parents, teachers) using questionnaire procedures, two of which will be presented as examples: The *Achenbach System of Empirically Based Assessment* (ASEBA), which has been most widely used worldwide (Achenbach & Rescorla, 2000, 2001, 2003) and the *Diagnostic System for Mental Disorders in Children and Adolescents based on the ICD-10 and DSM-5* (*Diagnostik-System für Psychische Störungen im Kindes- und Jugendalter nach ICD-10 und DSM-5*; DISYPS-III), which is widely used in German-speaking countries (Döpfner & Görtz-Dorten, 2017; Görtz-Dorten et al., 2022).

1.2.1 Achenbach System of Empirically Based Assessment (ASEBA)

The ASEBA is the most comprehensive system on which an evidence-based dimensional classification of mental disorders in childhood and adolescence has been built (Achenbach, 2020; Achenbach et al., 2008; Achenbach & Rescorla, 2000, 2001, 2003) and it has been adapted and reviewed for the German-speaking countries by the German Child Behavior Checklist working group (Döpfner et al., 2014; Plück et al., 2020). There are several broadband questionnaires available which assess behavioral problems, emotional problems, and somatic complaints of children and adolescents from the age of 1.5-18 years rated by parents and caregivers/teachers and from the age of 11 years rated by adolescents themselves. The school-age forms of the Child Behavior Checklist (Achenbach & Rescorla, 2001; Döpfner et al., 2014), which include the *Parent's Report Form on Child and Adolescent Behavior* (CBCL/6-18R), the *Teacher's Report Form on Child and Adolescent Behavior* (TRF/6-18R), and the *Youth Self-Report Form* (YSR/11-18R), form the core of this dimensional assessment system. Through multivariate statistics (e.g., factor analytic techniques), Achenbach and colleagues identified patterns of co-variation between signs and symptoms and formed eight syndrome scales and two higher-order broadband scales, namely, the *Internalizing Problems* and *Externalizing Problems* and an additional *Total Problems* scale. This dimensional structure of school-age forms has been widely confirmed across cultures in national and international studies of representative field samples (Döpfner et al., 2014; Ivanova et al., 2019; Ivanova, Achenbach, Dumenci, et al., 2007; Ivanova, Achenbach, Rescorla, Dumenci, Almqvist, Bathiche, et al., 2007; Ivanova, Achenbach, Rescorla, Dumenci, Almqvist, Bilenberg, et al.,

2007; Rescorla et al., 2007, 2012). For preschool age, similar dimensions could be identified and replicated across cultures (Ivanova et al., 2010, 2011; Plück et al., 2020). Notably, such 'bottom-up' paradigms prioritize the utilization of empirical data as the foundation for conceptualizations of psychopathology (Achenbach, 1966, 2020).

1.2.2 Diagnostic System for Mental Disorders in Children and Adolescents based on the ICD-10 and DSM-5 (DISYPS-III)

Most dimensional classification systems are based on questionnaires completed by parents, caregivers/teachers, or adolescents themselves. One of the few exceptions is the DISYPS, a comprehensive diagnostic system widely used in German-speaking countries that additionally includes clinical judgment (Döpfner & Görtz-Dorten, 2017; Görtz-Dorten et al., 2022). The DISYPS-III is a psychometrically sound diagnostic system for assessing mental disorders in children and adolescents according to ICD-10 and DSM-5 diagnostic criteria (Döpfner & Görtz-Dorten, 2017). The DISYPS systematically combines the assessment of three different rater perspectives. More specifically, parents, teachers, or other caregivers can complete proxy-rating forms (*Fremdbeurteilungsbogen*; FBB). Likewise, children and adolescents aged 11;0-17;11 years can assess themselves using the self-report form (*Selbstbeurteilungsbogen*; SBB). Clinicians can form their own clinical judgment using diagnostic checklists (*Diagnose-Checklisten*; DCL) or the newly developed semi-structured clinical interviews (*Interview-Leitfäden*; ILF), which are presented in more detail in Chapter 1.8. Besides symptomatology, the DISYPS assesses psychological distress and FI in the domains of academic performance, home life and family members, relationships with adults, as well as relationships with children and recreational activities. Notably, the DISYPS can be used to create both categorical diagnoses and dimensional symptom profiles via clinical judgments based on clinical explorations (DCL) or clinical interviews (ILF). A further distinguishing characteristic of the DISYPS is that all diagnostic instruments within a specific mental disorder domain consist of mostly identical items and share identical diagnostic scales. This design enables a systematic comparison of different perspectives from raters in a multi-informant approach (Döpfner & Petermann, 2012).

1.3 Multi-informant approach in the assessment of child and adolescent mental disorders

While adults themselves are usually the main informants for assessment data in adult mental health services, parents and other caregivers (e.g., teachers) are the main informants for

data in child and adolescent assessment (Achenbach, 2020; De Los Reyes et al., 2023). More than 30 years ago, Achenbach and colleagues (1987) conducted the first meta-analysis on the agreement of different informants in diagnosing mental disorders in children. Their findings revealed that the *"mean* rs *were .60 between similar informants (e.g., pairs of parents), .28 between different types of informants (e.g., parent/teacher), and .22 between subjects and other informants"* (Achenbach et al., 1987, p. 213). Successive meta-analyses have consistently shown similar disparities across informants in the assessment of child psychopathology (De Los Reyes et al., 2015, 2019). As a result, it became evident that no single informant could fully capture the important variations in how children function across different interaction partners and contexts, such as the home or school environment (Achenbach, 2020; De Los Reyes, 2011; De Los Reyes et al., 2013, 2023; Dirks et al., 2012; see also Thöne et al., 2021). Initially, it was postulated that these informant discrepancies were indicative of measurement error, invalidity, or rater bias (e.g., Roberts & Caspi, 2001), extensive theoretical and empirical work, along with ongoing discussions regarding optimal approaches to mental health assessment indicate that the significance of utilizing multi-informant assessments stems from their ability to capture distinct viewpoints offered by the reports of each informant (De Los Reyes, 2011; De Los Reyes et al., 2013, 2023; Dirks et al., 2012; Hunsley & Mash, 2007; Thöne et al., 2021, 2022). Above all, in their editorial statement, over 70 leading experts on informant discrepancies present a set of guidelines and future research directions that focus on the most common findings of multi-informant mental health assessments (De Los Reyes et al., 2023). Overall, a consensus among experts has emerged that discrepancies across informants ought to be *"embrace[d], not erase[d]"* (Dirks et al., 2012, p. 558).

1.4 Hierarchical systems of categorical and dimensional classification

As summarized above, the dimensional classification systems described so far are each based on a group of assessment instruments developed according to similar principles. In recent years, hierarchical categorical-dimensional systems have also been proposed which share a common structure across a variety of assessment instruments (Caspi et al., 2014; Caspi & Moffitt, 2018; Kotov et al., 2017; Krueger et al., 2018; Lahey et al., 2012, 2017, 2018; Martel, Pan, et al., 2017). One prominent example is the development of the *Hierarchical Taxonomy of Psychopathology* (HiTOP) by a large international consortium led by Kotov and colleagues (2017, 2021, 2022), whose goal is to develop an overarching evidence-based dimensional classification system. The HiTOP system is a collaborative effort among nosologists from various mental health disciplines aimed at enhancing the description, organization, and

measurement of psychopathology across the life span. The main goals of this consortium are twofold: (a) to gather data that can improve and refine the HiTOP system, and (b) to disseminate the resulting information to researchers and clinicians (Conway et al., 2019, 2021; Kotov et al., 2017, 2021, 2022; Krueger et al., 2018; Ruggero et al., 2019). This system, still under development, outlines six hierarchical layers, extending from *super spectra* at the highest level to *symptoms* at the lowest level (Kotov et al., 2022): *"*Super spectra *are higher-order dimensions, assumed to influence all spectra on the layer below. The HiTOP specifies six* spectra *(e.g. antagonistic externalizing), an array of* subfactors *(e.g. antisocial behavior),* syndromes and disorders *(e.g. conduct disorder [CD]), which are used synonymously with DSM-5 diagnoses at this point to facilitate communication,* components *(e.g. maladaptive traits) and* symptoms *(e.g. physical aggression) on the lowest level"* (Thöne et al., 2022, p. 845). Ultimately, the HiTOP consortium aims to improve the classification of psychopathology beyond the traditional ICD and DSM-based diagnostic systems (Kotov et al., 2022).

In recent times, there has been a significant focus on the highest level of the hierarchical structure, resulting in the exploration of a general factor of psychopathology known as the *p*-factor (Caspi et al., 2014; Caspi & Moffitt, 2018; Lahey et al., 2012, 2017, 2018; Martel, Pan, et al., 2017). Analogous to the general intelligence factor, the *p*-factor may reflect low to high severity of psychopathology (Caspi & Moffitt, 2018). The search for a *p*-factor was spurred by research showing that even Achenbach's higher-order externalizing and internalizing dimensions overlapped to a considerable degree ($r \approx .50$; see Achenbach, 2020; Achenbach et al., 2016; Krueger, 1999; Krueger et al., 1998; Krueger & Markon, 2006) and that a third higher-order dimension, psychotic experiences, could be identified in adult populations (Keyes et al., 2013; Wright et al., 2013). The notion is that this single dimension, the *p*-factor, might encompass shared characteristics among all mental disorders observed in clinical and community samples across various age groups (Caspi et al., 2014; Caspi & Moffitt, 2018; Lahey et al., 2012, 2017, 2018; Martel, Pan, et al., 2017). While the exact nature of the *p*-factor have not yet been comprehensively evaluated (Fried, 2020; van Bork et al., 2017; Watts et al., 2020), it was suggested that it may represent neuroticism or negative emotionality (e.g., Lahey et al., 2017), poor impulse control over emotions (e.g., Carver et al., 2017), or disordered thought processes (e.g., Caspi et al., 2014). From a clinical perspective, searching for a *p*-factor of psychopathology seems intriguing, since it may provide an explanation why transdiagnostic therapies are being tested with notable successes (Meier & Meier, 2018; Newby et al., 2015).

1.5 Latent factor models of psychopathology

Traditionally, latent factor models have been applied to characterize the simultaneous presence of signs and symptoms within specific disorders, as well as the interrelationships among diagnostic entities (Achenbach, 1966; Eaton, 2015; Krueger, 1999; Krueger et al., 1998; Krueger & Markon, 2006). Put simply, *"the key notion underlying factor analytic models is that some variables of theoretical interest cannot be observed directly; these unobserved variables are termed* latent variables *or* factors*. Although latent variables cannot be measured directly, information related to them can be obtained indirectly by noting their effects on observed variables believed to represent them"* (Byrne, 2005, p. 17). Although there is a variety of multivariate factor analyses for modeling and investigating the factor structures of constructs, the field of psychological research has predominantly relied on exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) as preferred methods (Byrne, 2005; Marsh et al., 2014). EFA represents primarily a hypothesis-generating method, which seems particularly suitable *"when links between the observed variables and their underlying factors are unknown or uncertain"* (Byrne, 2005, p. 17). In comparison, CFA represents a hypothesis-testing method, which is appropriate when researchers possess some prior knowledge or assumptions regarding the underlying latent variable structure (Byrne, 2005). Although CFA is widely used to assess the dimensional structure of psychopathology (Ringwald et al., 2023), it is not without limitations (cf. Thöne et al., 2021). From a conceptual perspective, CFA models are often criticized for being overly idealistic, simplistic, restrictive, since they assume the existence of 'pure factors', where variables only load onto a single predetermined latent factor, and cross-loadings are restricted to zero (Marsh et al., 2014; Morin et al., 2016). Considering that most variables are associated with multiple conceptually related latent factors, it is reasonable to expect some level of construct-relevant psychometric multidimensionality among the variables (Morin et al., 2016). For these reasons, exploratory structural equation modeling (ESEM) has been designed (Marsh et al., 2014; Morin et al., 2016). ESEM combines the strengths of EFA, which allows for cross-loadings, and CFA, which enables *a priori* model specification and assessment of goodness-of-fit indices (e.g., Thöne et al., 2021). When latent variable modeling approaches are applied to diagnostic data, they provide insights into latent factors that account for the observed connections among different dimensions of mental disorders (Carragher et al., 2015; Caspi & Moffitt, 2018; Eaton, 2015; Forbes, Sunderland, et al., 2021; Thöne et al., 2021).

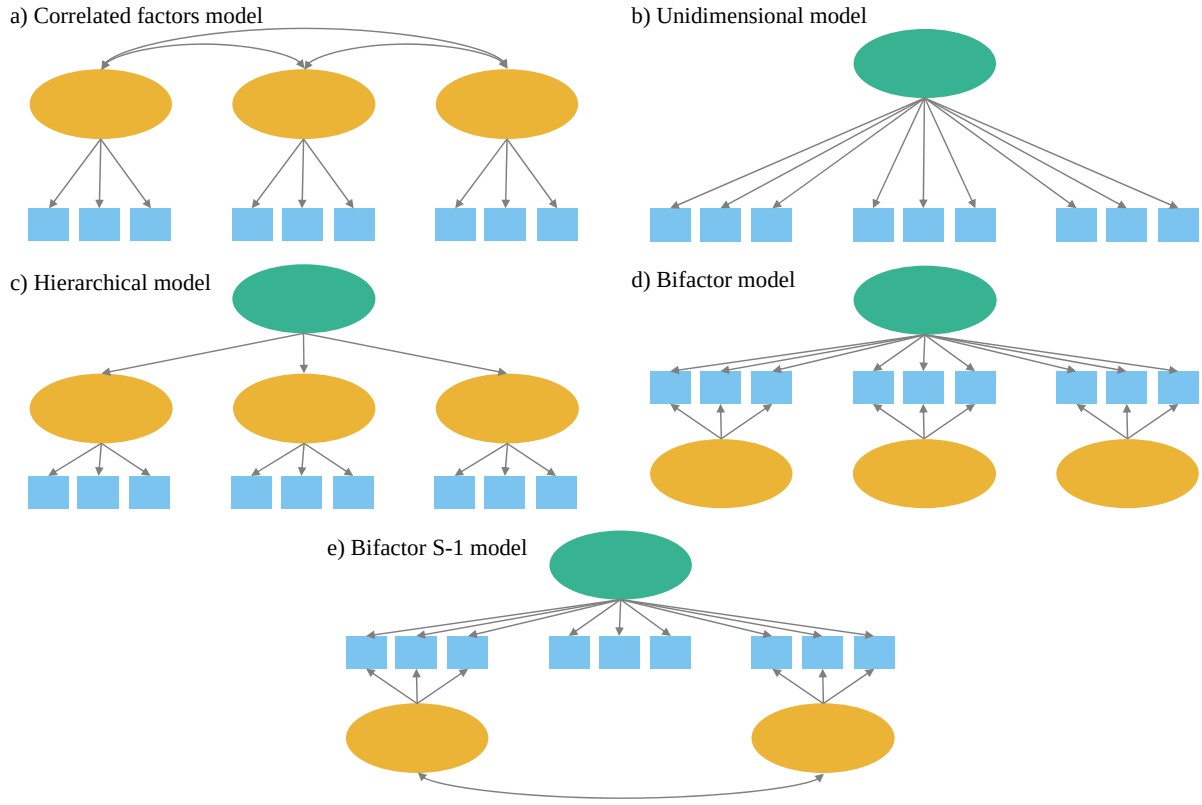
Latent factor models encompass a variety of types, each carrying its own advantages and disadvantages while providing distinct information that can effectively address different research questions (Brunner et al., 2012; Byrne, 2005; Lahey et al., 2021; Pettersson et al.,

2021; see also Thöne et al., 2021, 2022). The following is a brief outline of the factor models also used in this dissertation (see Study 2; Thöne et al., 2021), which are employed frequently, primarily focusing on investigating the interrelationships between dimensions of mental disorders (Figure 1). First-order correlated factor models offer a good starting point and provide initial insights into the mutual associations between latent constructs (Figure 1a; Eid, 2020). High correlations between the factors indicate that the first-order correlated factor model remains an incomplete model, as it does not account for the associations between dimensions (Brunner et al., 2012). In this case, a unidimensional model, reflecting that a single factor explains the variance across all observed variables, may capture the shared associations better than separate factors (Figure 1b). Alternatively, higher-order models or bifactor models can be applied, which assume some sort of overarching factor. In a higher-order factor model, the second-order factor (e.g., externalizing factor) signifies the variance that is common or shared between the first-order factors (e.g., ADHD, CD, ODD; Figure 1c). On the other hand, traditional bifactor models present an alternative perspective, positing a single common liability known as the g-factor, which accounts for the variance observed in all variable scores. Furthermore, these models incorporate a group of independent specific factors known as s-factors, which specifically impact a smaller subset of symptoms and disorders (Figure 1d). The term 'bifactor model' originated from intelligence research, where it proposed the existence of a general factor (g) shared by all items on a test, alongside specific factors that are common to a smaller subset of related items, representing independent cognitive modules (Caspi & Moffitt, 2018). The underlying concept behind bifactor models is that *"only with the general factor variance removed can we have a clear window into the remaining covariance patterns among the symptoms in our measure. Only with specific measures unconfounded by the general factor can we have a clear window into the etiological or prognostic associations"* (Hartman, 2021, p. 72). Importantly, there is a critical difference regarding the interpretation of the specific factors in hierarchical and bifactor models: *"lower-order factors in the hierarchical model represent the dimensionality of psychopathology* within *the general factor, while the specific factors in the bifactor model represent this dimensionality* beyond *the general factor"* (Hartman, 2021, p. 72). However, traditional bifactor models have faced theoretical and empirical concerns (Arias et al., 2018; Bonifay et al., 2017; Burns et al., 2020a; Eid, 2020; Eid et al., 2017; Heinrich et al., 2020, 2021; Reise, 2012; Thöne et al., 2022; Watts et al., 2019), leading to the development of bifactor S-1 models as an alternative (see also Chapter 6). Bifactor S-1 models define one first-order factor as the general reference factor (Figure 1e), enabling comparability of g-factors across studies. The remaining specific factors are

understood as residual factors, encompassing the true score variance that is *not* explained by the general reference factor (Burns et al., 2020a; Eid et al., 2017). The choice of a general reference factor should be based on theoretical principles or align with a specific facet of interest (Eid, 2020). As becomes clear, the appeal of latent factor analysis lies in its fundamental assumptions, suggesting that, comparable to medical conditions, there may exist a discernible cause that manifests as a set of symptoms. This perspective implies that identifying the underlying cause would aid in accurate diagnosis and treatment (Kotov et al., 2017).

Figure 1

Schematic illustrations to reflect on the positive statistical associations between mental disorders



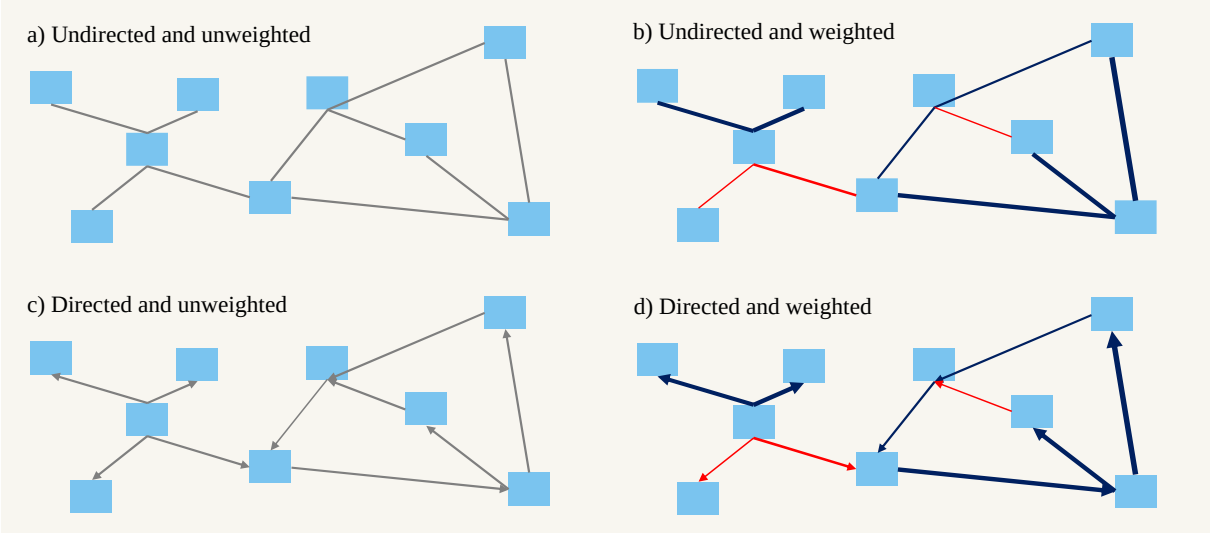
1.6 A network approach to psychopathology

In recent years, there has been significant progress in the field of network analysis, offering an alternative perspective to understand the patterns of covariation among symptoms or mental disorders, depending on the specific network model employed (Borsboom, 2008; Borsboom et al., 2011; Borsboom & Cramer, 2013; Cramer et al., 2010). Advocates of the network approach contend that traditional psychology has primarily embraced a latent variable framework, in which *"a mental disorder is viewed as a latent variable that* causes *a*

constellation of symptoms" (Cramer et al., 2010, p. 137). In contrast, the network perspective shifts the focus towards individual symptoms as the primary units of analysis, considering mental disorders as networks consisting of interconnected symptoms (Borsboom, 2008, 2017; Borsboom & Cramer, 2013). Graphical representations of the networks can display the associations between symptoms (Figure 2). Each psychometric network generally consists of circular nodes representing the symptoms being analyzed and linear edges representing a pairwise statistical association between each pair of nodes (Epskamp et al., 2018). As illustrated in Figure 2, edges can be either unweighted or weighted (i.e., the width of the edge reflects the strength of the relationship), undirected or directed (i.e., indicated by a line or by an arrow), and positive or negative (i.e., typically indicated by red or blue colors). Network models can incorporate not only symptoms but also other components like risk factors or FI domains to examine their associations with individual symptoms. This intuitive interpretation is further facilitated by the Fruchterman and Reingold algorithm, which is used in many psychometric networks, in which strongly related symptoms tend to be attracted to each other (i.e., cluster together) and symptoms with weaker relations tend to oppose each other (i.e., be positioned at the edges of the network). Thus, the network approach aims to comprehend mental disorders by emphasizing the relationships between symptoms, without relying on latent factors (Borsboom & Cramer, 2013; see also Junghänel et al., 2023). For instance, feeling tired may lead to evening rumination, which in turn can impair concentration. The collective presence of these symptoms and their interconnectedness is what is currently referred to as depression (cf. Junghänel et al., 2023). This example exemplifies the fundamental principle of network models, which posits that symptoms have direct causal influences on each other's occurrence (Borsboom & Cramer, 2013; van Bork et al., 2017). Within this context, symptoms are regarded as *"active agents in networks of interacting components rather than passive indicators of latent variables"* (van Bork et al., 2017, p. 767).

Figure 2

Different types of psychometric networks



Network theory holds substantial consequences for how we understand the diagnosis and treatment of mental disorders, as highlighted by Borsboom (2017). One such implication is the recognition of central (i.e., important) symptoms that have a strong influence on activating other symptoms or clusters within a network associated with a particular mental disorder (e.g., Junghänel et al., 2023). Targeting these central symptoms in therapeutic interventions may offer effective treatment strategies (Borsboom, 2017). Moreover, the individualized nature of network configurations aligns well with the concept of personalized treatment approaches (Borsboom, 2017). Accordingly, the principles of network theory also converge with the rationale behind cognitive-behavioral interventions (Borsboom, 2017). Cognitive-behavioral therapy, for instance, aims to modify dysfunctional cognitions (e.g., rumination) in order to bring about behavioral changes (e.g., withdrawal), reflecting the underlying principles of network theory (Borsboom, 2017).

Next, network analysis provides a fresh view on comorbidity, wherein symptoms shared by two disorders, such as sleep disturbances or difficulties to concentrate observed in both depression and general anxiety disorders, may simultaneously activate the networks associated with these co-occurring disorders (Borsboom, 2017; Cramer et al., 2010; Junghänel et al., 2023). Consequently, instead of perceiving comorbidity as a byproduct of a flawed diagnostic system caused by overlapping criteria (see also Chapter 1.1), network theory suggests that *"comorbidity is an intrinsic feature of mental disorders"* (Borsboom, 2017, p. 7) and that *"comorbidity should be seen as part of the flesh and bones of psychopathology"* (Borsboom, 2017, p. 8). In addition to exploring comorbidity, network analysis holds promise in addressing various other research inquiries at the symptom level, including the examination of associations

between symptoms and FI (Burns et al., 2022; Goh et al., 2020; Goh, Martel, et al., 2021). As previously stated, the diagnosis of mental disorders not only relies on the presence of symptoms with a specific frequency and severity but also takes into account the presence of FI or psychological distress (Palermo et al., 2008; Spitzer & Wakefield, 1999; Wakefield, 2007). FI encompasses a broad range of domains and can be defined as *"the extent of restriction in a child's ability to perform important daily life activities including physical, social, and personal activities due to their health condition or to specific symptoms"* (Palermo et al., 2008, p. 984). Although symptoms and FI are often interconnected, it is important to recognize that they are distinct constructs. In other words, the severity of symptoms does not necessarily correspond to higher levels of FI (Arildskov et al., 2022; Palermo et al., 2008; Rapee et al., 2012). Previous research has predominantly focused on examining the relationship between FI and symptomatology at the *scale level* (e.g., ADHD Inattention; Garner et al., 2013; Willcutt et al., 2012; Zoromski et al., 2015). However, only a limited number of studies have employed network analysis to delve into the individual connections between symptoms and various domains of FI, providing a more detailed understanding at the *symptom level* (Burns et al., 2022; Goh et al., 2020; Goh, Martel, et al., 2021; Thöne et al., 2023). Consequently, there is still a significant clinical research gap in investigating whether specific symptoms exert more substantial impairments compared to others. This inquiry is crucial due to the significant diversity in symptoms encountered by patients (Mota & Schachar, 2000; Zoromski et al., 2015).

Overall, latent factor models and the network approach to conceptualizing psychopathology are often presented in the scientific literature as a kind of dispute, with researchers from one camp challenging the methodology of the other (reviewed by Eaton, 2015; McNally, 2021). While some researchers enthusiastically embrace the novelty and potential of network models (see Contreras et al., 2019), others express skepticism regarding the replicability and reliability of network findings (Forbes et al., 2017a, 2019; Forbes, Wright, et al., 2021b). The strengths and limitations of both approaches are discussed in more detail in the Discussion section (Chapter 6), along with a proposed roadmap for future research.

1.7 Structured clinical interviews for child and adolescent mental disorders

As described in Chapter 1.2, clinical judgment provides an important rater perspective in the context of multi-informant assessment (Döpfner & Petermann, 2012). For this purpose, structured clinical interviews have been developed, which allow a systematic assessment of the diagnostic criteria (Frick et al., 2010; Leffler et al., 2015; Segal & Williams, 2014). The following outline of structured clinical interviews builds on the work by Thöne et al. (2020).

Structured clinical interviews are widely regarded as a gold standard for determining the presence of mental disorders in clinical settings (Hoyer & Knappe, 2012; Rettew et al., 2009), since they enhance the accuracy and reliability of diagnostic outcomes (Frick et al., 2010; Leffler et al., 2015; Segal & Williams, 2014). Regarding their structural characteristics, these interviews can be categorized into unstructured, semi-structured, and highly structured. Interviews characterized by a high degree of structure typically follow a categorical approach, since their aim is to establish categorical diagnoses (Frick et al., 2010; Leffler et al., 2015; Segal & Williams, 2014). Such interviews require minimal training but impose constraints on the interviewer's ability to thoroughly explore the patient's symptomatology. In most cases, closed-ended questions are used, allowing for a dichotomous assessment where clinical symptoms are categorized as either absent or present (Frick et al., 2010; Leffler et al., 2015; Segal & Williams, 2014). Examples of highly structured clinical interviews for the assessment in childhood and adolescence include the *Mini-International Neuropsychiatric Interview for Children and Adolescents* (Mini-KID; Sheehan et al., 2010), the *NIMH Diagnostic Interview Schedule for Children Version IV* (NIMH DISC-IV; Shaffer et al., 2000), and the *Children's Interview for Psychiatric Syndromes* which consists of separate versions for children (ChIPS; Weller et al., 1999b) and parents (P-ChIPS; Weller et al., 1999a). As reviewed by Leffler and colleagues (2015), structured interviews exhibit strong interrater reliability and satisfactory validity when utilized in clinical and community settings. However, these outcomes may be influenced by the high level of structure, which restricts interviewers' ability to form independent clinical judgments (Leffler et al., 2015; Segal & Williams, 2014). In contrast, semi-structured interviews offer greater flexibility by allowing for the exploration of symptoms, reasoned judgments, and the use of Likert scales for scoring responses. While this format necessitates more extensive training, it enables a dimensional approach that considers symptom severity (Döpfner & Petermann, 2012; Frick et al., 2010; Leffler et al., 2015). Examples of semi-structured clinical interviews for assessing children and adolescents include the *Schedule for Affective Disorders and Schizophrenia for School Aged Children* (K-SADS; Kaufman et al., 1997), the *Child and Adolescent Psychiatric Assessment* (CAPA; Angold & Costello, 2000), and the *Diagnostic Interview for Children and Adolescents* (DICA; Reich, 2000). Overall, semi-structured interviews demonstrate robust psychometric properties in terms of their reliability and validity in both clinical and community settings (reviewed by Leffler et al., 2015). Diagnostic agreement was generally satisfactory to good for most diagnoses (Leffler et al., 2015), but lower for some diagnoses, such as conduct disorder (CD; kappa = .50; Angold & Costello, 1995). One example of a semi-structured interview following a dimensional approach

is the *Semistructured Clinical Interview for Children and Adolescents* (SCICA; McConaughy & Achenbach, 2001). The SCICA covers a wide range of behavioral and emotional problems based on the internalizing-externalizing model introduced by Achenbach (1966). The SCICA demonstrates acceptable internal consistencies for all scales, except for *Anxious* and *Family Problems* ($\alpha < .70$; cf. Nunnally, 1978) and acceptable interrater reliability for the *Internalizing* ($r = .66$) and *Externalizing* scales ($r = .74$) in an outpatient Dutch sample ($N = 185$; Kasius, 1997). Since the SCICA provides empirically based syndromes, similar to the other instruments of the ASEBA system, this transdiagnostic interview does not adhere to the formulations of the ICD or DSM diagnostic criteria.

Considering the dynamic conceptualizations of psychopathology, there is an urgent need for clinical interviews that align with these changing requirements (cf. Thöne et al., 2020). One such evolving aspect pertains to state-of-the-art assessment tools needing the flexibility to determine categorical diagnoses while also establishing dimensional symptom profiles (e.g., Döpfner, 2022). Moreover, since clinical interviews are primarily conducted to determine a diagnosis, they should systematically assess diagnostic criteria according to the DSM or ICD systems.

1.8 Development and description of the clinical parent interviews DISYPS-ILF

To meet these changing needs, the *Clinical Parent Interviews According to the DSM-5 Diagnostic System of Mental Disorders for Children and Adolescents* (*Interview-Leitfäden zum Diagnostik-System für psychische Störungen nach DSM-5 für Kinder und Jugendliche;* DISYPS-ILF) were developed (Görtz-Dorten et al., 2022). The DISYPS-ILF are semi-structured diagnostic interviews covering a broad spectrum of current mental disorders in children and adolescents according to DSM-5 criteria. These interviews are a further product of the DISYPS and allow for a more structured exploration and assessment of diagnostic criteria in addition to the clinical exploration using the DCL from the DISYPS-III (see Chapter 1.2.2). The DISYPS-ILF can be conducted with caregivers (parents, educators, teachers) and with patients themselves from the age of eight. Overall, the following clinical interviews are available: *Interview for Externalizing Disorders* (*Interview-Leitfaden für Externale Störungen*; ILF-EXTERNAL), *Interview for Internalizing Disorders* (*Interview-Leitfaden für Internale Störungen*; ILF-INTERNAL), *Interview for Obsessive-Compulsive and Tic Disorders* (*Interview-Leitfaden für Zwangs- und Tic-Störungen;* ILF-ZWANG/TIC), *Interview for Contact Disorders* (*Interview-Leitfaden für Kontakt-Störungen;* ILF-KONTAKT). In addition, the *Interview for Screening of Mental Disorders* (*Interview-Leitfaden zum Screening*

psychischer Störungen; ILF-SCREEN) provides guidance for a wide range of mental health problems.

A noteworthy and distinctive aspect of the DISYPS-ILF is that both the clinical interviews and the report forms for parents, teachers/caregivers, and patients align with the DISYPS-III (Döpfner & Görtz-Dorten, 2017) and utilize the same diagnostic scales (cf. Thöne et al., 2020). This allows for a direct comparison of ratings between parents, teachers/caregivers, and patients with clinical assessments (cf. Thöne et al., 2020). Within the scope of this dissertation, the ILF-EXTERNAL was psychometrically evaluated and will therefore be briefly presented below.

1.8.1 DISYPS ILF-EXTERNAL

As outlined by Thöne et al. (2020), the ILF-EXTERNAL comprises a series of items, each assessing a symptom criterion based on the DSM-5. Clinicians assess each item utilizing a 4-point Likert scale, spanning from 0 (representing 'age-typical / not at all') to 3 (indicating 'very much'), where higher scores signify increased severity of symptoms. To assist clinicians in their judgment, a brief description of symptom severity is included for each score. Furthermore, for each item, an illustrative sentence is provided that exemplifies a child's behavior corresponding to a rating of 3. Item scores of 2 or higher are considered clinically relevant and indicative of meeting DSM-5 symptom criteria (cf. Thöne et al., 2020).

One notable advantage of the DISYPS-ILF is their ability to facilitate both categorical and dimensional assessments. From a categorical perspective, the ILF-EXTERNAL covers several DSM-5 diagnoses (including associated ICD-10 codes), including ADHD with its subtypes (‘combined type’, ‘predominantly inattentive type’, and ‘predominantly hyperactive-impulsive type’), oppositional defiant disorder (ODD), disruptive mood dysregulation disorder (DMDD), and CD with the specifier of limited prosocial emotions (cf. Thöne et al., 2020). Alternatively, from a dimensional approach, the corresponding scale scores can be calculated by averaging the item scores associated with each diagnosis, providing additional information on the severity level.

1.9 Interim summary

In summary, the literature reveals six main findings: (1) Dimensional models of psychopathology challenge existing categorical diagnostic systems and offer a potentially more accurate and parsimonious representation of mental disorders. (2) Incorporating input from multiple sources is crucial for evidence-based evaluation of child psychopathology, as diverse

informants may offer differing evaluations of symptom severity. (3) Through latent factor analysis, researchers have established latent dimensions of mental disorders based on the observed covariation of symptoms, such as the HiTOP model, which presents a hierarchical and transdiagnostic alternative to traditional classifications. (4) The network approach to psychopathology has advanced as an alternative viewpoint which is both specific by zooming in on individual symptom relations and by zooming out on the comorbidity of mental disorders. (5) Semi-structured interviews offer greater flexibility than highly structured interviews, by allowing for the exploration of symptoms, reasoned judgments, and the assessment of symptom severity and impairment in a dimensional manner, i.e., although this scoring format necessitates comprehensive training, it enables a dimensional perspective by considering the continuous range of symptom frequency and severity, as well as impairments. (6) The recently developed DISYPS-ILF are the first semi-structured clinical interviews in Germany for diagnosing mental disorders in children and adolescents, which specifically adhere to DSM-5 formulations. Consequently, it is crucial to comprehensively evaluate the psychometric properties of the DISYPS-ILF to enhance diagnostic assessment and treatment planning.

2 Thesis objectives

The main objective of this doctoral dissertation is to contribute to the establishment of a rigorous standard for diagnosing mental disorders in children and adolescents. Specifically, this dissertation provides a comprehensive psychometric evaluation of the clinical parent interview for diagnosing externalizing disorders in children and adolescents (ILF-EXTERNAL) from the DISYPS-ILF (Görtz-Dorten et al., 2022) with respect to the following aspects:

- To evaluate the reliability and validity of the ILF-EXTERNAL and, in particular, to examine categorical and dimensional diagnostic agreement among clinicians (Study 1)
- To examine the factorial structure of the ILF-EXTERNAL and systematically test the latent dimensions underlying externalizing symptoms using a multi-informant approach (Study 2)
- To enhance our understanding of the differential relationships between individual symptoms and domains of psychological distress and impairment (Study 3)

The first study of this cumulative dissertation (Chapter 3; Thöne et al., 2020) presents the psychometric properties of the ILF-EXTERNAL. Clinicians conducted the ILF-EXTERNAL with the parents or primary caregivers of 474 children with ADHD symptoms (see study description below). To determine interrater reliability at the dimensional level, intraclass correlation coefficients were calculated between the interviewers who administered the ILF-EXTERNAL and two independent raters. Agreement on categorical DSM-5 diagnoses was evaluated using Fleiss' kappa. Additional analyses were conducted to evaluate internal consistencies, item-total correlations, and the associations between symptom severity and FI. Convergent and divergent validity were evaluated using parent-ratings from the CBCL/6-18R and symptom checklists from the DISYPS-III (FBB/SBB-ADHS; FBB/SBB-SSV). Implications for categorical and dimensional concepts of mental disorders are discussed. This study (Thöne et al., 2020) provides an important contribution to a high-quality psychometric evaluation of clinical interviews with a focus on categorical and dimensional diagnostic agreement among clinicians .

The second study (Chapter 4; Thöne et al., 2021) aimed to investigate the applicability of various latent factor models, which were derived from the trait-impulsivity theory, to externalizing symptoms. More specifically, (a) a unidimensional model, (b) a first-order correlated factors model with five correlated factors, (c) a higher-order factor model with one second-order factor and five first-order factors, (d) a traditional bifactor model with a g-factor and five s-factors, and (e) a bifactor S-1 model with a general reference factor of hyperactivity-impulsivity symptoms were tested using CFA and ESEM. Additionally, the study examined the

measurement invariance of the prevailing factor models across different raters (clinicians, parents, teachers) and assessment methods (interviews, questionnaires). The discussion provides a critical comparison of the CFA and ESEM approaches, highlighting their respective advantages and disadvantages in elucidating latent dimensions of psychopathology. Moreover, the study critically evaluated the claimed superiority of bifactor models as the 'winning model' and provided future directions for modeling the dimensional structure of psychopathology. By combining factor-analytic approaches to symptomatology with etiological models, this study (Thöne et al., 2021) contributes to a deeper understanding of the multifactorial and dimensional nature of psychopathology in children.

The third study (Thöne et al., 2023) enhanced our understanding of the differential relationships between individual symptoms of ADHD and disruptive behavior disorders (i.e., ODD, CD, callous unemotional [CU] symptoms) and global FI as well as FI in specific domains such as academic performance; home life and family members; relationships with adults; relationships with children and recreational activities; and psychological distress. The ILF-EXTERNAL was employed by clinicians to evaluate the severity of individual symptoms and five functional impairment domains associated with symptoms of ADHD or symptoms of ODD/CD/CU, respectively. First, multiple linear regression analyses were conducted to determine the influence of individual ADHD symptoms or ODD/CD/CU symptoms on global FI. Second, two psychological networks were created to identify the most robust and distinct connections between individual symptoms of ADHD or ODD/CD/CU and multiple FI domains. The accuracy and stability of these networks were assessed using simulation studies and bootstrapping procedures. The discussion highlights clinical implications regarding the relationship between symptom severity and FI and provides future perspectives on the development of assessment instruments. This study (Thöne et al., 2023) contributes to the growing literature emphasizing the importance of analyzing the connections between individual symptoms and various domains of FI, underscoring the clinical utility of symptom-based approaches.

2.1 Study description

The data for all three studies were obtained from the ESCAschool *(Evidence-based, Stepped Care of ADHD in school-aged children)* project, which is a multicenter study conducted at nine different sites in Germany. The primary objective of the ESCAschool study *"is to assess the efficacy of a stepped care approach involving individually tailored adaptive treatment strategies"* (Döpfner et al., 2017, p. 3). As outlined in the study protocol, families

seeking treatment in the ESCAschool study had to need the following inclusion criteria: *"(i) child age 6;0 to 11;11 years; (ii) child attendance of school (including special schools); (iii) child meeting criteria for ADHD diagnosis according to the DSM-5; (iv) existence of informed consent of both parents or guardians and assent of the child. Exclusion criteria are: (i) child intelligence quotient (IQ) below average (IQ < 80); (ii) child clinical diagnosis of a pervasive developmental disorder, schizophrenia, bipolar disorder, severe depressive episode, epilepsy or heart disease; (iii) insufficient German language or reading skills of the parent with primary treatment involvement; (iv) current or planned intensive behaviour therapy for child ADHD or oppositional behaviour on a weekly basis; (v) known non-response of the child to all standard ADHD medication (methylphenidate, dexamphetamine, atomoxetine); (vi) psychotropic medication of the child other than for the treatment of ADHD, or neuroleptic medication other than for the treatment of disturbances of impulse control*" (Döpfner et al., 2017, p. 3).

This doctoral dissertation examined the baseline data (i.e., before any intervention) of N = 474 school-age children (age: M = 8.90, SD = 1.49; 19% females). To assess whether participants met the diagnostic criteria for ADHD, the screening procedure included employing the clinical parent interview for identifying externalizing disorders in children and adolescents (ILF-EXTERNAL; Görtz-Dorten et al., 2022). In case children were already taking ADHD medication before the study, parents provided descriptions of their child's behavior both when the medication was and wasn't administered. For the analyses of this dissertation, only the symptomatology of children without medication was considered. Clinical diagnoses of ADHD, ODD, CD, and CU symptoms were made based on the ILF-EXTERNAL following the guidelines of the DSM-5 (cf. Thöne et al., 2020). The sample used in the data analysis for this dissertation included children who did not meet the DSM-5 criteria for an ADHD diagnosis. These participants, referred to as screening negatives (n = 32, 72% males), exhibited subthreshold ADHD symptoms (cf. Thöne et al., 2020). Additionally, all clinicians employed a clinical diagnostic checklist (DCL-SCREEN) from the DISYPS-III to screen for comorbid symptoms (Döpfner & Görtz-Dorten, 2017).

3 Toward a dimensional assessment of externalizing disorders in children: Reliability and validity of a semi-structured parent interview

Thöne, A.-K., Görtz-Dorten, A., Altenberger, P., Dose, C., Geldermann, N., Hautmann, C., Jendreizik, L. T., Treier, A.-K., von Wirth, E., Banaschewski, T., Brandeis, D., Millenet, S., Hohmann, S., Becker, K., Ketter, J., Hebebrand, J., Wenning, J., Holtmann, M., Legenbauer, T., Huss, M., Romanos, M., Jans, T., Geissler, J., Poustka, L., Uebel-von Sandersleben, H., Renner, T., Dürrwächter, U., & Döpfner, M. (2020). Toward a dimensional assessment of externalizing disorders in children: reliability and validity of a semi-structured parent interview. *Front. Psychol.*, *11*, 1840. https://doi.org/10.3389/fpsyg.2020.01840

Abstract

Objective: This study assesses the reliability and validity of the DSM-5-based, semi-structured *Clinical Parent Interview for Externalizing Disorders in Children and Adolescents* (ILF-EXTERNAL). *Method:* Participant data were drawn from the ongoing ESCAschool intervention study. The ILF-EXTERNAL was evaluated in a clinical sample of 474 children and adolescents (aged 6 - 12 years, 92 females) with symptoms of attention-deficit/hyperactivity disorder (ADHD). To obtain interrater reliability, the one-way random-effects, absolute agreement models of the intraclass correlation (ICC) for single ICC(1,1) and average measurements ICC(1,3) were computed between the interviewers and two independent raters for 45 randomly selected interviews involving ten interviewers. Overall agreement on DSM-5 diagnoses was assessed using Fleiss' kappa. Further analyses evaluated internal consistencies, item-total correlations as well as correlations between symptom severity and the degree of functional impairment. Additionally, parents completed the German version of the *Child Behavior Checklist* (CBCL) and two DSM-5-based parent questionnaires for the assessment of ADHD symptoms and symptoms of disruptive behavior disorders (FBB-ADHS; FBB-SSV), which were used to evaluate convergent and divergent validity. *Results:* ICC coefficients demonstrated very good to excellent interrater reliability on the item and scale level of the ILF-EXTERNAL (scale level: ICC(1,1) = .83 - .95; ICC(1,3) = .94 - .98). Overall kappa agreement on DSM-5 diagnoses was substantial to almost perfect for most disorders ($.38 \leq \kappa \leq .94$). With some exceptions, internal consistencies ($.60 \leq \alpha. \leq 86$) and item-total correlations ($.21 \leq \text{rit} \leq .71$) were generally satisfactory to good. Furthermore, higher symptom severity was associated with a higher degree of functional impairment. The evaluation of convergent validity revealed positive results regarding clinical judgment and parent ratings (FBB-ADHS; FBB-SSV). Correlations between the ILF-EXTERNAL scales and the CBCL *Externalizing Problems* were moderate to high. Finally, the ILF-EXTERNAL scales were significantly more strongly associated with the CBCL *Externalizing Problems* than with the *Internalizing Problems*, indicating divergent validity. *Conclusion:* In clinically referred, school-age children, the ILF-EXTERNAL demonstrates sound psychometric properties. The ILF-EXTERNAL is a promising clinical interview and contributes to high-quality diagnostics of externalizing disorders in children and adolescents.

Keywords: structured interview, ADHD, ODD, externalizing disorders, reliability, intraclass correlation coefficient, validity

Introduction

Structured clinical interviews are considered to be the gold standard for diagnosing mental disorders (Hoyer & Knappe, 2012; Nordgaard et al., 2013; Rettew et al., 2009). Accumulating evidence suggests that structured interviews lead to improved diagnostic accuracy and reliability (Frick et al., 2010; Leffler et al., 2015; Segal & Williams, 2014), which can in turn enhance the quality of treatment decision making (Galanter & Patel, 2005). In clinical research, structured interviews are especially used to screen participants for study inclusion or to evaluate psychotherapeutic outcomes (Hoyer & Knappe, 2012; Segal & Williams, 2014). Besides their use in research, such interviews have increasingly found their way into clinical practice as part of a comprehensive and standardized diagnostic process (Frick et al., 2010; Hoyer & Knappe, 2012; Segal & Williams, 2014). Moreover, clinicians in training can also benefit from these instruments, as they cover diagnostic criteria in a systematic manner (Frick et al., 2010; Leffler et al., 2015; Segal & Williams, 2014).

In terms of their degree of structure, clinical interviews can be classified into highly structured versus semi-structured. While the highly structured interviews require only a minimum of training, they leave little flexibility for the interviewer to explore and rate the patient's symptomatology. Typically, closed-ended questions form a dichotomous assessment, that is, a clinical symptom is either present or absent (Frick et al., 2010; Leffler et al., 2015; Segal & Williams, 2014). Examples of highly structured clinical interviews for assessing children and adolescents include the *NIMH Diagnostic Interview Schedule for Children Version IV* (NIMH DISC-IV; Shaffer et al., 2000), the *Children's Interview for Psychiatric Syndromes* which encompasses separate child (ChIPS; Weller et al., 1999b) and parent versions (P-ChIPS; Weller et al., 1999a), and the *Mini-International Neuropsychiatric Interview for Children and Adolescents* (Mini-KID; Sheehan et al., 2010). Most of these structured interviews have yet to be revised and validated for DSM-5 (American Psychiatric Association, 2013). As reviewed by Leffler et al. (2015), the current versions of these interviews show high interrater reliability (IRR) and good validity in community and clinical samples. However, such findings may also be attributable to the high degree of structure, and the inherent limited scope for interviewers to form their own clinical judgment (Leffler et al., 2015).

By comparison, semi-structured interviews allow the interviewer to inquire about symptoms, make informed judgments, and score responses in a more flexible manner (e.g. Likert-type scales). While this scoring format requires more extensive training, it can follow a dimensional approach by taking into account the severity of symptoms (Döpfner & Petermann, 2012; Frick et al., 2010; Leffler et al., 2015). Therefore, different interviewers may form

disparate judgments, which can in turn result in lower IRR compared to their highly structured counterparts. One of the most prominent semi-structured clinical interviews is the *Schedule for Affective Disorders and Schizophrenia for School Aged Children* (K-SADS; Kaufman et al., 1997), which mainly aims at an early diagnosis of affective disorders but also includes sections on other common mental disorders. Both the parents and their child can be interviewed at the same time. Different editions of the K-SADS exist and the instrument has been evaluated in a variety of populations, with overall good psychometric evidence. Available diagnostic interrater agreement on DSM-IV or DSM-5 externalizing disorders ranges from moderate to almost perfect agreement for several cross-cultural K-SADS adaptations with generally higher agreement in clinical samples (de la Peña et al., 2018; Ghanizadeh et al., 2006; Kim et al., 2004; Nishiyama et al., 2020; Ulloa et al., 2006) than in the community population (Birmaher et al., 2009; Y. L. Chen et al., 2017). Kariuki and colleagues (2018) evaluated the attention-deficit/hyperactivity disorder (ADHD) module of the K-SADS in a large community sample and obtained moderate to substantial intraclass correlation (ICC) coefficients for the subdimensions of ADHD. With regard to convergent validity, small to moderate correlations were found between clinical diagnoses and the broadband parent-rated *Child Behavior Checklist* (CBCL) questionnaire (Birmaher et al., 2009; Brasil & Bordin, 2010; Y. L. Chen et al., 2017; Kim et al., 2004). Furthermore, correlations between clinical diagnoses and the corresponding scales of the CBCL were generally higher than divergent correlations (Birmaher et al., 2009; Y. L. Chen et al., 2017). Another semi-structured interview is the *Child and Adolescent Psychiatric Assessment* (CAPA; Angold & Costello, 2000), which covers the full range of common mental disorders. With a duration of up to 120 minutes, it can be very time-consuming to administer (Leffler et al., 2015). In a test-retest study of the CAPA, ICC coefficients for DSM-III-R symptom scale scores ranged from .50 for oppositional defiant disorder (ODD) to .98 for substance abuse / dependence in self-reports of clinically referred children and adolescents (Angold & Costello, 1995). Furthermore, the CAPA interview has shown good construct validity in relation to ten formulated criteria (Angold & Costello, 2000).

Overall, semi-structured clinical interviews provide a valuable tool for diagnosing mental disorders in children and adolescents (Nordgaard et al., 2013). However, given the evolving conceptualizations of psychopathology, there is a current need for clinical interviews to meet these changing requirements. One such evolving conceptualization considers whether diagnostic domains are best characterized as discrete categories (such as in the DSM-5) or whether they should follow a dimensional approach (Coghill & Sonuga-Barke, 2012; Döpfner & Petermann, 2012). Consequently, state-of-the-art assessment instruments should have the

flexibility to allow both a categorical assessment and follow a dimensional approach which allows for varying degrees of severity and functional impairment. To the best of our knowledge, there is no diagnostic system available which meets all of the following criteria: a DSM-5-based, semi-structured clinical interview for externalizing disorders which follows both a categorical and dimensional approach by assessing symptom severity and functional impairment on a Likert scale and includes parallel parent forms with the exact same diagnostic scales.

To meet these criteria, we developed a comprehensive set of clinical parent and patient interviews *Diagnostic System of Mental Disorders in Children and Adolescents – Interview* (DISYPS-ILF; Görtz-Dorten et al., in press) which are part of the German *Diagnostic System of Mental Disorders in Children and Adolescents based on the ICD-10 and DSM-5* (DISYPS-III; Döpfner & Görtz-Dorten, 2017). Of these interviews, the *Clinical Parent Interview for Externalizing Disorders in Children and Adolescents* (ILF-EXTERNAL) covers diagnostic criteria according to the DSM-5 for the following externalizing disorders: ADHD, with the subtypes "combined type", "predominantly inattentive type", and "predominantly hyperactive-impulsive type"; ODD; conduct disorder (CD), with the specifier limited prosocial emotions; and disruptive mood dysregulation disorder (DMDD); for further details, see Materials and Methods. Besides this categorical assessment, the ILF-EXTERNAL also allows clinical symptoms to be viewed from a dimensional standpoint. A further distinguishing and novel characteristic of the ILF-EXTERNAL is that both the interview and the rating scales for parents, teachers, and patients correspond to the same diagnostic system DISYPS-III (Döpfner & Görtz-Dorten, 2017) and therefore have the exact same diagnostic scales. This allows a specific comparison of ratings of parents, teachers, and patients with clinical judgments. In addition, we sought to psychometrically evaluate this interview in a clinical sample of children with externalizing problems, as this group of children represents the prospective target group for clinical assessment using the ILF-EXTERNAL.

Currently, the ILF-EXTERNAL is being conducted in the multicenter consortium ESCAlife (ESCAlife: Evidence-Based Stepped Care of ADHD along the Lifespan). The purpose of this consortium is to evaluate adaptive interventions for patients diagnosed with ADHD, including 3-6-year-old preschool children (ESCApreschool; Becker et al., 2020), 6-12-year-old school children (ESCAschool; Döpfner et al., 2017), and 12-17-year-old adolescents (ESCAadol; Geissler et al., 2018).

The overall aim of this study is to present the newly developed clinical parent interview ILF-EXTERNAL and its psychometric properties, including (1) descriptive statistics for all

scales, (2) internal consistencies and item-total correlations, (3) IRR on the item and scale level, (4) overall agreement on DSM-5 diagnoses, (5) associations between symptom severity and the degree of functional impairment, and (6) convergent and divergent validity in a clinical sample of school-age children with ADHD symptoms.

Materials and methods

Measures

During the ESCAschool study, the below-mentioned measures were collected at several main assessment points (cf. Döpfner et al., 2017). In the present study, measures at baseline (i.e. before any intervention) were analyzed.

Clinical Parent Interview for Externalizing Disorders in Children and Adolescents (ILF-EXTERNAL)

The clinical parent interview ILF-EXTERNAL (Görtz-Dorten et al., in press) is part of the DISYPS-III (Döpfner & Görtz-Dorten, 2017). The ILF-EXTERNAL comprises a set of items, each of which explores a DSM-5 symptom criterion. Following a semi-structured approach, clinicians give their own judgment by rating each item on a 4-point Likert scale ranging from 0 (age-typical / not at all), to 3 (very much), with higher scores indicating higher symptom severity. To aid clinical judgment, a short description of the symptom severity is provided for each score. Also, an example sentence of a child's behavior representing a rating of 3 is given for each item. Item scores of 2 and higher are interpreted as clinically relevant and considered to fulfill the DSM-5 symptom criteria. The ILF-EXTERNAL consists of 18 items assessing ADHD symptoms which can be aggregated into two scales, *Inattention* (nine items) and *Hyperactivity-Impulsivity* (nine items). Together, these 18 items form the *ADHD Symptoms* scale. Additionally, five items assess functioning and psychological strain associated with ADHD symptoms and form the *ADHD Functional Impairment* scale. Moreover, the ILF-EXTERNAL consists of 36 items assessing oppositional and disruptive symptoms which are aggregated to the following scales: *ODD Symptoms* (eight items) and *CD Symptoms* (15 items), which together form the scale *ODD/CD Symptoms* (23 items). Further items form the scales *Disruptive Mood Dysregulation* (five items, three of which are also part of the *ODD Symptoms* scale) and *Limited Prosocial Emotions* (11 items). In addition, five items assess functioning and psychological strain associated with ODD and CD symptoms and form the *ODD/CD Functional Impairment* scale (see Supplemental Table 1 in the online supplementary material for a more detailed description of the items forming each scale). Scale scores are computed by

averaging the associated item scores. In the present study, the items assessing aggressive and antisocial symptoms from the age of 11 (B06 to B15) were excluded from further analyses due to an obvious floor effect, resulting in the shortened scales *CD Symptoms - short version* (five items) and *ODD/CD Symptoms - short version* (13 items).

Child Behavior Checklist for Ages 6-18 (CBCL/6-18R)

To examine convergent and divergent validity, information from the German CBCL/6-18R was used (Arbeitsgruppe Deutsche Child Behavior Checklist, 1998; Döpfner et al., 2014). Originally developed by the Achenbach group (Achenbach, 1991; Achenbach & Rescorla, 2001), the German CBCL/6-18R is a broadband questionnaire comprising 120 items developed to assess behavioral and emotional problems in children and adolescents. Parents rate their child's behavior on a 3-point scale (0 = "not true", 1 = "somewhat or sometimes true", 2 = "very true or often true"). The items form eight syndrome scales and three broadband scales (*Externalizing Problems, Internalizing Problems, Total Problems*). The German CBCL/6-18R has demonstrated at least satisfactory internal consistencies for the eight syndrome scales with slightly higher values in a large clinical sample than in a community sample (Döpfner et al., 2014). Exceptions are the scales *Thought Problems* ($\alpha < .70$ in both samples) and *Somatic Complaints* ($\alpha = .65$ in the community sample). Internal consistencies were good for the *Externalizing Problems* and *Internalizing Problems* ($\alpha > .80$) and excellent for the *Total Problems* ($\alpha > .90$) in both samples. In cross-cultural analyses, Rescorla et al. (2007) found that parents' ratings were similar across 31 societies including Germany, indicating the multicultural robustness of the CBCL. Furthermore, the configural invariance of the 8-syndrome structure of the CBCL was confirmed in large cross-cultural studies including Germany (Ivanova et al., 2019; Ivanova, Achenbach, Dumenci, et al., 2007). In the present study, the raw scale scores of the eight syndrome scales and the *Internalizing Problems* and *Externalizing Problems* were used.

Symptom Checklist for Attention-deficit/hyperactivity Disorder (FBB-ADHS)

The German *Symptom Checklist for Attention-deficit/hyperactivity Disorder* (FBB-ADHS) is part of the DISYPS-III (Döpfner & Görtz-Dorten, 2017). This questionnaire consists of 27 items which form identical scales to those in the ILF-EXTERNAL and an additional six items assessing the child's competencies. All items are rated on a 4-point Likert scale ranging from 0 ("not at all") to 3 ("very much"). Psychometric evaluations support the reliability and validity of the FBB-ADHS (Döpfner et al., 2008; Erhart et al., 2008). The present analyses

included the scales *Inattention*, *Hyperactivity-Impulsivity*, *ADHD Symptoms*, and *ADHD Functional Impairment.*

Symptom Checklist for Disruptive Behavior Disorders (FBB-SSV)

The German *Symptom Checklist for Disruptive Behavior Disorders* (FBB-SSV) is also part of the DISYPS-III (Döpfner & Görtz-Dorten, 2017). The structure and assessment are the same as outlined for the FBB-ADHS. The FBB-SSV includes 46 items which also form identical scales to those in the ILF-EXTERNAL and an additional 12 items assessing the child's competencies. Psychometric evaluations of the FBB-SSV revealed positive results regarding reliability and validity (Görtz-Dorten et al., 2014). The scales *ODD Symptoms, CD Symptoms, ODD/CD Symptoms, Disruptive Mood Dysregulation, Limited Prosocial Emotions*, and *ODD/CD Functional Impairment* were used in the present study. For the sake of consistency with the scales *CD Symptoms - short version* and *ODD/CD Symptoms - short version* of the ILF-EXTERNAL, the items assessing aggressive and antisocial symptoms from the age of 11 (B06 to B15) from the FBB-SSV were also excluded from further analyses.

Participants and procedure

Data collection was based on the ongoing ESCAschool intervention study (target N = 521), which is part of the research consortium ESCAlife and involves nine study centers located in Germany (Cologne, Essen, Göttingen, Hamm, Mainz, Mannheim, Marburg, Tübingen, Würzburg). The ESCAschool study investigates an evidenced-based, individualized, stepwise-intensifying treatment program based on behavioral and pharmacological interventions for children diagnosed with ADHD. For further details on the procedures, including inclusion and exclusion criteria, please refer to Döpfner et al. (2017). In the present study, the ILF-EXTERNAL was evaluated using ESCAschool baseline data from 474 children (age range 6 - 12 years, M = 8.9, SD = 1.5, 92 females). The assessment of the ILF-EXTERNAL baseline data is part of a screening to check the participants' eligibility for the ESCAschool study. The screening was conducted at two successive appointments no longer than eight weeks apart. During the screening, the ILF-EXTERNAL was administered to the parents and was either video- or audio-recorded. About one third of the children (32.5%) were receiving ADHD medication prior to the study. In these cases, parents were asked to describe their child's behavior with and without medication, resulting in two ratings for each item. For the present analyses, the children's symptomatology without medication was analyzed. Besides children diagnosed with ADHD, the present sample also included children who did not meet criteria for

an ADHD diagnosis (i.e. so-called screening negatives of the ESCAschool study). These screening negatives (n = 32, including 9 females) were characterized by subclinical ADHD symptomatology, which allowed us to capture the full spectrum of ADHD symptoms. Descriptive statistics (M, SD) for all ILF-EXTERNAL scales considering only the screening negatives are reported in Table 2. As can be seen, although these children did not fulfill inclusion criteria for the ESCAschool treatment study, they nevertheless exhibited symptoms of externalizing behavior problems. Clinical diagnoses of ADHD and comorbid externalizing disorders were based on the outcome of the ILF-EXTERNAL. To assess comorbid symptoms, all clinicians applied a clinical diagnostic checklist (DCL-SCREEN) from the DISYPS-III (Döpfner & Görtz-Dorten, 2017). All parents and children gave their assent and written informed consent, and each participating study site received ethical approval (Döpfner et al., 2017). Participant data are presented in Table 1.

Subsample for the analysis of interrater reliability

To obtain IRR, a subsample of 45 interviews of the ILF-EXTERNAL was chosen (for the characteristics of this subsample, see Table 1). More specifically, we empirically determined the required sample size as recommended by published guidelines on IRR studies (Kottner et al., 2011). We selected a method for sample size calculation for the ICC coefficient (Zou, 2012), which estimates the required sample (N) to achieve a reliability coefficient (ρ) that is not less than a prespecified value (ρ_0) with a prespecified assurance probability. The calculations revealed that a minimum of 42 interviews rated by two additional raters (k = 3) is required to ensure that the lower limit of a 95% one-sided confidence limit for ρ = .80 is no less than ρ_0 = .65 with 80% assurance probability based on the ICC one-way random-effects model (Zou, 2012). Subsequently, 45 interviews (five interviews from one clinician from each of the nine study sites) were randomly selected using the *select cases* function in SPSS. Inclusion criteria for the interview recordings were as follows: A video- or audio-recording had to be present for both parts of the interview, the recordings needed to have sufficient audio quality, the clinical assessment had to follow the ILF-EXTERNAL, and, if possible, both parts of the interview should be conducted by the same interviewer. If it was not possible to rate an interview recording due to violation of the inclusion criteria, another recording from the same interviewer was randomly selected. For one study site, there were only four recordings available from one interviewer; in this case, we therefore included one recording by another interviewer from the same study site. In short, the subsample to obtain IRR consists of 45 recordings of the ILF-EXTERNAL conducted by ten interviewers from nine study sites. Typically, interviewers

conducted the first part of the interview, assessing ADHD symptoms, at the first appointment and the second part, assessing ODD/CD symptoms, at the second appointment. For the ADHD part, 37 (82.2%) interviews were conducted with the mother, three (6.7%) with the father, and five (11.1%) with both parents. Similarly, for the ODD/CD part, 40 (88.9%) interviews were conducted with the mother, three (6.7%) with the father, and two (4.4%) with both parents. Regarding the duration of the interviews, the ADHD part had a mean length of 42 min (*SD* = 19 min, range 15 to 88 min) and the ODD/CD part had a mean length of 35 min (*SD* = 20 min, range 5 to 98 min). Thirty-eight interviews were video-recorded and seven were audio-recorded. Regarding ADHD diagnosis, 28 children were diagnosed with ADHD combined type, 14 children with ADHD inattentive type, two children with ADHD hyperactive-impulsive type and only one child was below the cut-off for any ADHD diagnosis. Hence, this sample does not capture the full spectrum of ADHD symptomatology but rather represents a clinical sample of different ADHD subtypes. Sample characteristics are reported in Table 1.

Interview training

All interviewers who were involved in recruiting patients for the ESCAschool study were trained psychologists or educationists with a Master's degree, PhD candidates, or in training to become a child and adolescent psychotherapist/psychiatrist. During the ESCAschool study, all interviewers received a standardized training on administering and scoring the ILF-EXTERNAL, including watching a practice video. All interviewers were encouraged to consult their supervisor if they experienced any difficulties regarding the assessment with the ILF-EXTERNAL. Furthermore, two independent raters were asked to rate a subsample of 45 recordings of the ILF-EXTERNAL to obtain IRR. Both independent raters were PhD students at the University of Cologne and were completing their training as child and adolescent psychotherapists. In addition to the ESCAschool training on the ILF-EXTERNAL outlined above, both raters participated in a one-day workshop in which they discussed the administration of the ILF-EXTERNAL, including detailed information on the scoring of each item. Both raters were then asked to independently code three practice videos randomly selected from the ESCAschool study, after which they received elaborate feedback from their supervisor and discussed potential difficulties when rating the recordings. Both raters were instructed not to discuss the interviews with each other during the rating process.

Statistical analysis

All statistical analyses were performed using SPSS Version 26 (SPSS Inc, Chicago, IL, USA) if not stated otherwise. A first check of the data revealed no considerable floor or ceiling effects of the ILF-EXTERNAL item frequencies (except for the items that had been excluded previously). If more than 10% of the items forming a particular scale were missing, this scale was not computed for the affected participant due to a possible bias of the results (Bennett, 2001). This listwise exclusion criterion was also applied to the scales of the parent questionnaire data. A summary of the valid cases for each analysis is provided in the respective tables.
Besides descriptive statistics (mean scores, standard deviations) for all ILF-EXTERNAL scales, Cronbach's alpha was computed, with values of > .70 indicating acceptable internal consistency (Nunnally, 1978). Moreover, the corrected item-total correlations were calculated, with values of > .30 considered acceptable (Field, 2018).

The ICC coefficient (McGraw & Wong, 1996; Shrout & Fleiss, 1979) was computed to assess IRR between the interviewers and both independent raters. The ICC is one of the most common metrics when assessing IRR of continuous data (Hallgren, 2012; Koo & Li, 2016; LeBreton, James & Senter, Jenell, 2008). It should be noted that different formulas exist, each involving distinct assumptions about their calculations and therefore leading to different interpretations (Koo & Li, 2016). We computed the ICC one-way random-effects, absolute agreement model for single rater/measurements ICC(1,1) as well as for measures based on a mean-rating ICC(1,3) with their 95% confidence intervals (CIs). The ICC one-way model was chosen because the physical distance between study centers prevented the same interviewer from measuring all participants which would otherwise qualify for the two-way measurement models (Koo & Li, 2016). Furthermore, we believe that the single rater/measurements model ICC(1,1) is more appropriate than average measures, given that the clinical outcome of ILF-EXTERNAL should be based on one clinician and not on the average information obtained from multiple clinicians (Koo & Li, 2016). Nevertheless, we also present average measurements ICC(1,*k*) to ensure comparability of our results across studies. We also calculated the IRR for both independent raters for all scales of the ILF-EXTERNAL using the two-way random-effects models for single ICC(2,1) and for average ICC(2,2) measurements (McGraw & Wong, 1996; Shrout & Fleiss, 1979). To interpret ICC coefficients, different benchmarks are commonly cited. Cicchetti (1994) provided the following guidelines for interpreting ICC coefficients: poor ≤ .40; fair = .41 – .59; good = .60 – .74; and excellent ≥ .75. However, other authors proposed more stringent guidelines: poor ≤ .50; moderate = .51 – .75; good = .76 – .90; and excellent ≥ .91 (Koo & Li, 2016). The results are therefore presented

using both .75 and .91 as interpretations of "excellent" reliability. Additionally, to obtain a further estimate on the degree of agreement, pairwise percent agreement was calculated based on integer scale scores (Wirtz & Caspar, 2002) using MATLAB and Statistics Toolbox Release 2018b. It should be noted that percentages of agreement do not correct for agreements that would be expected by chance and therefore, may overestimate the degree of agreement (Wirtz & Caspar, 2002).

Overall agreement on DSM-5 diagnoses was assessed using Fleiss' kappa (Fleiss, 1971), which is a statistical measure to assess agreement between multiple raters (i.e. the interviewers and both raters) on categorical variables (i.e. the presence or absence of a disorder). While Fleiss' kappa is a chance-corrected measure, it is dependent on the base rate of each disorder. Especially when the base rate of a disorder is low ($n < 10$), corresponding kappa values should only be interpreted with caution. The presence or absence of a DSM-5-based disorder was derived from the raw interview item scores by symptom counts. For example, if at least four items from the *ODD Symptoms* scale were scored with 2 or higher, these scores were considered to fulfill the diagnosis of ODD. Following common research practice, other exclusion criteria (such as not making the diagnosis of ODD in the presence of DMDD or such as only specifying limited prosocial emotions in the presence of CD) were ignored (Angold & Costello, 1995; de la Peña et al., 2018). Fleiss' kappa was calculated between the interviewers and two raters. To interpret kappa values, Landis and Koch (1977) suggested the following benchmarks: slight ≤ .20; fair = .21 – .40; moderate = .41 – .60; substantial = .61 – .80; almost perfect agreement ≥ .81.

Pearson product-moment correlations were computed between the ILF-EXTERNAL scales *ADHD Symptoms*, *ODD/CD Symptoms - short version* and the corresponding scales *ADHD Functional Impairment*, *ODD/CD Functional Impairment* in order to describe the relationship between symptom severity and the degree of functional impairment. To test for significant differences between pairs of correlations, the *cocor* software package for the R programming language (R 3.6.2) was applied (Diedenhofen & Musch, 2015). More specifically, we compared the magnitude of two dependent correlation coefficients with overlapping variables (i.e., the correlations have one variable in common) based on Steiger's (1980) modification of Dunn and Clark's (1969) *z*-transformation.

Additionally, Pearson product-moment correlations were computed between all ILF-EXTERNAL scales and the corresponding scales in the parent forms (FBB-ADHS; FBB-SSV) in order to evaluate convergent validity between clinical judgment and parent ratings. Two-sided paired samples *t*-tests were used group comparisons between the average scores of the

ILF-EXTERNAL scales and the corresponding scales of the parent forms. This analysis allowed us to investigate whether clinician-rated scale scores on the ILF-EXTERNAL differed significantly from ratings on the corresponding parent-rated scales.

To further assess convergent and divergent validity, Pearson product-moment correlations were computed between the ILF-EXTERNAL scales and the eight syndrome scales as well as the *Externalizing Problems* and *Internalizing Problems* of the CBCL/6-18R. The R *cocor* package and Steiger's test (Steiger, 1980) were again applied to compare the magnitude of two dependent correlations. In particular, we determined whether the correlations of a particular ILF-EXTERNAL scale (e.g. *Inattention*) with the CBCL/6-18R broadband scales (*Externalizing Problems* and *Internalizing Problems*) differed significantly.

Results

Scale characteristics

Table 2 summarizes the mean scores, standard deviations, internal consistencies (Cronbach's alpha) and the ranges of the item-total correlations for all ILF-EXTERNAL scales. The lowest mean score was observed for the scale *CD Symptoms - short version* (M = 0.40, SD = 0.43), and the highest mean score for the scale *Inattention* (M = 1.95, SD = 0.48). As can be expected, given the clinical sample of children with ADHD symptoms, average scale scores were generally higher on the ADHD scales than on the ODD/CD scales. Cronbach's alpha coefficients for the ILF-EXTERNAL symptom scales were generally acceptable to good, with the exception of the scale *CD Symptoms - short version* (α = .60). The scales comprising *Functional Impairment* showed questionable internal consistency for ADHD (α = .62) and very good internal consistency for ODD/CD (α = .86). Item-total correlations were generally satisfactory ($.21 \leq r_{it} \leq .71$) with some exceptions. The following items demonstrated item-total correlations below r_{it} = .30 (ADHD items: *A01 Careless, A06 Concentration, F05 Interferes with educational activities*. ODD/CD items: *B03 Cruel to animals, B05 Steals without confrontation, C04c Manipulates*). However, excluding any of these items did not noticeably change the Cronbach's alpha of the respective scales.

Interrater reliability

Table 3 presents the IRR of the ILF-EXTERNAL scales, according to the ICC one-way random-effects, absolute agreement model for single ICC(1,1) and average measures ICC(1,3), their respective 95% confidence intervals, and pairwise percent agreement. Regarding the ILF-

EXTERNAL symptom scales, all ICC(1,1) coefficients were greater than .75, indicating excellent IRR according to Cicchetti (1994) or, following a more stringent interpretation, good to excellent IRR (Koo & Li, 2016). Furthermore, all ICC(1,3) coefficients of the average measurement model were greater than .90, indicating excellent IRR of the ILF-EXTERNAL symptom scales (Cicchetti, 1994; Koo & Li, 2016). Regarding the ILF-EXTERNAL scales assessing functional impairment, both the *ADHD Functional Impairment* scale (ICC(1,1) = .89; ICC(1,3) = .96) and the *ODD/CD Functional Impairment* scale (ICC(1,1) = .92; ICC(1,3) = .97) demonstrated ICC values in the upper range, indicating very good to excellent IRR by the single and average measurement model (Cicchetti, 1994; Koo & Li, 2016). In addition, pairwise percent agreement was consistently higher than 80%, indicating high agreement between the interviewers and both raters. For the interested reader, results on the IRR on the item level are reported in the online supplementary material (Supplemental Table 1). Furthermore, we calculated the IRR for both independent raters for all scales of the ILF-EXTERNAL using the two-way random-effects models for single ICC(2,1) and for average measurements ICC(2,2) (McGraw & Wong, 1996; Shrout & Fleiss, 1979). The results show that all ICC coefficients were .90 or greater using single and average measures, indicating excellent IRR of the ILF-EXTERNAL scales (Cicchetti, 1994; Koo & Li, 2016). The results are summarized in the online supplementary material, (Supplemental Table 2).

Agreement on DSM-5 diagnoses

Table 4 presents overall agreement on DSM-5 diagnoses assessed using Fleiss' kappa values, their corresponding 95% confidence intervals, and pairwise percent agreement. Following the benchmarks of Landis and Koch (1977), diagnostic agreement ranged from fair (ADHD hyperactive-impulsive type: $\kappa = .38$), through moderate (DMDD: $\kappa = .55$), substantial (ADHD combined type: $\kappa = .71$; ADHD inattentive type: $\kappa = .71$; any ADHD: $\kappa = .74$) to almost perfect agreement (ODD: $\kappa = .82$; conduct disorder: $\kappa = .94$; with its specifier limited prosocial emotions: $\kappa = .82$). However, due to the low base rate of the diagnoses ADHD hyperactive-impulsive type ($n = 2$) and CD ($n = 6$) in the subsample, agreement on these two disorders should be interpreted with caution. In particular, pairwise percent agreement mainly seems to reflect agreement by chance. With regard to the remaining DSM-5 diagnoses, agreement could be estimated more reliably.

Correlations between ILF-EXTERNAL symptom scales and functional impairment

Regarding the association between symptom severity and the degree of functional impairment, Pearson correlations revealed a moderate to large (r = .50) association between the scales *ADHD Symptoms* and *ADHD Functional Impairment*. In turn, there was a strong positive association between the scales *ODD/CD Symptoms - short version* and *ODD/CD Functional Impairment* (r = .67). Furthermore, the scale *ADHD Symptoms* correlated significantly more strongly with the scale on functional impairment associated with ADHD than with the *ODD/CD Functional Impairment* scale (z = 3.92, p < .001). Likewise, the scale *ODD/CD Symptoms* correlated significantly more strongly with the scale on ODD/CD-related functional impairment than with the *ADHD Functional Impairment* scale (z = -5.41, p < .001).

Convergent and divergent validity

Table 5 compares the ILF-EXTERNAL scales and the corresponding scales of the parent forms (FBB-ADHS; FBB-SSV). Pearson correlations were moderate to high and significant (.57 ≤ r ≤ .78, p < .001), indicating convergent validity between clinical judgment and parent ratings. Overall, ratings on most ILF-EXTERNAL scale scores differed significantly from ratings on the corresponding scales of the parent forms (p < .05), with the exception of the scales *ADHD Symptoms* (p = .051) and *CD Symptoms - short version* (p = .260). Furthermore, mean scale scores on the parent forms were higher than the corresponding clinical judgment (exceptions: *ADHD Symptoms* and *Hyperactivity-Impulsivity*).

In addition, Table 6 summarizes Pearson correlations between the ILF-EXTERNAL scales and the eight CBCL/6-18R syndrome scales as well as the CBCL/6-18R broadband scales *Externalizing Problems* and *Internalizing Problems*. Overall, correlations between the ILF-EXTERNAL scales and the CBCL *Externalizing Problems* were moderate to high and significant (.33 ≤ r ≤ .69, p < .001). As can be expected, the highest observed correlations of the CBCL *Externalizing Problems* were with the ILF-EXTERNAL scales *ODD Symptoms*, *CD Symptoms - short version*, and *ODD/CD Symptoms - short version* scales (.58 ≤ r ≤ .69). Furthermore, the ILF-EXTERNAL *Inattention* scale was most strongly associated with the CBCL syndrome scale *Attention Problems* (r = .39). As can be expected, the ILF-EXTERNAL scales were more strongly associated with the CBCL *Externalizing Problems* than the *Internalizing Problems*. When comparing the correlation coefficients of both CBCL problem scales, we found that all ILF-EXTERNAL scales were significantly more strongly associated with the CBCL *Externalizing Problems* (.001 ≤ p ≤ .005). Taken together, these results provide support for the convergent and divergent validity of the ILF-EXTERNAL.

Discussion

This study presents the DSM-5-based, semi-structured, clinical parent interview ILF-EXTERNAL and its psychometric properties in a clinical sample of school-age children with ADHD symptoms. The results suggest that the ILF-EXTERNAL is a promising and overall reliable and valid clinical interview for diagnosing externalizing disorders in children and adolescents.

Regarding scale reliability, Cronbach's alpha coefficients for the ILF-EXTERNAL scales were generally acceptable to good. Accordingly, those items which were aggregated to form a particular scale predominantly seem to measure a common construct. One exception is the *CD Symptoms - short version* scale (α = .60). Similar internal consistency of the *CD Symptoms* scale was reported for the DISC version 2.3 (α = .59, cf. Frick et al., 2010). We believe that for the following reasons, this rather low internal consistency is unsurprising: First, we excluded the items B06 to B15, assessing aggressive and antisocial symptoms from the age of 11, which resulted in a shortened scale of only five items. Second, with a low mean score (*M* = 0.40; *SD* = 0.43), the scores of the remaining items of this shortened scale displayed a skewed distribution. Third, these symptoms represent a heterogeneous group of symptoms, which may have impaired the reliability of this scale (Frick et al., 2010). Similarly, the *ADHD Functional Impairment* scale demonstrated low internal consistency (α = .62), which might also be explained by the heterogeneity of the items. However, the *ODD/CD Functional Impairment* scale showed very good internal consistency (α = .86). In addition, item-total correlations were generally satisfactory with some exceptions. Although some items demonstrated item-total correlations below r_{it} = .30, excluding any of these items did not noticeably change the Cronbach's alpha of the respective scales.

Having calculated the ICC one-way random-effects model for single ICC(1,1) and average ICC(1,3) measurements, ICC coefficients demonstrated "very good" to "excellent" IRR for all scales (Cicchetti, 1994; Koo & Li, 2016). Most IRR studies on broadband clinical interviews assessing children and adolescents did not provide IRR results on the scale level. One previous study assessed externalizing symptoms in children and adolescents using a modified ADHD-ODD scale of the K-SADS (Jans et al., 2009). This modified scale was based on a dichotomous assessment of the DSM-IV-based ADHD and ODD criteria, leading to a sum score. Pearson correlations revealed a strong positive association (*r* = .98) between the sum scores of the interviewers and the sum scores from independent raters. However, it should be noted that ICC might be a more appropriate measure to assess IRR than Pearson correlations. While the Pearson correlation coefficient indicates the strength of the linear relationship

between two variables, a high correlation may be observed even though agreement is poor (Bland & Altman, 1986; Gisev et al., 2013). Another study assessed IRR of the ADHD subdimensions in the K-SADS using ICC (Kariuki et al., 2018). The results indicated moderate to good IRR for the inattentive subtype (ICC = .76), hyperactive-impulsive subtype (ICC = .41), combined type (ICC = .77), and any ADHD type (ICC = .64). While the authors calculated the one-way random-effects model, it remains unclear whether they relied on single or average measurements, which limits the interpretation of their results. Although our ICC coefficients were consistently higher on all ADHD scales, a comparison with the aforementioned study must be treated with caution for the following reasons: First, the authors validated the ADHD subdomains in a community sample, while our results were based on clinically referred children. Second, the authors only obtained IRR estimates from 20 children, while we empirically calculated our required sample size and based our IRR results on twice as many children. Overall, our study demonstrates high IRR and addresses the aforementioned research gap, providing valuable information regarding the psychometric quality on the scale level. These findings were largely confirmed even on the single-item level (see online supplementary material, Supplemental Table 1).

Diagnostic agreement between the interviewers and both independent raters was "substantial" to "almost perfect" for most disorders with the exceptions of ADHD hyperactive-impulsive type and DMDD (Landis & Koch, 1977). With regard to diagnosing ADHD and its subtypes, we found substantial agreement for any ADHD diagnosis, for ADHD combined type, and for ADHD inattentive type. However, these results should be discussed within the scope of the subsample. The composition of this subsample may have influenced agreement estimates, particularly because of the high base rate of ADHD diagnoses (i.e. 44/45 children). Although both independent raters were not aware of this high base rate, the sole fact that almost all children exhibited clinically relevant symptoms (i.e. scorings of 2 or 3 on each item) may have led to an uneven distribution of item scorings and thus, possible overestimation of agreement on ADHD diagnoses. For example, the "perfect" pairwise agreement of 100% for any ADHD diagnosis (κ = .74) rather seems to reflect an overestimation of agreement due to sampling issues. Concerning diagnostic agreement on ADHD hyperactive-impulsive type, we found rather low Fleiss' kappa agreement (κ = .38) but almost perfect pairwise agreement (95.6%). Although this finding might seem somewhat perplexing, it can be explained as follows: Considering that Fleiss' kappa is influenced by the base rate of observations (Wirtz & Caspar, 2002), the agreement on ADHD hyperactive-impulsive type seems to primarily reflect sampling issues due its very low base rate (n = 2) in our subsample. This low base rate, in turn, influences

pairwise percent agreement which does not correct for agreement that would be expected by chance. For example, even if both raters agreed on *no* ADHD hyperactive-impulsive diagnosis for all 45 participants, they still would have demonstrated agreement in 43/45 cases.

As a newly developed clinical interview with a semi-structured format, it is particularly essential to compare diagnostic interrater agreement of the ILF-EXTERNAL with that from other semi-structured interviews. The degree of agreement on any ADHD diagnosis was comparable with other findings in clinical samples using the K-SADS ($.42 \leq \kappa \leq .92$; de la Peña et al., 2018; Ghanizadeh et al., 2006; Kim et al., 2004; Nishiyama et al., 2020; Ulloa et al., 2006). Furthermore, our results regarding diagnostic agreement on ADHD subtypes were also relatable to previous literature. Having calculated kappa agreement using the MINI-KID interview in a clinical sample, Sheehan et al. (2010) reported almost perfect agreement for ADHD combined type ($\kappa = .90$) and ADHD inattentive type ($\kappa = .93$) and substantial agreement for ADHD hyperactive-impulsive type ($\kappa = .65$). Interestingly, high diagnostic agreement on diagnosing ADHD combined type ($\kappa = .86$) and ADHD inattentive type ($\kappa = .78$) was also reported in a clinical sample of children with ADHD symptoms (Power et al., 2004).

With regard to comorbid externalizing disorders, the degree of diagnostic agreement was comparable with other findings in clinical samples using the K-SADS for ODD ($.69 \leq \kappa \leq .80$; de la Peña et al., 2018; Ghanizadeh et al., 2006), DMDD ($\kappa = .53$; de la Peña et al., 2018), and CD ($.78 \leq \kappa \leq 1.0$; de la Peña et al., 2018; Ghanizadeh et al., 2006; Ulloa et al., 2006). Although our results concerning CD should be interpreted with caution due to its low base rate in the subsample ($n = 6$), these results were also in line with previous studies reporting the highest agreement on this diagnosis (Ghanizadeh et al., 2006; Ulloa et al., 2006). We suggest that this finding may be attributable to the clinical presentation of CD symptoms, which are clear to observe and unambiguous to score. Agreement on the specifier limited prosocial emotions was classified if symptoms in at least two out of four categories were considered as clinically relevant. While previous research observed fair agreement ($\kappa = .29$; de la Peña et al., 2018), we found very high diagnostic agreement on this specifier ($\kappa = .82$), which again, may be attributable to our sample characteristics.

The ranges of diagnostic agreement reported in the literature might arise from differences in the administration of the interview (e.g. parents or children as primary informant), the respective study samples (e.g. children or adolescents), methodological issues (e.g. number of raters or amount of training received on administering the interview), or the sample population (community vs. clinical) and its characteristics (e.g. base rates of disorders). Notably, diagnostic agreement is often higher in clinical than in community samples (Y. L.

Chen et al., 2017). One basic criticism of clinical samples is that they typically only include patients with clear and severe symptoms. Consequently, the patients' symptoms can be easily recognized and scored, which may lead to overestimated reliability results, an effect which is also referred to as spectrum bias (Ranshoff & Feinstein, 1978).

Overall, while these reliability results and their corresponding coefficients yield important empirical findings, these labels do not indicate their practical or clinical relevance (Kottner et al., 2011). In other words, even though we obtained very good to excellent IRR and diagnostic agreement results, discrepancies between ratings nevertheless occurred, which warrant further discussion. We critically explored discrepancies between the interviewers and both raters and propose the following reasons for rater disagreement: 1) In terms of the administration of the ILF-EXTERNAL, we noted that some interviewers explored the frequency and intensity of each symptom more thoroughly than did others. This possible lack of clinical information may have affected the scorings of both independent raters. Moreover, 2) noise disturbances during the recordings may have affected the raters, and 3) information variance (i.e. the interviewers may have integrated information prior to the interview into their ratings), as well as 4) interpretation variance (i.e. different raters may have subjective ideas about weighting of symptoms) might have arisen (see also Hoyer & Knappe, 2012).

A further finding was that higher symptom severity was associated with a higher degree of functional impairment. This result highlights the importance of the current DSM practice of considering a clinical significance criterion (Spitzer & Wakefield, 1999), which requires symptoms to be associated with clinically significant psychological strain and functional impairment in social, occupational, or other areas of life to warrant a diagnosis (American Psychiatric Association, 2013). Results from a large meta-analysis confirmed the relationship between ADHD subtypes and multiple domains of functional impairment (Willcutt et al., 2012).

Regarding convergent and divergent validity, we found moderate to strong correlations between the ILF-EXTERNAL scales and the scales of the German CBCL/6-18R covering similar symptoms. Furthermore, the ILF-EXTERNAL scales were significantly more strongly associated with the CBCL *Externalizing Problems* than with the *Internalizing Problems*, indicating construct validity. These results are largely consistent with previous studies reporting small to moderate relations between the CBCL and clinical diagnoses from semi-structured interviews for the assessment of clinical symptoms in children and adolescents (Birmaher et al., 2009; Brasil & Bordin, 2010; Y. L. Chen et al., 2017; Kim et al., 2004). Moreover, correlations of the ILF-EXTERNAL scales with the corresponding CBCL scales were generally higher than correlations with the non-corresponding CBCL scales. Similar findings have also

been reported in the community population (Birmaher et al., 2009; Y. L. Chen et al., 2017). However, limitations of these findings are that they often rely solely on broad diagnostic categories such as "ADHD" without specification of its subtypes (Birmaher et al., 2009; Y. L. Chen et al., 2017; Kim et al., 2004), "any disruptive disorder" (Brasil & Bordin, 2010), or that their results are based on small (i.e. less than $N = 100$) sample sizes (Brasil & Bordin, 2010; Kim et al., 2004). We therefore extended these findings by reporting validity results on diagnostic scales in a larger sample. A further strength of our study is that we included parent forms (FBB-ADHS; FBB-SSV) which cover the same DSM-5 symptoms as the ILF-EXTERNAL. This distinguishing and novel characteristic allowed us to specifically compare ratings between parental and clinical judgments. While our results indicate moderate to substantial convergence between parent ratings and clinical judgments, we believe that this convergence is not sufficiently strong to argue that raters could be seen as interchangeable. In contrast, Boyle et al. (2017) challenged that structured clinical interviews may be replaced by self-completed problem checklists as a time- and cost-effective alternative. One basic criticism was that *"the dependence on respondents in these interviews is similar to the dependence on respondents completing a checklist on their own except for the potential error introduced by interviewer characteristics and interviewer–respondent exchanges"* (Boyle et al., 2017, p. 2). While we agree with this view inasmuch as clinical interviews should provide additional value to questionnaire data such as problem checklists, close inspection of our results revealed the following: Although we found moderate to large correlations between clinician and parent ratings, comparisons of the absolute scale scores revealed significant differences between the ratings on several scales. This indicates that both perspectives are complementary and that both are necessary for an informed clinical diagnosis. On top of that, similar recommendations are made by the German interdisciplinary evidence- and consensus-based (S3) guidelines on the clinical assessment of ADHD (Association of Scientific Medical Societies in Germany AWMF, 2018).

In terms of limitations, one drawback of the present study is that parents were the only informants for both the interview and the questionnaires. Hence, no information was available from the children themselves. However, we believe that this limitation is surmountable given that parents are typically better informants regarding their children's externalizing behavior problems than their children.

Another significant aspect to consider is the composition of the subsample for the analysis of IRR. We concede that the high base rate of ADHD diagnoses may have influenced interrater agreement. As percentages agreement do not correct for agreements that would be

expected by chance, they may overestimate the degree of agreement. In particular, the almost perfect percentages of agreement on some diagnoses rather seem to reflect an overestimation due to chance agreement and sampling issues.

While agreement between parent and teacher ratings on childhood diagnoses is typically quite low (Willcutt et al., 2012), studies investigating interrater agreement between interviewers using clinical interviews yield higher estimates. We concede that these higher agreement estimates may be explained as follows: 1) Intensive rater trainings on the administration and scoring of a clinical interview may lead to more homogenous ratings, and thus, higher rates of agreement. 2) Within research settings, it is common practice to classify agreement on diagnoses based on raw interview item scores by symptom counts. However, this approach may overestimate diagnostic agreement because additional criteria for an informed clinical diagnosis are not further considered. 3) As the interviews are video- or audio-recorded, the interviewers and raters have the exact same informants (e.g. parents) with the exact same information. This approach results in higher agreement estimates compared to other forms of reliability, e.g. test-retest reliability where the same informant is interviewed twice but may provide different information (Angold & Costello, 1995).

Finally, the factor structure of the ILF-EXTERNAL has not yet been validated. While this clinical interview comprises a set of items with each item exploring a DSM-5 symptom criterion, it remains unclear whether this DSM-5-based factor structure can be replicated empirically. For this reason, a follow-up study exploring the factor structure of the ILF-EXTERNAL using correlated factor models and bifactor models is planned. Nevertheless, it should be noted that the factor structure of the corresponding DISYPS parent forms, FBB-ADHS and FBB-SSV, has been confirmed (Erhart et al., 2008; Görtz-Dorten et al., 2014).
We suggest that future studies evaluating psychometric properties of structured clinical interviews should include ratings of symptom severity on the scale level as part of a dimensional approach. Ideally, specific aspects covering functioning and psychological strain could also be included.

Conclusions

The aim of this study was to assess the reliability and validity of a DSM-5-based, semi-structured parent interview for diagnosing externalizing disorders in children and adolescents. In clinically referred, school-age children, the ILF-EXTERNAL demonstrates sound psychometric properties in terms of IRR on the item and on the scale level, rater agreement on most DSM-5 diagnoses, internal consistency, and convergent and divergent validity. In line

with current literature and the DSM practice to consider functional impairment as prerequisite for making a diagnosis, higher symptom severity was associated with a higher degree of functional impairment. Having developed a comprehensive set of clinical parent and patient interviews (DISYPS-ILF), we hope to contribute to a high-quality standard of diagnosing mental disorders in children and adolescents.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Ethical approval has been obtained for the study centre Cologne by the University of Cologne (ID 15–216), for the study centre Essen by the University Duisburg-Essen (ID 17–7404-BO), for the study centre Hamm by the Ruhr- University Bochum (ID 15–5564), for the study centre Göttingen by the University Medical Center Göttingen (ID 3/3/17), for the study centre Mainz by the Federal Medical Association of Rhineland-Palatinate (ID 837.237.17 [11071]), for the study centre Mannheim by the Ruprecht-Karls- University Heidelberg (ID 2015-646 N-MA), for the study centre Marburg by the Philipps-University Marburg (ID Studie 03/16), for the study centre Tübingen by the Eberhard-Karls-University Tübingen (ID 791/2015BO2), and for the study centre Würzburg by the Julius-Maximilians-University Würzburg (ID 332/15_z). Written informed consent to participate in this study was provided by the participants legal guardian/next of kin.

Author contributions

A-KTh developed the first draft of the manuscript, rated the data (Rater 1), analyzed the data, and was involved in the recruitment and data acquisition of the ESCAschool study site in Cologne and was co-author of the DISYPS-ILF. MD was principal investigator of the ESCAschool trial, developed the basic ESCAschool study design, was part of the ESCAlife-Consortium, was head of the Cologne recruiting center for the ESCA children trials (ESCApreschool, ESCAschool, and ESCAadol), developed the DISYPS-III system, and the DISYPS- ILF, and critically revised the manuscript. AG-D developed the DISYPS-III system and the DISYPS-ILF, was involved in rater training of the ESCAschool study, and critically revised the manuscript. PA was a scientific staff member at the Cologne study site, was involved in patient recruitment and data acquisition, and critically revised the manuscript. CD heads the telephone-assisted self-help trial performed in one treatment arm of the ESCAschool study and critically revised the manuscript. NG rated the data (Rater 2) and critically revised the manuscript. CH was involved in the development of the ESCAschool research proposal and organization of the study, and critically revised the manuscript. LJ and A-KTr coordinated ESCAschool, contributed to the management and organization of the study, and critically revised the manuscript. EW coordinated the Cologne study site, contributed to the implementation of the study, provided supervision for patient treatment, and critically revised the manuscript. TB coordinated the ESCAlife consortium and was the co-principal investigator in the ESCAschool study, and made substantial contributions to the conception and design of the ESCAschool research proposal, heads the Mannheim study site of ESCAschool, and critically revised the manuscript. DB made substantial contributions to the conception and design of the ESCAschool research proposal and the neuropsychological research battery, heads the sub-project ESCAbrain, and critically revised the manuscript. SM heads the

Mannheim study site of ESCAschool, coordinates the ESCAlife consortium, made substantial contributions to the conception of the neuropsychological research battery and the neurofeedback protocol, and critically revised the manuscript. SH coordinated the ESCAlife consortium, made substantial contributions to the conception of the neuropsychological research battery and the neurofeedback protocol, and critically revised the manuscript. KB heads the Marburg study site of ESCAschool, was PI for the ESCApreschool study, is a member of the ESCAlife consortium, was a co-applicant of the ESCAlife research project, made substantial contributions to the conception and design of the ESCAlife study, and critically revised the manuscript. JK coordinated ESCAschool at the Marburg study site, contributed to the management and organization of the study, was involved in patient recruitment and data acquisition, and critically revised the manuscript. JH heads the Essen study site of ESCAschool, made substantial contributions to the conception and design of the ESCAschool study, and critically revised the manuscript. JW coordinated ESCAschool at the Essen study site, contributed to the management and organization of the study, and critically revised the manuscript. MHo heads the Hamm/Bochum study site of ESCAschool, was involved in the development of the ESCAschool research proposal, and critically revised the manuscript. TL heads the research department of the Hamm/Bochum study site of ESCAschool, contributed to the implementation ofthe study in Hamm, supervised the realization of the trial in Hamm, and critically revised the manuscript. MHu heads the Mainz study site of ESCAschool, was involved in the implementation of the ESCAlife projects, and critically revised the manuscript. MR was involved in the planning of the ESCAlife research projects and application for funding, was co- PI for the ESCAadol trial, was involved in the implementation of the ESCAlife projects at the Würzburg study site, and critically revised the manuscript. TJ was involved in the planning of the ESCAlife research projects and application for funding, was co-PI for the ESCAadol trial and was involved in the implementation of the ESCAlife projects at the Würzburg study site, and critically revised the manuscript. JG coordinated the Würzburg study site, contributed to the management and organization of the ESCAlife projects, and critically revised the manuscript. LP heads the Göttingen study site of ESCAschool, was involved in the implementation of the ESCAlife projects, and critically revised the manuscript. HU coordinated ESCAschool at the Göttingen study site, contributes to the management and organization of the study, was involved in patient recruitment and data acquisition, and critically revised the manuscript. TR heads the Tübingen study site of ESCAschool, contributed to the implementation of the ESCAschool study, and critically revised the manuscript. UD coordinates the Tübingen study site of ESCAschool, contributes to the management and organization of the study, was involved in patient recruitment and data acquisition, and critically revised the manuscript. All authors gave final approval of the last version of the manuscript and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding

The ESCAschool study was funded by the German Federal Ministry of Education and Research (BMBF) Grant 01EE1408. This work was supported by the research consortium on ADHD, ESCA-Life, funded by the German Federal Ministry of Education and Research (FKZ 01EE1408E).

Acknowledgments

We thank all staff members who are participating in the ESCA-school study in nine different study recruiting centers and who are involved in recruitment, study organization and management, blind rating, diagnostics and therapy as well as the ESCAbrain study group and the CTU. These are (sorted by recruiting center affiliation and in alphabetical order): Study

center Cologne: Ruth Beckmann, Caroline Betcher, Jana Bretthauer, Corinna Broekmans, Yvonne Buntrock, Mareike Bürger, Dominique Deli, Elvan Dogan, Katrin Floß, Anna Fronhofen, Claudia Ginsberg, Jana Harenkamp, Julia Kellner, Claudia Kinnen, Marie-Theres Klemp, Hedda Körner, Constanze Kitzig, Katrin Krugmann, Jessica Leite, Sarah Meier, Judith Mühlenmeister, Lea Ostrowski, Imke Partzsch, Jeannine Pawliczek, Daniela Perri, Judith Ratayczak, Stephanie Schürmann, Ellen Settegast, Johanna Spicher, Susanne Stollewerk, Laura Wähnke, Annabelle Warschburger, Friederike Waschau, Tanja Wolff Metternich-Kaizman, Sara Zaplana Labarga. Study center Essen: The authors would like to gratefully thank all the participating patients and their families. Study center Göttingen: Fatmeh Al-Debi, Jan Schulz. Study center Hamm: Miriam Davids, Natalie Deux, Mareen Dökel, Silvia Eißing, Regina Herdering, Carina Huhn, Lara Kaffke, Inken Kirschbaum-Lesch, Franziska Martin, Nina Müller, Nicola Nolting, Daniela Pingel, Carina Schulz, Karen Schumann, Birthe Wagner, Daniela Walkenhorst. Study center Mainz: The authors would like to gratefully thank all the participating patients and their families. Study center Mannheim: Karina Abenova, Brigitta Gehrig, Monja Groh, Christine Igel, Anna Kaiser, Theresa Nickel, Angelina Samaras, Marie-Therese Steiner, Nina Trautmann, Matthias Winkler, Mirjam Ziegler. Study center Marburg: Anette Becker, Viktoria Birkenstock, Doreen Blume, Anita Dehnert, Silas Deutsch, Sabine Finkenstein, Charlotte Finger, Claudia Freitag, Nicole Grau, Wiebke Haberhausen, Julia Häusser, Daria Kasperzack, Franziska Körfgen, Johannes Kresse, Jana Langhammer, Christopher Mann, Tanja Mingebach, Thomas Mooz, Jens Pfeiffer, Maren Rumpf, Alisa Samel, Bastian Schrott, Anna Schuh, Isabell Schulz, Thomas Stehr, Rebecca Stein, Anne von Stiphout, Linda Weber, Svenja-Katharina Weber, Anne-Kathrin Wermter, Maximiliane Werther. Study center Tübingen: Yael Amling, Anna Hertle, Natalie Herrmann, Melinda Mross, Anja Pascher, Alena Roth, Priska Schneider, Anja Schöllhorn, Ida Steinacker. Study center Würzburg: Nadja Becker, Christoph Biohlawek, Lea Brosig, Kathrin Delius, Adrian Dernbach, Anja Diehl, Simone Dorsch, Lisa Eidenschink, Stefanie Fekete, Zuzana Fouskova, Julia Gläser, Jana Halmagyi, Verena Hartlieb, Silke Hauck, Julia Häusler, Christin Heim, Sonja Hetterich, Michaela Linke, Annette Nowak, Corinna Otte, Isabel Paul, Katharina Peters, Hanna Schwarz, Carolin Steinmetz, Kristin Wehrmann, Johannes Weigl, Lea Zwilling. Additional ESCAbrain staff members, who are involved in ESCA-school: Clinical Trials Unit Freiburg: Veronika Frommberger, Alexander Hellmer, Lydia Herbstritt, Maria Huber, Carolin Jenkner, Stefanie Lade, Patrick Müller, Ralf Tostmann, Barbara Schilling, Carla Schneider. Further thanks to Sarah Mannion for English-language proofreading.

Conflict of interest

A-KTh, AG-D, and MD were involved in the development of the DISYPS-ILF and will receive royalties after publication of this instrument from the publisher Hogrefe. AG-D and MD are AKiP supervisors and lecturers and received income as heads of the School for Child and Adolescent Behavior Therapy at the University of Cologne and royalties from treatment manuals, books and psychological tests published by Guilford, Hogrefe, Enke, Beltz, and Huber. MD received consulting income and research support from Lilly, Medice, Shire, Janssen Cilag, Novartis, and Vifor. AG-D is head of the AKiP Research and Evaluation department. TB served in an advisory or consultancy role for Lundbeck, Medice, Neurim Pharmaceuticals, Oberberg GmbH, Shire, and Infectopharm; and received conference support or speaker's fees from Lilly, Medice, and Shire; and received royalties from Hogrefe, Kohlhammer, CIP Medien, and Oxford University Press. DB served as an unpaid scientific advisor for an EU-funded neurofeedback trial unrelated to the present work. KB has been involved in research/clinical trials with Eli Lilly (≤2011) and Shire (2010), was on the Advisory Board of Eli Lilly/Germany (≤2014), a member of the Scientific Committee of Shire (≤2012) and was paid for public speaking by Eli Lilly (<2011) and Shire. These activities do not bias the objectivity of this manuscript (in her opinion), but are mentioned for the sake of completeness. MHo served in an

advisory role for Shire and Medice and received conference attendance support or was paid for public speaking by Medice, Shire and Neuroconn. He receives research support from the German Research Foundation and the German Ministry of Education and Research. He receives royalties as editor in chief of the German Journal for Child and Adolescent Psychiatry and for text books from Hogrefe. MHu served as a member of the advisory boards of Eli Lilly and Co., Engelhardt Arzneimittel, Janssen-Cilag, Medice, Novartis, Shire, and Steiner Arzneimittel within the past 5 years; served as a consultant to Engelhardt Arzneimittel, Medice, and Steiner Arzneimittel; received honoraria from Eli Lilly and Co., Engelhardt Arzneimittel, Janssen-Cilag, Medice, Novartis, and Shire; and received unrestricted grants for investigator- initiated trials from Eli Lilly and Co., Medice, Engelhardt Arzneimittel, and Steiner Arzneimittel. LP served in an advisory or consultancy role for Shire, Roche and Infectopharm; and received speaker's fees from Shire and royalties from Hogrefe, Kohlhammer, and Schattauer. HU served in an advisory or consultancy role for Medice; and received speaker's fees from Shire and Medice.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Table 1

Sample characteristics

	Subsample for the analysis of interrater reliability (*n* = 45)	Total sample (*N* = 474)
Age: mean *(SD)*	9.2 (1.6)	8.9 (1.5)
Male: *n* (%)	34 (75.6)	382 (80.6)
Diagnosis *n* (%)		
No ADHD diagnosis	1 (2.2)	32 (6.8)
ADHD – combined type	28 (62.2)	208 (43.9)
ADHD – predominantly inattentive type	14 (31.1)	184 (38.8)
ADHD – predominantly hyperactive-impulsive type	2 (4.4)	50 (10.5)
Comorbidities *n* (%)		*n* = 454 - 465
Internalizing disorders:		
- Anxiety	2 (4.4)	29 (6.4)
- Depression	2 (4.4)	15 (3.1)
Externalizing disorders:		
- Oppositional defiant disorder	23 (51.1)	166 (36.6)
- Disruptive mood dysregulation disorder	6 (13.3)	40 (8.8)
- Conduct disorder	5 (11.1)	28 (6.2)
Other disorders:		
- Obsessive-compulsive disorder	1 (2.2)	2 (0.4)
- Tic disorder	4 (8.9)	24 (5.2)
- Autism spectrum disorder	0	2 (0.4)
Medication *n* (%)		*n* = 462
ADHD medication	17 (37.8)	150 (32.5)
Parents' primary language *n* (%)		*n* = 458
German	42 (93.3)	429 (93.7)
Highest parents' graduation *n* (%)		*n* = 455
Higher-track school	24 (53.3)	261 (57.4)
Vocational school	2 (4.4)	26 (5.7)
Medium-track school	14 (31.1)	123 (27.0)
Lower-track school	4 (8.9)	43 (9.5)

Note. Clinical diagnoses of ADHD and comorbid externalizing disorders were based on the semi-structured clinical interview ILF-EXTERNAL conducted with the parents. Further comorbid symptoms were assessed using a clinical diagnostic checklist. ADHD = attention-deficit/hyperactivity disorder.

Table 2

Scale characteristics, Cronbach's alpha (α) and range of item-total correlations of the ILF-EXTERNAL

	Total sample (N = 474)					Screening negatives (n = 32)	
Scale	*k* (items)	α	Item-total *r* (range)	Mean (*SD*)	*N*	Mean *(SD)*	*n*
ADHD Symptoms	18	.84	.21 - .62	1.80 (0.50)	474	1.09 (0.51)	32
- Inattention	9	.71	.29 - .49	1.95 (0.48)	474	1.35 (0.49)	32
- Hyperactivity-Impulsivity	9	.87	.50 - .68	1.64 (0.73)	474	0.84 (0.62)	32
ADHD Functional Impairment	5	.62	.24 - .48	1.61 (0.59)	472	1.03 (0.49)	31
ODD/CD Symptoms - short version	13	.84	.25 - .62	0.85 (0.50)	450	0.55 (0.55)	23
- ODD Symptoms	8	.82	.41 - .61	1.12 (0.63)	451	0.78 (0.68)	24
- CD Symptoms - short version	5	.60	.22 - .47	0.40 (0.43)	450	0.22 (0.37)	23
Disruptive Mood Dysregulation	5	.83	.53 - .67	1.08 (0.73)	452	0.73 (0.65)	24
Limited Prosocial Emotions	11	.77	.23 - .58	0.50 (0.42)	444	0.37 (0.30)	22
ODD/CD Functional Impairment	5	.86	.61 - .71	0.93 (0.78)	442	0.49 (0.73)	21

Note. Screening negatives are participants who have been screened for eligibility of the ESCAschool study and are characterized by subclinical ADHD symptoms. ADHD = attention-deficit/hyperactivity disorder; CD = conduct disorder; *k* = number of items which form a particular scale; ODD = oppositional defiant disorder.

Table 3

Interrater reliability of the ILF-EXTERNAL scales

Scale	ICC(1,1)	95% CI	ICC(1,3)	95% CI	Pairwise percent agreement	*n*
ADHD Symptoms	.91	.87 - .95	.97	.95 - .98	88.1	45
- Inattention	.83	.74 - .90	.94	.89 - .96	85.2	45
- Hyperactivity-Impulsivity	.95	.91 - .97	.98	.97 - .99	82.2	45
ADHD Functional Impairment	.89	.82 - .94	.96	.93 - .98	80.7	39
ODD/CD Symptoms - short version	.94	.90 - .96	.98	.97 - .99	91.1	45
- ODD Symptoms	.94	.90 - .96	.98	.96 - .99	83.7	45
- CD Symptoms - short version	.90	.85 - .94	.97	.94 - .98	88.2	44
Disruptive Mood Dysregulation	.90	.85 - .94	.97	.94 - .98	83.7	45
Limited Prosocial Emotions	.93	.89 - .96	.98	.96 - .99	86.7	41
ODD/CD Functional Impairment	.92	.86 - .96	.97	.95 - .99	85.2	31

Note. ADHD = attention-deficit/hyperactivity disorder; CD = conduct disorder; CI = confidence interval; ICC = intraclass correlation; ICC(1,1) = one-way random-effects, absolute agreement model for single rater/measurements; ICC(1,3) = one-way random-effects, absolute agreement model based on a mean-rating; ODD = oppositional defiant disorder.

Table 4

Agreement on DSM-5 diagnoses in the subsample for the analysis of interrater reliability

DSM-5 Diagnosis	Fleiss' Kappa	95% CI	Pairwise percent agreement	Base rate *n*
Any ADHD	.74	.74 - .75	100	44
ADHD: combined type	.71	.70 - .71	86.7	29
ADHD: predominantly inattentive type	.74	.74 - .75	89.6	13
ADHD: predominantly hyperactive-impulsive type[a]	.38	.37 - .38	95.6	2
Oppositional defiant disorder	.82	.82 - .83	91.1	24
Disruptive mood dysregulation disorder	.55	.54 - .55	88.2	10
Conduct disorder[a]	.94	.94 - .95	98.5	6
- Specifier: Limited prosocial emotions	.82	.81 - .82	91.1	18

Note. The presence of a disorder was derived from the raw interview item scores by symptom counts.

[a] Diagnostic agreement should be interpreted with caution due to a low base rate ($n < 10$). Sample size $n = 45$: ADHD = attention-deficit/hyperactivity disorder.

Table 5

Comparisons of the ILF-EXTERNAL scales and the corresponding parent forms (FBB-ADHS, FBB-SSV)

Scale	ILF-EXTERNAL: Mean (SD)	FBB (parents): Mean (SD)	*r*	Paired samples *t*-test	*p*	*n*
ADHD Symptoms	1.83 (0.48)	1.79 (0.56)	.69*	$t(425) = 1.95$	.051	426
- Inattention	1.98 (0.45)	2.03 (0.57)	.58*	$t(420) = -2.35$	.019	421
- Hyperactivity-Impulsivity	1.67 (0.71)	1.59 (0.73)	.78*	$t(422) = 3.78$	< .001	423
ADHD Functional Impairment	1.62 (0.57)	1.74 (0.69)	.59*	$t(400) = -4.16$	< .001	401
ODD/CD Symptoms - short version	0.84 (0.49)	0.97 (0.54)	.74*	$t(411) = -6.84$	< .001	412
- ODD Symptoms	1.12 (0.63)	1.38 (0.70)	.72*	$t(404) = -10.21$	< .001	405
- CD Symptoms - short version	0.40 (0.43)	0.42 (0.44)	.65*	$t(407) = -1.18$	.260	408
Disruptive Mood Dysregulation	1.06 (0.72)	1.21 (0.72)	.67*	$t(411) = -4.72$	< .001	412
Limited Prosocial Emotions	0.49 (0.41)	0.68 (0.54)	.63*	$t(406) = -8.96$	< .001	407
ODD/CD Functional Impairment	0.94 (0.78)	1.39 (0.82)	.57*	$t(380) = -11.75$	< .001	381

Note. ADHD = attention-deficit/hyperactivity disorder; CD = conduct disorder; ODD = oppositional defiant disorder; ILF-EXTERNAL = Clinical Parent Interview for Diagnosing Externalizing Disorders in Children and Adolescents; FBB = parent-rated symptom checklists for the assessment of ADHD symptoms and symptoms of disruptive behavior disorders.
* $p < .001$.

Table 6

Correlations of the ILF-EXTERNAL scales and the Child Behavior Checklist (CBCL/6-18) syndrome scales

	CBCL/6-18											
Scale	Anxious / Depressed	Withdrawn / Depressed	Somatic Complaints	Social Problems	Thought Problems	Attention Problems	Rule-Breaking Behavior	Aggressive Behavior	**Externalizing Problems**	**Internalizing Problems**	Externalizing vs. Internalizing Problems	
											z	*p*
ADHD Symptoms	.21***	.02	.12*	.37***	.26***	.31***	.38***	.50***	.49***	.17***	6.65	< .001
-Inattention	.16***	.16***	.14**	.29***	.21***	.39***	.29***	.32***	.33***	.19***	2.79	.005
-Hyperactivity-Impulsivity	.18***	-.07	.08	.31***	.21***	.17***	.32***	.47***	.44***	.10*	6.90	< .001
ODD/CD Symptoms - short version	.22***	.12*	.15**	.35***	.21***	.21***	.59***	.68***	.69***	.21***	11.17	< .001
-ODD Symptoms	.24***	.13**	.15**	.35***	.21***	.21***	.52***	.64***	.64***	.23***	9.27	< .001
-CD Symptoms - short version	.11*	.06	.08	.24***	.14**	.16**	.56***	.54***	.58***	.11*	10.03	< .001
Disruptive Mood Dysregulation	.26***	.15**	.13**	.32***	.21***	.18***	.39***	.59***	.55***	.24***	6.69	< .001
Limited Prosocial Emotions	.20***	.21***	.10*	.29***	.21***	.25***	.44***	.43***	.46***	.22***	4.99	< .001

Note. sample size $n = 407$; ADHD = attention-deficit/hyperactivity disorder; CD = conduct disorder; ODD = oppositional defiant disorder; ILF-EXTERNAL = Clinical Parent Interview for Diagnosing Externalizing Disorders in Children and Adolescents.
*$p < .05$; **$p < .01$; ***$p < .001$

4 Disentangling symptoms of externalizing disorders in children using multiple measures and informants

Thöne, A.-K., Junghänel, M., Görtz-Dorten, A., Dose, C., Hautmann, C., Jendreizik, L. T., Treier, A.-K., Vetter, P., von Wirth, E., Banaschewski, T., Becker, K., Brandeis, D., Dürrwächter, U., Geissler, J., Hebebrand, J., Hohmann, S., Holtmann, M., Huss, M., Jans, T., Ketter, J., Legenbauer, T., Millenet, S., Poustka, L., Renner, T., Romanos, M., Uebel-von Sandersleben, H., Wenning, J., Ziegler, M., & Döpfner, M. (2021). Disentangling symptoms of externalizing disorders in children using multiple measures and informants. *Psychol. Assess.*, *33*(11), 1065-1079. https://doi.org/10.1037/pas0001053

Abstract

The trait impulsivity theory suggests that a single, highly heritable externalizing liability factor, expressed as temperamental trait impulsivity, represents the core vulnerability for externalizing disorders. The current study sought to test the application of latent factor models derived from this theory to a clinical sample of children. Participants were 474 German children (age 6 - 12 years, 81% male) with symptoms of attention-deficit/hyperactivity disorder and externalizing behavior problems participating in an ongoing multicenter intervention study. Using confirmatory factor analyses (CFA) and exploratory structural equation modeling (ESEM), we evaluated several factor models of externalizing spectrum disorders (unidimensional; first-order correlated factors; higher-order factor; fully symmetrical bifactor; bifactor S-1 model). Furthermore, we assessed our prevailing factor models for measurement invariance across raters (clinicians, parents, teachers) and assessment modes (interview, questionnaires). While both CFA and ESEM approaches provided valuable insights into the multidimensionality, ESEM solutions were generally superior, since they showed a substantially better model fit and less biased factor loadings. Among the models tested, the bifactor S-1 CFA/ESEM models, with a general hyperactivity-impulsivity reference factor, displayed a statistically sound factor structure and allowed for straightforward interpretability. Furthermore, these models showed the same organization of factors and loading patterns, but not equivalent item thresholds across raters and assessment modes, highlighting cross-situational variability in child behavior. Our findings are consistent with the assumption of the trait impulsivity theory that a common trait, presented as hyperactivity-impulsivity symptoms, underlies all externalizing disorders.

Public significance statement: This study evaluated the factor structure of externalizing symptoms in children and found that a common trait, expressed as hyperactivity-impulsivity symptoms, may underlie all externalizing disorders. These findings are consistent with predictions derived from the etiological trait impulsivity theory and highlight the multifactorial structure of psychopathology.

Keywords: externalizing disorders, ADHD, bifactor models, ESEM, trait impulsivity theory

Introduction

Externalizing spectrum disorders encompass several mental disorders including attention-deficit/hyperactivity disorder (ADHD), oppositional defiant disorder (ODD), conduct disorder (CD) as well as substance use disorders and antisocial personality disorder (Beauchaine et al., 2017). Although the DSM-5 (American Psychiatric Association, 2013) categorizes these disorders as distinct phenomena, different lines of research provide support for common etiological mechanisms. Among individuals diagnosed with externalizing disorders, high rates of concurrent comorbidity (Angold et al., 1999; Costello et al., 2003) and heterotypic developmental continuity (Beauchaine & McNulty, 2013; Costello et al., 2003), as well as overlapping genetic and neural vulnerabilities (Andersson et al., 2020; Shen et al., 2020) indicate shared etiological pathways. The concept of an externalizing spectrum has therefore gained much attention and is further supported by the Hierarchical Taxonomy of Psychopathology perspective (HiTOP; Kotov et al., 2017). Moreover, the concept of an externalizing spectrum is consistent with the Research Domain Criteria (RDoC) initiative, highlighting a dimensional approach to mental disorders by focusing on transdiagnostic neurobiological vulnerabilities (Insel et al., 2010).

The etiology of externalizing spectrum disorders has been conceptualized by the trait impulsivity theory as described by Beauchaine and colleagues (Beauchaine et al., 2017; Beauchaine & McNulty, 2013). According to this theory, trait impulsivity is assumed to arise from a dysfunction in the mesolimbic reward system and is expressed behaviorally as early-onset symptoms of the hyperactive-impulsive and combined types of ADHD. In psychological terms, trait impulsivity manifests itself as a reduced delay discount, taking actions without further thought or planning ahead, and impairments in self-control (cf. Beauchaine et al., 2017). Depending on the presence of high-risk environmental factors, trait impulsivity increases the likelihood of developing ODD and CD in childhood as well as the likelihood, across development, of progressing to increasingly severe externalizing disorders such as substance use disorders and antisocial personality disorder (Beauchaine et al., 2017; Beauchaine & McNulty, 2013). It has to be acknowledged, however, that various definitions of impulsivity exist, which range from more specific operationalizations to personality-based accounts. While these definitions have merit in certain contexts, Beauchaine et al. (2017) prefer to define impulsivity using DSM-5 derived ADHD scales of hyperactivity-impulsivity (HI) symptoms over more circumscribed measures due to their hereditary nature (approximately .80), associations with well-replicated molecular and genetic neural substrates, and shared progression along the externalizing spectrum as an ontogenic process (Beauchaine et al., 2017;

Beauchaine & Hinshaw, 2020). Notably, symptoms of inattention (IN) are not included in Beauchaine et al.'s definition of trait impulsivity. This notion is consistent with genetic evidence that IN symptoms of ADHD are etiologically distinct from the HI and combined presentations of ADHD (for a meta-analysis, see Nikolas & Burt, 2010).

The latent factor structure of externalizing spectrum disorders

Traditionally, latent factors have been modeled using exploratory factor analysis (EFA) or confirmatory factor analysis (CFA). While EFA allows all items to freely load on different factors, the CFA approach restricts items to only load on their target factor, thereby constraining all cross-loadings to zero (Morin et al., 2016). One possible reason for the popularity of the CFA approach may be that it allows for the evaluation of an *a priori* defined factor structure using goodness-of-fit indices. The trait impulsivity theory offers several hypotheses regarding the latent factor structure of ADHD and ODD, as well as related externalizing disorders. First-order correlated factor models provide initial insights into the mutual correlations among dimensions or disorders. Due to the hierarchical nature of the externalizing spectrum, higher-order and bifactor models are of particular interest. In a higher-order factor model, the shared variance of the first-order factors (e.g. ADHD, ODD, CD) is accounted for by a second-order factor (e.g. externalizing factor). Hence, the general factor (g-factor) explains the common variance of the first-order specific factors (s-factors). However, the proportionality constraint, that is, the proportions of variance explained by the general and specific constructs being constrained to be the same, limits the value of higher-order models in providing insights into the relationship between general and specific facets of psychopathology (Gignac, 2016). By contrast, bifactor models are not limited by the proportionality constraint, since the unique contributions of the indicators of the g-factor and s-factors are directly teased apart (Gignac, 2016). In a fully symmetrical bifactor model, all items directly load on a g-factor, i.e., the g-factor accounts for variance in all item scores. In addition, several uncorrelated s-factors are modeled (Reise, 2012). These s-factors account for further unique variance in specific sets of items which is not attributable to the g-factor (see our application in Fig. 1). In other words, the interpretation of the s-factors differs between the higher-order and bifactor models, in that the s-factors in bifactor models are interpreted as residuals relative to the g-factor.

In recent years, many research efforts have focused on examining the bifactor CFA structure of externalizing disorders in children, adolescents, and adults across community and clinical samples (Arias et al., 2018 see their Table 1 for 27 such studies). So far, only a handful of studies have examined bifactor models of ADHD and ODD symptoms together in clinical

(Martel et al., 2010; Rodenacker et al., 2018) and community samples (Burns et al., 2014; Lee et al., 2016). These studies found good model fit for bifactor models with a g-factor and two s-factors (ADHD and ODD; Martel et al., 2010), three s-factors (ADHD IN, ADHD HI, and ODD; Burns et al., 2014; Rodenacker et al., 2018), or in conjunction with a separate factor for sluggish cognitive tempo (Lee et al., 2016). Rodenacker et al. (2018) preferred an incomplete bifactor model, excluding the s-factor ADHD HI. Overall, most studies considered a fully symmetrical bifactor model to offer a better conceptualization of externalizing psychopathology than competing factor models. For bifactor models of ADHD and/or ODD symptoms, several studies demonstrated measurement invariance across sample settings (e.g. Rodenacker et al., 2016), gender (Lee et al., 2016; Rodenacker et al., 2016, 2018), or informants (e.g. Burns et al., 2014).

However, recent theoretical and empirical works have expressed concerns with regard to evaluating and interpreting a fully symmetrical bifactor model. These concerns include the bifactor model's tendency for superior model fit, which may be a symptom of overfitting (Bonifay et al., 2017). Moreover, the meaning of the g-factor may not be comparable across studies, as it strongly depends on the items included in the analyses (Eid et al., 2017; Reise, 2012). Furthermore, fully symmetrical bifactor models often yield anomalous or inadmissible parameter estimates (e.g. negative variance estimates, negative factor loading estimates), which impede a meaningful interpretation of the respective factors (Eid et al., 2017; Heinrich et al., 2020). On top of this, omega factor reliability coefficients, which many studies failed to take into account, reveal that s-factors are often weakly defined or even vanish empirically (Arias et al., 2018; Rodriguez et al., 2016). For example, Arias et al. (2018) conducted further analyses on published ADHD/ODD bifactor studies and found that s-factors often did not acquire sufficient stability to represent empirically interpretable constructs. These results particularly concerned the HI or hyperactivity s-factors and were most apparent in community samples. Similar difficulties with the fully symmetrical bifactor model have also been reported for various psychological constructs (Eid et al., 2017; Heinrich et al., 2020; see also Rodriguez et al., 2016 for a critical evaluation of such studies). Overall, these issues cast doubt on fully symmetrical bifactor models and their clinical significance when examining the structure of psychological constructs.

One approach that seeks to resolve the concerns raised above is the bifactor S-1 model, which was derived from the perspective of stochastic measurement theory (see Eid et al., 2017, for the formal background). In contrast to a fully symmetrical bifactor model, in a bifactor S-1 model, one first-order factor is *a priori* defined as the general reference factor, rendering g-

factors comparable across studies. For the items of the general reference factor, no s-factor is modeled (see Fig. 1). The remaining symptoms are contrasted statistically against the general reference factor. Specific factors are thus interpreted as residual factors, i.e. they represent true score variance not being shared with the general reference factor (Burns et al., 2020a; Eid et al., 2017). The decision for a general reference factor should be theoretically derived or correspond to a reference facet of special interest (Eid, 2020). Burns et al. (2020a) were the first to apply a bifactor S-1 model to ADHD and ODD symptom ratings. Based on theoretical assumptions of the trait impulsivity theory, the authors modeled HI as the general reference factor, while the IN and ODD symptoms were represented by specific residual factors. Burns et al. (2020a) found that their bifactor S-1 model provided excellent model fit to parent and teacher ratings and resolved anomalous findings related to their fully symmetrical bifactor model. In a reanalysis of the data presented by Rodenacker et al. (2018), Junghänel et al. (2020) were able to replicate the findings of Burns et al. (2020a) in a clinical sample of children with externalizing behavior problems. That is, applying a bifactor S-1 model to parent ratings of ADHD and ODD symptoms resolved anomalous findings related to the associated fully symmetrical bifactor model and allowed for an unambiguous interpretation, which was consistent with predictions of the trait impulsivity theory.

Since CFA has been increasingly criticized for its approach of constraining item cross-loadings to zero, which may not constitute a realistic proposition, exploratory structural equation modeling (ESEM) has been developed (Marsh et al., 2014; Morin et al., 2016). ESEM combines the advantages of EFA (allowing cross-loadings) and CFA (*a priori* model specification; model fit indices), and can also be applied to bifactor models (assessing general and specific factors). Thus, ESEM offers a state-of-the-art analysis technique to disentangle the underlying sources of construct-relevant psychometric multidimensionality (Morin et al., 2016). With regard to its application to externalizing symptoms, ESEM has proven to be effective for validating ADHD/ODD instruments (Burns et al., 2013) and examining the latent factor structure in ADHD adults (Gomez et al., 2021). Furthermore, bifactor ESEM models offered a better solution than bifactor CFA models for ADHD teacher ratings of preschool children (Arias et al., 2016) and in an ADHD adult community sample (Gomez & Stavropoulos, 2020). By contrast, Rodenacker et al. (2017) compared CFA and ESEM models across informants and could not find consistent evidence for the ESEM solutions in clinically referred children.

Overall, five general findings emerge from the literature, namely that (1) fully symmetrical bifactor models have often mistakenly been chosen as the “winning” model,

despite serious statistical and interpretational difficulties; (2) bifactor S-1 models may offer a promising alternative to fully symmetrical bifactor models for disentangling the factor structure of externalizing spectrum disorders; (3) ESEM is an advanced method to provide valuable insights into the multidimensionality of psychological constructs; (4) more research using clinical interviews, which represent the gold standard for diagnosing mental disorders, is of particular interest and importance; and (5) more empirical research is needed that includes further symptoms in addition to ADHD and ODD in order to cover a broader range of the externalizing spectrum.

Aim of the study

The purpose of our study was to systematically test the underlying factor structure of externalizing symptoms using CFA and ESEM models. Following current evidence, we considered not only ODD symptoms as defined by the DSM-5, but also symptoms of affective dysregulation, as proposed in the subtype ODD with chronic irritability/anger (ODD-AD; Evans et al., 2017). In addition, we included CD and callous-unemotional (CU) symptoms in order to extend previous research and enhance the understanding of the externalizing spectrum. To the best of our knowledge, this is the first study to examine hypotheses regarding the latent factor structure of ADHD, ODD, CD, and CU symptoms derived from the trait impulsivity theory in a clinical sample of children. Beyond this, we evaluated the factor reliability using omega coefficients and tested the measurement invariance of general and specific facets of externalizing symptoms across raters or informants (clinicians, parents, teachers) and assessment modes (interview, questionnaires).

Hypotheses

We evaluated several factor models of externalizing spectrum disorders (unidimensional; first-order correlated factors; higher-order factor; fully symmetrical bifactor; bifactor S-1 model) obtained from CFA and ESEM measures (see Fig. 1). In line with the results of previous studies on the structure of ADHD and ODD symptoms (Arias et al., 2018), and due to the comorbidity among externalizing disorders (Angold et al., 1999; Costello et al., 2003), we expected that our first-order CFA and ESEM models would provide a better model fit than the unidimensional CFA model. Furthermore, we assumed that our ESEM models would provide a more realistic solution than CFA due to the fallible nature of indicators (Morin et al., 2016) as well as previously observed cross-loadings between externalizing symptoms (Arias et al., 2016; Burns et al., 2013; Gomez et al., 2021; Gomez & Stavropoulos, 2020; Rodenacker et

al., 2017). Moreover, we proposed that our higher-order and bifactor CFA/ESEM models would be more adequate than our first-order factor models because they incorporate the coexistence of global and specific constructs (Morin et al., 2016), which is in line with the trait impulsivity theory (Beauchaine et al., 2017). Additionally, we expected our bifactor S-1 CFA/ESEM models with HI as the general reference factor to resolve possible problematic results associated with the fully symmetrical bifactor models (Burns et al., 2020a; Heinrich et al., 2020). Our decision to select HI as the general reference factor was based on theoretical assumptions from the trait impulsivity theory (Beauchaine et al., 2017; Beauchaine & McNulty, 2013) and empirical studies on the application of a bifactor S-1 model to ADHD/ODD symptom ratings (Burns et al., 2020a; Junghänel et al., 2020). Finally, we hypothesized that our prevailing factor models would be invariant across different raters or informants, as this was previously found for the factor structure of ADHD and ODD symptoms (Burns et al., 2014, 2020a), and additionally across assessment modes (interview, questionnaires).

Methods

Participants and procedure

Data for the current analyses were derived from the ongoing ESCAschool multicenter study (ESCAschool: Evidence-based, Stepped Care of ADHD in school-aged children), which is part of the ESCAlife research consortium. A total of nine study centers located in Germany (Cologne, Essen, Göttingen, Hamm, Mainz, Mannheim, Marburg, Tübingen, Würzburg) participated ESCAschool. The ESCAschool study investigates an evidence-based, individualized, stepwise- intensifying treatment program based on behavioral and pharmacological interventions for children who meet criteria for an ADHD diagnosis according to the DSM-5. Further details on the background and procedures, including inclusion and exclusion criteria, are outlined in (Döpfner et al., 2017). For the present study, the following informants were considered: clinicians, parents, and teachers. We analyzed baseline data (i.e. before any intervention) of 474 school-age children (age range 6 - 12 years, M = 8.9, SD = 1.5, 81% males). The screening to check the participants' eligibility for the multicenter study was based on the *Clinical Parent Interview for Externalizing Disorders in Children and Adolescents* (ILF-EXTERNAL; Görtz-Dorten et al., 2022) and is described elsewhere (Thöne et al., 2020). The baseline data set also included children who did not meet criteria for an ADHD diagnosis (i.e. screening negatives of the ESCAschool study; n = 32, 72% males). These screening negatives were characterized by subclinical ADHD symptomatology (Thöne et al., 2020). Clinical diagnoses of ADHD and externalizing disorders were based on the interview ILF-

EXTERNAL. All clinicians assessed comorbid symptoms other than externalizing disorders using a clinical diagnostic checklist (DCL-SCREEN) from the German *Diagnostic System of Mental Disorders in Children and Adolescents based on the ICD-10 and DSM-5* (DISYPS-III; Döpfner & Görtz-Dorten, 2017). The participants' clinical and demographic characteristics are reported in Table 1. All parents, children, and teachers were informed verbally and received information sheets about the ESCAschool study. All parents and all children gave their assent and written informed consent and, likewise, all teachers provided their written informed consent for study participation. Ethical approval was obtained by each participating study site separately (Döpfner et al., 2017).

Measures

The children's parents or primary caregivers underwent the ILF-EXTERNAL, a semi-structured diagnostic interview based on the DSM-5, which is part of the DISYPS-III (Döpfner & Görtz-Dorten, 2017). Each item is rated by the clinicians on a 4-point Likert scale ranging from 0 (age-typical / not at all) to 3 (very much), with higher scores reflecting greater symptom severity. Further details on the administration and scoring are reported in Thöne et al. (2020). The ILF-EXTERNAL assesses IN symptoms (nine items) and HI symptoms (nine items) according to the DSM-5 to evaluate an ADHD diagnosis. Furthermore, the ILF-EXTERNAL assesses ODD symptoms (eight items), CD symptoms (15 items), disruptive mood dysregulation symptoms (five items in total; two items assessing affective dysregulation and three items assessing irritability/anger associated with ODD), and CU symptoms (11 items). The CD items evaluating aggressive and antisocial symptoms in participants aged 11 years or older were excluded from further analyses due to obvious floor effects. Psychometric evaluations support the interrater reliability [scale level: ICC(1,1) = .83 - .95; ICC(1,3) = .94 - .98] as well as the convergent and divergent validity of the ILF-EXTERNAL scale scores (Thöne et al., 2020). In the current sample, internal consistencies (Cronbach's alpha: $.60 \leq \alpha \leq .87$) and item-total correlations ($.22 \leq r_{it} \leq .68$) were generally satisfactory to good (see Table S1).

The German *Symptom Checklist for Attention-Deficit/Hyperactivity Disorder* (FBB-ADHS) and the *Symptom Checklist for Disruptive Behavior Disorders* (FBB-SSV) from the DISYPS-III (Döpfner & Görtz-Dorten, 2017) were completed by the parents and teachers. These symptom checklists provide symptom scales identical to those in the ILF-EXTERNAL. All items are rated on a 4-point Likert scale ranging from 0 (age-typical / not at all) to 3 (very much). The FBB-ADHS assesses IN symptoms (nine items) and HI symptoms (11 items). The

FBB-SSV assesses ODD symptoms (eight items), AD symptoms (six items, three of which also assess ODD symptoms), CD symptoms (16 items), and CU symptoms (11 items). Some of these symptoms are divided into an (a) and a (b) item, i.e. the respective symptom is assessed by two items. As the evaluation of measurement invariance requires the same set and number of items, we excluded two items of the FBB-ADHS (B03b *extreme internal restlessness*; B05a *permanently extremely restless*) and two items of the FBB-SSV SSV (B01a *starts fights with siblings;* D01a *severe temper outbursts)* from the analysis. Thereby, we retained that item which was more closely associated with the corresponding item from the ILF-EXTERNAL. Moreover, similarly to the ILF-EXTERNAL, the aggressive and antisocial items from the CD symptoms (items: B06 – B15) were also excluded. Psychometric evaluations support the reliability and validity of the FBB-ADHS scale scores (Döpfner et al., 2008; Erhart et al., 2008) and the FBB-SSV scale scores (Görtz-Dorten et al., 2014). Descriptive information for clinician, parent, and teacher ratings in the current sample is summarized in Table S1.

Statistical analyses

Model estimation: Graphical representations of the factor models considered in this study are depicted in Fig. 1. Regarding our CFA models, items only loaded on their respective factor, while cross-loadings were restricted to zero. Regarding our ESEM models, items freely loaded on their target factor, whereas cross-loadings were estimated to be as close to zero as possible with the oblique target rotation (first-order correlated factors; bifactor S-1 model) or the orthogonal target rotation (fully symmetrical bifactor model) procedure (Morin et al., 2016). The higher-order ESEM model was estimated using ESEM-within-CFA (Morin & Asparouhov, 2018).

All latent construct analyses were performed using Mplus version 8.4 for Mac (L. K. Muthén & Muthén, 2017). Given the ordinal structure of our data (4-point Likert scale), the *weighted least squares means and variance adjusted* estimator (WLSMV; Delta parameterization) was chosen. The WLSMV was specifically proposed for ordinal data, mainly because it does not make distributional assumptions about the observed items (C. H. Li, 2016). Missing data were handled using pairwise present analysis, as this is the default in Mplus for WLSMV (L. K. Muthén & Muthén, 2017). The amount of missing data per item varied from 0% (IN, HI), through 5% (ODD-AD, CD), to 6% (CU). Covariance coverage was greater than 94% for each item of the ILF-EXTERNAL.

Evaluating model fit: Global goodness of fit was assessed using the χ^2 test of exact model fit, the comparative fit index (CFI), the Tucker-Lewis index (TLI), the root mean square error of approximation (RMSEA) with corresponding 90% confidence intervals, and the standardized root mean square residual (SRMR). Model fit was considered as acceptable when RMSEA and SRMR were ≤ .08, and as good when CFI and TLI were ≥ .95 and RMSEA and SRMR were ≤ .05 (Hooper et al., 2008; Hu & Bentler, 1999). Considering that the χ^2 test is overly sensitive to sample size, more weight was given to CFI, TLI, RMSEA, and SRMR. Nevertheless, we wish to emphasize that these fit indices should only be interpreted as rough guidelines and that the statistical and theoretical conformity of the factor models should be considered as well (cf. Morin et al., 2016). In addition, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) were calculated using *maximum likelihood estimation with robust standard errors* (MLR) and Monte Carlo simulation (1,500 integration points) to evaluate whether one model provides a better fit than another. These information criteria compare competing models, and balance improvement in model fit against overparameterization, with smaller values being preferred (cf. Burnham & Anderson, 2004). So far, it is not possible to calculate AIC and BIC using MLR estimator with categorical indicators from ESEM models in Mplus. Therefore, we had to calculate AIC and BIC without the categorical command for all factor models to ensure that AIC and BIC were comparable across CFA and ESEM models.

The explained common variance was computed to evaluate the degree of unidimensionality of our bifactor models. For the g-factor, explained common variance is the ratio of variance explained by the g-factor divided by the variance explained by the g-factor plus the s-factors (Reise, 2012). In addition, we calculated omega factor reliability coefficients to evaluate whether the g-factor and s-factors truly represent the target constructs of interest (Rodriguez et al., 2016). Coefficient omega ω (McDonald, 1999; Revelle & Zinbarg, 2009) assesses the amount of variance in item scores attributable to the g-factor and the s-factors taken together. Moreover, omega hierarchical general ω_H (McDonald, 1999; Zinbarg et al., 2005) expresses the amount of variance attributable to the g-factor, and omega hierarchical subscale ω_S (Reise et al., 2013) displays the amount of variance attributable to an s-factor. As such, ω_S provides an estimate of how much reliable variance exists beyond that due to the g-factor and whether it is reasonable to interpret s-factors.

On the item level, consistency and specificity measures were obtained for our bifactor S-1 CFA model. While consistency refers to the amount of true score variance in each symptom

accounted for by the g-factor, specificity represents the amount of true score variance in each symptom attributable to a specific domain (Eid et al., 2017).

Measurement invariance

The prevailing models were tested for measurement invariance (Meredith, 1993; Vandenberg & Lance, 2000) across raters or informants (clinicians, parents, teachers) and assessment modes (interview, questionnaires). Configural invariance examines whether the basic model configuration (i.e. the same items belonging to the same factors) is the same in both groups. Metric invariance additionally requires equal item loadings across groups. Scalar invariance additionally requires item intercepts to be the same across groups. Difference tests for CFI, TLI, RMSEA, and SRMR were used to evaluate whether the imposed constraints were invariant across groups (WLSMV estimator; Theta parameterization). The same aforementioned cut-off values were considered as valid to assess configural invariance. Additionally, it has been suggested that a change in CFI of ≤ -.01 (Cheung & Rensvold, 2002), a change in SRMR of ≤ .30 for metric invariance or a change in SRMR of ≤ .01 for scalar invariance (F. F. Chen, 2007), or equal or better fit of TLI and RMSEA (Marsh et al., 2010), indicate that two models have equivalent fit.

Results

Factor analytic evaluations

Goodness-of-fit indices of the alternative factor models for clinician ratings are summarized in Table 2. Additional findings regarding parent and teacher ratings can be found in the supplement (Tables S9 – S15).

Regarding our unidimensional model, model fit indices did not meet the aforementioned cut-off criteria, resulting in a relatively poor model fit. These results indicate that models with correlated factors may provide more insights into the multidimensionality of the externalizing spectrum. Indeed, we found that our first-order ESEM/CFA models provided a better model fit than our unidimensional model in terms of model fit indices and information criteria. Furthermore, all standardized CFA factor loadings and ESEM target factor loadings were significant (Table S2).

With respect to our first-order CFA and ESEM solutions, we found lower factor correlations for ESEM (|*r*| = .07 to .36, *M* = .26) than CFA (|*r*| = .30 to .75, *M* = .47; Table S3). Examining ESEM factor loadings revealed well-defined factors with overall substantial target

factor loadings (|λ| = .21 to 1, *M* = .55; Table S2). Similarly, several significant cross-loadings were present, mostly involving the CU items, although few loadings were substantial (> .20).

Since our higher-order and bifactor CFA/ESEM models incorporate the coexistence of a g-factor and s-factors, we proposed that they would be more adequate for modeling externalizing symptoms. However, model fit indices of the higher-order CFA model were only marginally acceptable, indicating that this model might not capture the data in the most ideal way. Examining the factor correlations revealed that the second-order factor explained a lot of the shared correlations between the first-order factors (|*r*| = .54 to .86, *M* = .70; Table S4), although moderate to high residual variances from the first-order factors were present (σ^2 = .28 to .70, *M* = .50), indicating that a considerable amount of variance was not shared with the second-order factor. Regarding our higher-order ESEM model, model estimation resulted in a better representation of the data than CFA according to improvement in fit indices. These results support that relevant cross-loadings are present, supporting the need to rely on ESEM. Moreover, the superiority of the CFA/ESEM higher-order models further support the appropriateness of incorporating hierarchically superior constructs.

Our fully symmetrical bifactor CFA model demonstrated adequate model fit. However, this model produced some anomalous results yielding a small, negative residual variance (σ^2 = -.05) of one item in the CD factor (item B02: *bullies, threatens, or intimidates*). Although we carefully checked modification indices and tried fixing the negative variance to zero, the resulting model could not be properly estimated. Our fully symmetrical bifactor ESEM model demonstrated a well-defined g-factor with mostly substantial factor loadings (|λ| = .09 to .72, *M* = .42; Table S5). Over and above the g-factor, the s-factors were well-defined through mostly substantial target factor loadings (|λ| = .06 to .79, *M* = .44) and omega factor reliability coefficients (ω_S = .23 to .70) suggesting that they indeed tap into relevant specificity and add information to the g-factor. Notably, only a small number of substantial cross-loadings (> .20) could be observed, indicating that items have very little overlap between the different constructs. Despite the many advantages a bifactor ESEM model offers, one serious caveat still remains, namely the interpretation of the g-factor (Eid et al., 2017; further discussed below).

Given the drawbacks of fully symmetrical CFA/ESEM bifactor models, we constructed bifactor S-1 CFA/ESEM models with a general HI reference factor. Both bifactor S-1 models demonstrated adequate model fit and both models resulted in a well-defined general reference factor (CFA: (|λ| = .01 to .79, *M* = .36; ESEM: |λ| = .05 to .78, *M* = .40) and s-factors (CFA: |λ| = .25 to .79, *M* = .54; ESEM: |λ| = .16 to .86, *M* = .46; Table S6). Residual factor correlations can be interpreted meaningfully as partial correlations (CFA: (|*r*| = .10 to .67, *M* = .38; ESEM:

(|*r*| = .01 to .27, *M* = .16; Table S7), as the theoretical rationale behind the bifactor S-1 model does not assume that the g-factor accounts for all common variance (Eid et al., 2017). Omega factor reliability coefficients further supported the bifactor S-1 CFA/ESEM models for the g-factor (CFA: ω_H = .67; ESEM: ω_H = .73) and the s-factors (CFA: ω_S = .56 to .75; ESEM: ω_S = .24 to .73). These results demonstrate that a substantial amount of reliable variance exists beyond that attributable to the g-factor, and indicate that the s-factors can be interpreted reliably in both bifactor S-1 models. Consistency and specificity values for each symptom item from the bifactor S-1 CFA model show that while some true score variance was associated with the general HI reference factor, more true score variance for all symptoms was associated with their respective s-factor (Table S8). Overall, both bifactor S-1 models provided a statistically valid factor structure, which can be interpreted in a straightforward manner and is consistent with the trait impulsivity theory. We conclude that both the more parsimonious bifactor S-1 CFA model and the more realistic ESEM solution offered the best conceptualization of externalizing spectrum disorders.

Measurement invariance

Given that we considered our bifactor S-1 CFA/ESEM models as the optimally fitting models for our symptoms of externalizing disorders due to their statistical and theoretical conformity, we next tested these models for measurement invariance across raters or informants (clinicians, parents, teachers) and assessment modes (interview, questionnaires). For both bifactor S-1 CFA/ESEM models, we found that metric invariance, but not scalar invariance, was supported (Table 3). These results imply that the bifactor S-1 CFA/ESEM models present the same organization of factors and loading patterns as well as equal item loadings and cross-loadings across groups, but there is some variation in the individual item thresholds across groups. In other words, we note discrepancies across raters and informants with regard to individual symptom severity across settings.

Discussion

This study tested the underlying factor structure of externalizing symptoms in clinically referred children using CFA and ESEM models. When comparing our CFA and ESEM models, the ESEM solutions were generally superior, since they showed a substantially better model fit and more interpretable, less biased factor loadings and lower factor correlations. Generally, when this is the case, ESEM is considered a better model than CFA as it results in a clearer differentiation between the factors (Marsh et al., 2014; Morin et al., 2016). These findings fit

well with the increasing support for ESEM models, which may provide a more realistic representation of externalizing symptoms (Arias et al., 2016; Gomez et al., 2021; Gomez & Stavropoulos, 2020; Rodenacker et al., 2017). As outlined in the introduction and demonstrated in this study, CFA might often be overly restrictive in that items must only load on their target factors, but not on other, conceptually related factors (Marsh et al., 2014; Morin et al., 2016).

In terms of model fit indices, the fully symmetrical bifactor ESEM model demonstrated the best model fit and resolved anomalous parameter estimates related to its CFA counterpart. These observed statistical anomalies related to the bifactor CFA model are consistent with methodological concerns expressed by Eid (2017) and empirical work showing a variety of anomalous results associated with the application of fully symmetrical bifactor models to externalizing symptoms (Arias et al., 2018; Burns et al., 2020a; Rodenacker et al., 2018). We acknowledge that bifactor ESEM models provide an advanced, powerful analysis technique for modeling the fallible nature of indicators and construct-relevant psychometric multidimensionality due to hierarchically superior constructs (Morin et al., 2016). However, we would also like to call attention to one serious drawback of bifactor models, namely that the meaning of a g-factor varies across studies and with that the meaning of the s-factors, since they are defined in terms of residual variance related to the g-factor (Eid et al., 2017; Reise, 2012). For example, a g-factor of the externalizing spectrum changes its meaning depending on the number of domains included in the respective study, thus impeding the comparability across studies. This causes considerable problems relating to the synthesis of research through literature reviews, meta-analyses, or replication studies (Eid et al., 2017; Reise, 2012).

For these reasons, we also constructed bifactor S-1 CFA and ESEM models with a general HI reference factor (Burns et al., 2020a; Eid et al., 2017). Our decision to choose HI as reference factor was guided by previous research (Burns et al., 2020a; Junghänel et al., 2020) and by the trait impulsivity theory which states that *"hyperactive–impulsive symptoms of ADHD are the purest behavioral manifestations of trait impulsivity"* (Beauchaine et al., 2017, p. 347). Both bifactor S-1 models demonstrated appropriate model fit, but more importantly, displayed a statistically sound factor structure and unambiguous and meaningful interpretability. Although cross-loadings were present in the ESEM model, these did not undermine the definitions of the s-factors. Moreover, some of the cross-loadings seem reasonable given that they tap into similar aspects of the target constructs (e.g. items loading onto ODD-AD and CD s-factors). These novel insights are in line with other recent studies which applied the bifactor S-1 CFA model to ADHD/ODD symptoms in community (Burns et al., 2020a) and clinical samples of children and adolescents (Junghänel et al., 2020). Notably,

our study extends these previous findings by considering a wider range of externalizing symptoms and incorporating ESEM models. Compared to Burns et al. (2020a), we found lower loadings on the general reference factor and, accordingly, higher loadings on the s-factors. Furthermore, our consistencies of the items were lower and, accordingly, the specificities were higher. Moreover, all s-factors explained a higher amount of reliable variance than they did in the aforementioned community sample (Burns et al., 2020a). The observed differences may be attributable to a broader range of externalizing symptoms in our models or to different manifestations of externalizing symptomatology in clinical versus community samples. It is conceivable that s-factors explain more reliable variance in clinical samples due to greater response variability (see also Burns et al., 2014; Lee et al., 2016; Rodenacker et al., 2016, 2018). Our results were quite similar to those of Junghänel et al. (2020), thus supporting our argument that s-factors might account for more reliable variance in clinical settings.

With regard to CU symptoms, we observed low, and some non-significant, factor loadings on the general reference factor and high loadings on the s-factor in our bifactor S-1 CFA and ESEM models. Accordingly, item consistencies were low and item specificities were high in the CFA model. This finding is of clinical relevance because it suggests that CU symptoms may capture different aspects of the externalizing spectrum. For instance, it has been proposed that increased CU traits in children and adolescents with aggressive/antisocial problems may designate a subgroup that is especially severe and violent, with a poor prognosis regarding treatment outcome (Frick et al., 2014a, 2014b). In other words, children and adolescents with increased CU traits may exhibit different profiles (e.g. in terms of biological, cognitive, emotional, and social characteristics) compared to aggressive/antisocial youth without increased CU traits, suggesting that these two subgroups may benefit from different psychotherapeutic interventions (Frick et al., 2014a, 2014b).

In a final step, we evaluated our bifactor S-1 CFA/ESEM models' measurement invariance. We found that scalar invariance was not confirmed, indicating that individual symptom severity differs across settings. These findings are of particular interest since cross-informant discrepancies had been theorized to reflect some kind of invalidity or rater bias (for a review, see De Los Reyes et al., 2013). However, there is growing recognition that such informant discrepancies may rather reflect how children's behavior varies meaningfully across settings (e.g. at school vs. at home). Hence, it may be more reasonable to assume that different informants will provide diverging estimates of children's symptom severity. In fact, a large-scale meta-analysis (De Los Reyes et al., 2015) demonstrated low to moderate correspondence (externalizing symptoms: r = .30) across multiple informants (parents, teachers, self-reports).

Likewise, reporter-specific variance across parents and teachers was reported in children with ADHD (Vitoratou et al., 2019) and externalizing behavior problems (King et al., 2018). As synthesized by Martel et al. (2017), there seem to be various explanations for cross-informant discrepancies, including cross-situational variability in child behavior, differential demands across contexts, individual perspective of raters, different attributions of child behavior, rater's characteristics (e.g. psychopathology), visibility of behavior (e.g. overt vs. covert), and externally generated rating bias. Consequently, an emerging consensus in the field is that such cross-informant discrepancies should be *"embrace[d], not erase[d]"* (Dirks et al., 2012, p. 558).

With regard to future directions of modeling the factor structure of psychopathology, some authors have raised important issues about the application of bifactor S-1 models (J. D. Burke & Johnston, 2020; Willoughby, 2020). These issues involve (1) overfitting of the bifactor S-1 model similar to the fully symmetrical bifactor model, (2) arbitrary selection of different reference facets, (3) questions about the interpretation of the g-factor, (4) additional value of the bifactor S-1 model, (5) application of the bifactor S-1 model to item-level data, and (6) application of bifactor models to find a general psychopathology factor (J. D. Burke & Johnston, 2020; Burns et al., 2020b; Willoughby, 2020). Burns et al. (2020b) thoroughly addressed these issues in a well-thought-out response clarifying these six issues as follows: (1) In the event that overfitting may be a concern, a restricted bifactor S-1 model can be estimated. This restricted model results in the same number of parameters and exact model fit as the correlated first-order factors model, thereby avoiding the danger that the bifactor S-1 model may be undeservedly chosen as the "better" model. (2) It is important to carefully consider *a priori* which facet should serve as reference, because changing the reference facet will change the interpretation of results. Burns et al. (2020b) explain that this reasoning can be related to changing the reference group in regression analysis, which affects the interpretation of regression coefficients. (3) Notably, the g-factor in a fully symmetrical bifactor model changes its meaning from a general to a specific factor when one s-factor collapses or when indicators have zero factor loadings on their s-factors (Burns et al., 2020a). As collapsing HI factors have often been reported (cf. Arias et al., 2018), the resulting g-factor turns into a facet-specific HI factor. (4) Applying a bifactor S-1 model can answer different research questions than the fully symmetrical bifactor model, providing additional value. For example, the factor structure within ODD symptoms could be disentangled by modeling HI symptoms as the general reference domain, with specific ODD irritability and defiant residual factors. (5) Problems related to the fully symmetrical bifactor model are not limited to item-level (symptom) data but

occur whenever a fully symmetrical bifactor model is used with structurally different facets or raters (Eid et al., 2017). (6) Although the desire to find a general factor of psychopathology may be understandable, Burns et al. (2020b) discourage the use of fully symmetrical bifactor models for this research purpose for the reasons outlined above. Instead, the authors highlight that a bifactor S-1 model yields a g-factor with well-defined meaning across studies, and may enhance our view on psychopathology.

In terms of limitations, our clinical sample mainly includes children with an ADHD diagnosis (94%). Hence, this sample does not capture the full spectrum of externalizing disorders, but rather represents a sample of different ADHD subtypes with varying severity and associated externalizing behavior problems. Therefore, our results require replication in more representative samples of externalizing spectrum disorders. In this regard, it would also be informative to include more severe externalizing disorders (e.g. antisocial behavior, substance use) into the externalizing spectrum (Krueger et al., 2007). Ideally, future studies should also include longitudinal data in order to track the developmental trajectory of externalizing spectrum disorders. In addition, analyzing the relationship between different facets and external criterion variables would provide further value and demonstrate the prevailing model's practical relevance (e.g. Burns et al., 2020a).

Moreover, it should be noted that latent factor models based solely on behavioral symptoms cannot provide etiological insights given the complexity of neurobiological vulnerabilities and neural circuitries. Nevertheless, we believe that consistency with predictions derived from etiological theories is an important step toward a dimensional model of mental disorders. We therefore encourage future research to evaluate multiple units of analysis within the RDoC framework (Beauchaine & Hinshaw, 2020; Insel et al., 2010). Such research would not only shed new light on the hierarchical factor structure of mental disorders but would also be highly relevant for advancing more target-oriented psychotherapeutic interventions.

Finally, our findings may also have implications for the organization of externalizing disorders within diagnostic classification systems. In the DSM-5, ADHD is categorized in the neurodevelopmental disorders section, whereas ODD and CD are found in the disruptive, impulse-control, and conduct disorders category. Furthermore, disruptive mood dysregulation disorder is categorized as a standalone diagnosis in the section on depressive disorders, although this is subject to contentious debate among clinicians and researchers (Evans et al., 2017). Placing these mental disorders into different categories seems to imply different etiological processes. However, as indicated by etiological theories (Beauchaine et al., 2017; Beauchaine & McNulty, 2013), genetic research (Andersson et al., 2020; Shen et al., 2020),

and factor analytic accounts within our study and previous work (Burns et al., 2014, 2020a; Junghänel et al., 2020; Lee et al., 2016; Rodenacker et al., 2018), separating disorders across DSM-5 categories may hinder transdiagnostic research, thereby slowing down potential diagnostic and therapeutic advances.

Conclusion

The current study was informed by the trait impulsivity theory and aimed to elucidate the latent factor structure of externalizing disorders using CFA and ESEM models. While both CFA and ESEM approaches provided valuable insights into the multidimensionality of the externalizing spectrum, ESEM solutions were generally superior, since they showed a substantially better model fit and more interpretable, less biased factor loadings and lower factor correlations. Among the models tested, the bifactor S-1 CFA and ESEM models, with a general HI factor and four specific residual factors (symptoms of IN, ODD-AD, CD, CU) provide a psychometrically sound and clearly interpretable conceptualization of symptoms of externalizing spectrum disorders in clinically referred children. These models demonstrate the same organization of factors and loading patterns, but not equivalent item thresholds across raters and informants, highlighting cross-situational variability in child behavior. Furthermore, our findings are consistent with predictions derived from the etiological trait impulsivity theory, which assumes an externalizing liability, expressed as temperamental trait impulsivity, to represent the core vulnerability for externalizing disorders. Our study, which links factor analytic accounts of symptoms to etiological models, enriches the understanding of the dimensional and multifactorial structure of psychopathology in children.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

Ann-Kathrin Thöne analyzed the data and developed the first draft of the manuscript. She was involved in the recruitment and data acquisition of the ESCAschool study site in Cologne and is co-author of the DISYPS-ILF. Manfred Döpfner is the principal investigator of the ESCAschool trial, developed the basic ESCAschool study design, is part of the ESCAlife consortium, is head of the Cologne recruiting center for the ESCA children trials (ESCApreschool, ESCAschool, and ESCAadol), developed the DISYPS-III system, developed the DISYPS-ILF, contributed to the interpretation of the results, and critically revised the manuscript. Michaela Junghänel contributed to the interpretation of the results and critically revised the manuscript. Anja Görtz-Dorten developed the DISYPS-III system, developed the DISYPS-ILF, contributed to the interpretation of the results, and critically revised the manuscript. Christina Dose contributed to the management and organization of the study, heads the telephone-assisted self- help trial performed in one treatment arm of the ESCAschool study,

and critically revised the manuscript. Christopher Hautmann was involved in the development of the ESCAschool research proposal and organization of the study and critically revised the manuscript. Lea Teresa Jendreizik and Anne- Katrin Treier coordinate ESCAschool, contribute to the management and organization of the study, and critically revised the manuscript. Paula Vetter is a scientific staff member at the Cologne study site, was involved in patient recruitment and data acquisition, and critically revised the manuscript. Elena von Wirth coordinates the ESCAschool study, contributed to the implementation of the study, provided supervision for patient treatment, and critically revised the manuscript. Tobias Banaschewski coordinates the ESCAlife consortium and is co-principal investigator in the ESCAschool study. He made substantial contributions to the conception and design of the ESCA- school research proposal, heads the Mannheim study site of ESCAschool, and critically revised the manuscript. Katja Becker heads the Marburg study site of ESCAschool, is PI for the ESCApreschool study, is a member of the ESCAlife consortium, is a co-applicant of the ESCAlife research project, made substantial contributions to the conception and design of the ESCAlife study, and critically revised the manuscript. Daniel Brandeis made substantial contributions to the conception and design of the ESCAschool research proposal and the neuropsychological research battery, heads the sub-project ESCAbrain, and critically revised the manuscript. Ute Dürrwächter coordinates the Tübingen study site ofESCAschool, contributes to the management and organization of the study, was involved in patient recruitment and data acquisition, and critically revised the manuscript. Julia Geissler coordinates the Würzburg study site, contributed to the management and organization of the ESCAlife projects, and critically revised the manuscript. Johannes Hebebrand heads the Essen study site of ESCAschool, made substantial contributions to the conception and design of the ESCAschool study, and critically revised the manuscript. Sarah Hohmann coordinates the ESCAlife consortium, made substantial contributions to the conception of the neuro- psychological research battery and the neurofeedback protocol, and critically revised the manuscript. Martin Holtmann heads the Hamm/Bochum study site of ESCAschool, was involved in the development of the ESCAschool research proposal, and critically revised the manuscript. Michael Huss heads the Mainz study site of ESCAschool, was involved in the implementation of the ESCAlife projects, and critically revised the manuscript. Thomas Jans was involved in the planning of the ESCAlife research projects and application for funding, is co-PI for the ESCAadol trial, was involved in the implementation of the ESCAlife projects at the Würzburg study site, and critically revised the manuscript. Johanna Ketter coordinates ESCAschool at the Marburg study site, contributes to the management and organization of the study, was involved in patient recruitment and data acquisition, and critically revised the manuscript. Tanja Legenbauer heads the research department of the Hamm/Bochum study site of ESCAschool, contributed to the implementation of the study in Hamm, supervised the realization of the trial in Hamm, and critically revised the manuscript. Sabina Millenet heads the Mannheim study site of ESCAschool, coordinates the ESCAlife consortium, made substantial contributions to the conception of the neuropsychological research battery and the neurofeedback protocol, and critically revised the manuscript. Luise Poustka heads the Göttingen study site of ESCAschool, was involved in the implementation of the ESCAlife projects, and critically revised the manuscript. Tobias Renner heads the Tübingen study site of ESCAschool, contributed to the implementation of the ESCA- school study, and critically revised the manuscript. Marcel Romanos was involved in the planning of the ESCAlife research projects and application for funding, is co-PI for the ESCAadol trial, was involved in the implementation of the ESCAlife projects at the Würzburg study site, and critically revised the manuscript. Henrik Uebel-von Sandersleben coordinates ESCA- school at the Göttingen study site, contributed to the management and organization of the study, was involved in patient recruitment and data acquisition, and critically revised the manuscript. Jasmin Wenning coordinates ESCAschool at the Essen study site and contributed to the management and organization of the study and

critically revised the manuscript. Mirjam Ziegler is a scientific staff member at the Mannheim study site, contributed to the implementation of the study, was involved in patient recruitment and data acquisition, and critically revised the manuscript.

Conflicts of interest

Ann-Kathrin Thöne, Anja Görtz-Dorten, and Manfred Döpfner were involved in the development of the DISYPS-ILF and will receive royalties from the publisher Hogrefe after the publication of this instrument. Anja Görtz-Dorten and Manfred Döpfner are AKiP supervisors and lecturers and received income as heads of the School for Child and Adolescent Behavior Therapy at the University of Cologne and royalties from treatment manuals, books, and psychological tests published by Guilford, Hogrefe, Enke, Beltz, and Huber. Manfred Döpfner received consulting income and research support from Lilly, Medice, Shire, Janssen Cilag, Novartis, and Vifor. Anja Görtz-Dorten is head of the AKiP Research and Evaluation department. Tobias Banaschewski served in an advisory or consultancy role for Lund- beck, Medice, Neurim Pharmaceuticals, Oberberg GmbH, Takeda/Shire, and Infectopharm. He received conference support or speaker's fees from Lilly, Medice, and Takeda/Shire. He received royalties from Hogrefe, Kohlham- mer, CIP Medien, and Oxford University Press. Katja Becker has been involved in research/clinical trials with Eli Lilly (≤2011) and Shire (≤2010), was on the Advisory Board of Eli Lilly/Germany (≤2014), a member of the Scientific Committee of Shire (≤2012), and was paid for public speaking by Eli Lilly (<2011) and Shire (2015). These activities do not bias the objectivity of this manuscript (in her opinion), but are mentioned for the sake of completeness. Daniel Brandeis served as an unpaid scientific advisor for an EU-funded neurofeedback trial unrelated to the present work. Martin Holtmann served in an advisory role for Shire and Medice and received conference attendance support or was paid for public speaking by Medice, Shire, and Neuroconn. He receives research support from the German Research Foundation and the German Ministry of Education and Research. He receives royalties as Editor-in-Chief of the German Journal for Child and Adolescent Psychiatry and for textbooks from Hogrefe. Michael Huss has served as a member of the advisory boards of Eli Lilly and Co, Engelhardt Arzneimittel, Janssen-Cilag, Medice, Novartis, Shire, and Steiner Arzneimittel within the past 5 years; served as a consultant to Engelhardt Arzneimittel, Medice, and Steiner Arzneimittel; received honoraria from Eli Lilly and Co, Engelhardt Arzneimittel, Janssen-Cilag, Medice, Novartis, and Shire; and received unrestricted grants for investigator-initiated trials from Eli Lilly and Co, Medice, Engelhardt Arzneimittel, and Steiner Arzneimittel. Luise Poustka served in an advisory or consultancy role for Shire, Roche, and Infectopharm. She has received speaker's fees from Shire and royalties from Hogrefe, Kohlhammer, and Schattauer. Henrik Uebel-von Sandersleben served in an advisory or consultancy role for Medice. He has received speaker's fees from Shire and Medice. Michaela Junghänel, Chris- tina Dose, Christopher Hautmann, Lea Teresa Jendreizik, Anne-Katrin Treier, Paula Vetter, Elena von Wirth, Ute Dürrwächter, Julia Geissler, Johannes Hebebrand, Sarah Hohmann, Thomas Jans, Johanna Ketter, Tanja Legenbauer, Sabina Millenet, Tobias Renner, Marcel Romanos, Jasmin Wenning, and Mirjam Ziegler are not aware of any biases that might be perceived as affecting the objectivity of this manuscript and declare that they have no competing interests.

Ethics approval and consent to participate

Ethical approval was obtained for the study center Cologne by the University of Cologne (ID 15–216), for the study center Essen by the University of Duisburg-Essen (ID 17–7404-BO), for the study center Hamm by the Ruhr-University Bochum (ID 15–5564), for the study center Göttingen by the University Medical Center Göttingen (ID 3/3/17), for the study center Mainz by the Federal Medical Association of Rhineland- Palatinate (ID 837.237.17 [11071]), for the study center Mannheim by the Ruprecht-Karls-University Heidelberg (ID 2015-646 N-MA),

for the study center Marburg by the Philipps-University Marburg (ID Studie 03/16), for the study center Tübingen by the Eberhard-Karls-University Tübingen (ID 791/2015BO2), and for the study center Würzburg by the Julius- Maximilians-University Würzburg (ID 332/15_z).

Patients, parents or guardians, and (if applicable) teachers were informed verbally and received information sheets about the scope and the relevance of the study as well as the study procedures. All participants were able to clarify questions about the study with research staff members. For study participation, it was required that both the patient and his/her parents or guardians gave their assent and informed consent. Likewise, teachers had to give their informed consent for study participation.

Acknowledgements

We thank all staff members who participated in the ESCAschool study in nine different study recruiting centers and who were involved in recruitment, study organization and management, blind rating, diagnostics and therapy, as well as the ESCAbrain study group and the CTU. These are (sorted by recruiting center affiliation and in alphabetical order) as follows:

Study center Cologne: Ruth Beckmann, Caroline Betcher, Jana Bretthauer, Corinna Broekmans, Yvonne Buntrock, Mareike Bürger, Dominique Deli, Elvan Dogan, Lisa Firmenich, Katrin Floß, Anna Fronhofen, Nina Geldermann, Claudia Ginsberg, Jana Harenkamp, Saskia Heintzmann, Julia Kellner, Claudia Kinnen, Marie-Theres Klemp, Hedda Körner, Constanze Kitzig, Katrin Krugmann, Jessica Leite, Sarah Meier, Ronja Meingast, Judith Mühlenmeister, Sophie Niestroj, Lea Ostrowski, Imke Partzsch, Jeannine Pawliczek, Daniela Perri, Judith Ratayczak, Stephanie Schürmann, Ellen Settegast, Johanna Spicher, Susanne Stollewerk, Katharina Verhüls- donk, Laura Wähnke, Christine Wälde, Annabelle Warschburger, Friederike Waschau, Tanja Wolff Metternich-Kaizman, and Sara Zaplana Labarga. Study center Essen: We would like to gratefully thank all the participating patients and their families. Study center Göttingen: Fatmeh Al-Debi and Jan Schulz. Study center Hamm: Miriam Davids, Natalie Deux, Mareen Dökel, Silvia Eißing, Regina Herdering, Carina Huhn, Lara Kaffke, Inken Kirschbaum-Lesch, Franziska Martin, Nina Müller, Nicola Nolting, Daniela Pingel, Carina Schulz, Karen Schumann, Birthe Wagner, and Daniela Walkenhorst. Study center Mainz: We would also like to gratefully thank all the participating patients and their families. Study center Mannheim: Karina Abenova, Brigitta Gehrig, Monja Groh, Raphael Gutzweiler, Christine Igel, Anna Kaiser, Theresa Nickel, Angelina Samaras, Anne Schreiner, Anne Schröter, Marie-Therese Steiner, Nina Trautmann, Lisa Wagner, and Matthias Winkler. Study center Marburg: Anette Becker, Viktoria Birkenstock, Doreen Blume, Anita Dehnert, Silas Deutsch, Sabine Finkenstein, Charlotte Finger, Claudia Freitag, Nicole Grau, Wiebke Haberhausen, Julia Häusser, Daria Kasperzack, Franziska Körfgen, Johannes Kresse, Jana Langhammer, Chris- topher Mann, Tanja Mingebach, Thomas Mooz, Jens Pfeiffer, Maren Rumpf, Alisa Samel, Bastian Schrott, Anna Schuh, Isabell Schulz, Thomas Stehr, Rebecca Stein, Anne von Stiphout, Linda Weber, Svenja- Katharina Weber, Anne-Kathrin Wermter, and Maximiliane Werther. Study center Tübingen: Yael Amling, Anna Hertle, Natalie Herrmann, Melinda Mross, Anja Pascher, Alena Roth, Priska Schneider, Anja Schöllhorn, and Ida Steinacker. Study center Würzburg: Nadja Becker, Christoph Biohlawek, Lea Brosig, Kathrin Delius, Adrian Dernbach, Anja Diehl, Simone Dorsch, Lisa Eidenschink, Stefanie Fekete, Zuzana Fouskova, Julia Gläser, Jana Halmagyi, Verena Hartlieb, Silke Hauck, Julia Häusler, Christin Heim, Sonja Hetterich, Michaela Linke, Annette Nowak, Corinna Otte, Isabel Paul, Katharina Peters, Hanna Schwarz, Carolin Steinmetz, Kristin Wehrmann, Johannes Weigl, and Lea Zwilling. Additional ESCAbrain staff members, who were involved in ESCA-school: Clinical Trials Unit Freiburg: Veronika Frommberger, Alexander Hellmer, Lydia Herbstritt, Maria Huber, Carolin Jenkner, Stefanie Lade, Patrick Müller, Barbara Schilling, Carla Schneider, and Ralf Tostmann. Further thanks to Sarah Mannion for English-language proofreading.

Table 1

Sample characteristics

	Total Sample (*N* = 474)
Age Mean (*SD*)	8.9 (1.5)
Male *n* (%)	382 (81)
Primary Diagnosis *n* (%)	
No ADHD diagnosis	32 (7)
ADHD – combined type	208 (44)
ADHD – predominantly inattentive type	184 (39)
ADHD – predominantly hyperactive-impulsive type	50 (11)
Comorbidities *n* (%)	*n* = 454 - 465
Internalizing disorders:	
- Anxiety	29 (6)
- Depression	15 (3)
Externalizing disorders:	
- Oppositional defiant disorder	166 (37)
- Conduct disorder	28 (6)
- Disruptive mood dysregulation disorder	40 (9)
Other disorders:	
- Obsessive-compulsive disorder	2 (< 1)
- Tic disorder	24 (5)
- Autism spectrum disorder	2 (< 1)
Medication *n* (%)	*n* = 462
ADHD medication	150 (33)
Parents' Primary Language *n* (%)	*n* = 458
German	429 (94)
Parents' Highest Educational Attainment *n* (%)	*n* = 455
Higher-track school	261 (57)
Vocational school	26 (6)
Medium-track school	123 (27)
Lower-track school	43 (10)

Note. Clinical diagnoses of ADHD and externalizing behavior disorders were based on the semi-structured *Clinical Parent Interview for Externalizing Disorders in Children and Adolescents* (ILF-EXTERNAL) conducted with the parents. Additional comorbid symptoms were evaluated using a clinical diagnostic checklist. ADHD = attention-deficit/hyperactivity disorder.

Table 2

Goodness-of-fit statistics and information criteria for alternative factor models of the externalizing spectrum

Model	χ^2 (*df*)	CFI	TLI	RMSEA (90% CI)	SRMR	AIC [a]	BIC [a]
Unidimensional	3535.343* (902)	.721	.707	.078 (.076, .081)	.109	47773.523	48322.803
First-Order CFA	1646.250* (892)	.920	.915	.042 (.039, .045)	.072	46160.296	46751.187
First-Order ESEM	1230.448* (736)	.948	.933	.038 (.034, .041)	.050	45985.638	47225.678
Higher-Order CFA	1752.630* (897)	.909	.904	.045 (.042, .048)	.076	46179.307	46749.393
Higher-Order ESEM	1228.856* (741)	.948	.934	.037 (.034, .041)	.050	45989.849	47209.082
Bifactor CFA	1548.872* (858)	.927	.919	.041 (.038, .044)	.069	45970.805	46703.177
Bifactor ESEM	1103.364* (697)	.957	.952	.035 (.031, .039)	.045	45897.184	47299.511
Bifactor S-1 CFA	1595.585* (861)	.922	.914	.042 (.039, .046)	.068	46143.283	46863.172
Bifactor S-1 ESEM	1230.481* (736)	.948	.933	.038 (.034, .041)	.050	45985.638	47225.678

Note. df = Degrees of freedom; CFI = Comparative fit index; TLI = Tucker-Lewis index; RMSEA = Root mean square error of approximation; CI = Confidence interval; SRMR = Standardized root mean square residual; AIC = Akaike information criterion; BIC = Bayesian information criterion; CFA = Confirmatory factor analysis; ESEM = Exploratory structural equation modeling.

[a] Information criteria were calculated using *maximum likelihood estimation with robust standard errors* (MLR) for continuous indicators to ensure that AIC and BIC were comparable across CFA and ESEM models. *$p < .001$.

Table 3

Measurement invariance results for the bifactor S-1 CFA model and bifactor S-1 ESEM solution

Model	Clinicians vs. Parents					Clinicians vs. Teachers					Parents vs. Teachers					All Informants Combined				
	χ^2 (*df*)	CFI	TLI	RMSEA	SRMR	χ^2 (*df*)	CFI	TLI	RMSEA	SRMR	χ^2 (*df*)	CFI	TLI	RMSEA	SRMR	χ^2 (*df*)	CFI	TLI	RMSEA	SRMR
Bifactor S-1 CFA																				
Configural	3384.494* (1722)	.930	.923	.046 (.044, .049)	.066	3384.069* (1560)	.925	.917	.056 (.054, .059)	.071	3599.219* (1560)	.927	.919	.061 (.059, .064)	.070	5114.124* (2340)	.927	.919	.055 (.053, .057)	.068
Metric	3186.080* (1796)	.941	.938	.041 (.039, .044)	.071	3267.329* (1630)	.932	.929	.052 (.049, .055)	.078	3228.595* (1630)	.943	.939	.053 (.050, .056)	.075	4709.459* (2489)	.941	.939	.048 (.046, .050)	.075
Scalar	3784.040* (1923)	.921	.923	.046 (.044, .048)	.072	3661.379* (1751)	.921	.922	.054 (.052, .057)	.079	3500.526* (1751)	.937	.938	.054 (.051, .056)	.076	5559.225* (2722)	.925	.929	.052 (.050, .054)	.077
Bifactor S-1 ESEM																				
Configural	2576.827* (1472)	.953	.940	.041 (.038, .043)	.047	2364.365* (1322)	.957	.944	.046 (.043, .049)	.046	2551.849* (1322)	.956	.942	.052 (.049, .055)	.043	3667.875* (1983)	.955	.942	.047 (.044, .049)	.045
Metric	2518.025* (1667)	.964	.959	.034 (.031, .036)	.055	2479.722* (1507)	.960	.954	.042 (.039, .045)	.059	2407.475* (1507)	.968	.963	.041 (.038, .045)	.055	3679.525* (2353)	.965	.961	.038 (.036, .040)	.059
Scalar	2925.832* (1794)	.952	.949	.037 (.035, .040)	.060	2923.710* (1628)	.946	.943	.046 (.044, .049)	.061	2662.748* (1628)	.963	.961	.043 (.040, .046)	.057	4532.936* (2595)	.949	.949	.044 (.042, .046)	.062

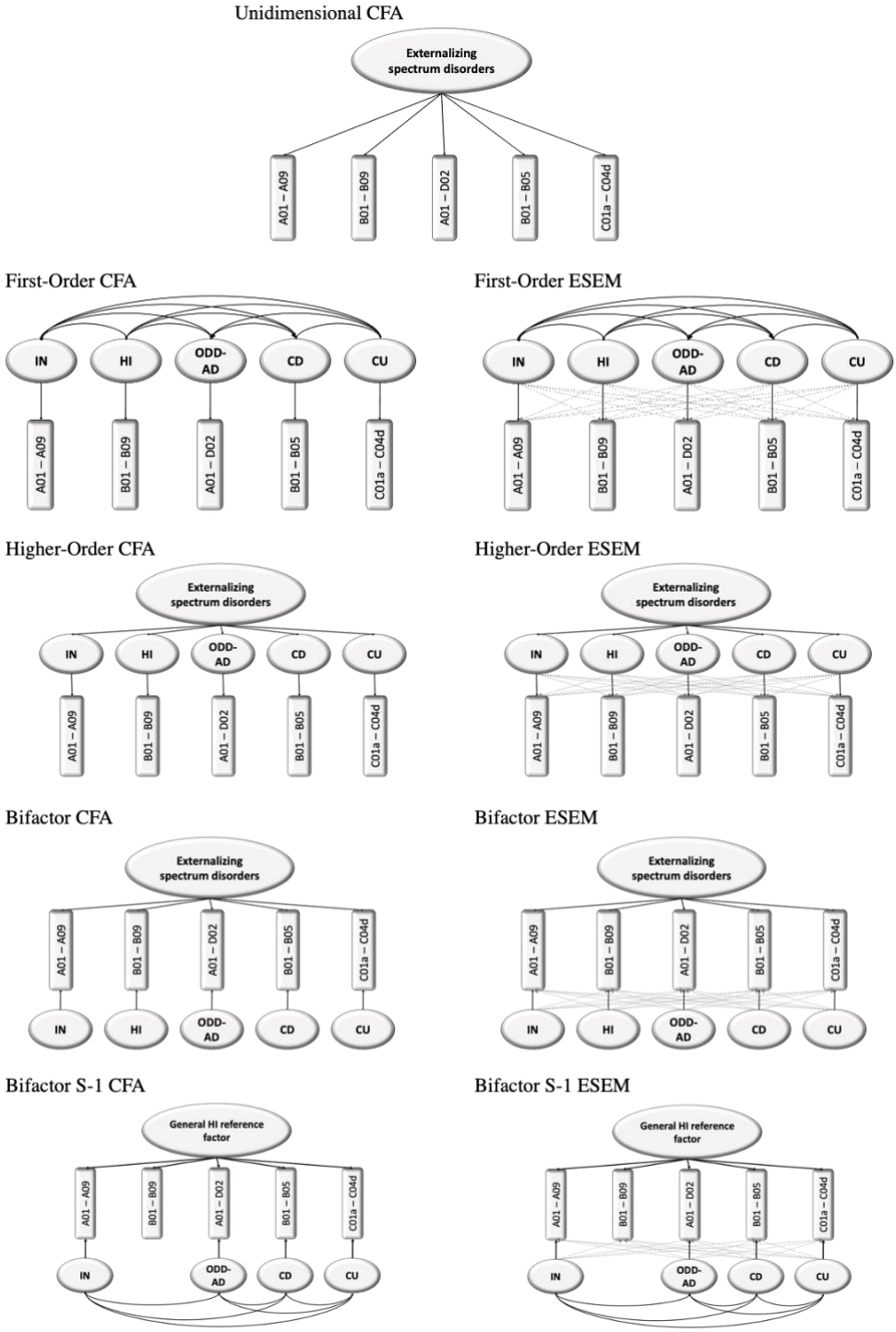
Note. Clinician ratings were based on the *Clinical Parent Interview for Externalizing Disorders in Children and Adolescents* (ILF-EXTERNAL). Parents and teachers completed symptom checklists for the assessment of ADHD symptoms and symptoms of disruptive behavior disorders (FBB-ADHS; FBB-SSV). For comparison with teacher ratings, two items (B03: *Cruel to animals*, B05: *Steals without confrontation*) were omitted due to zero variance. CFA = Confirmatory factor analysis; ESEM = Exploratory structural equation modeling; *df* = Degrees of freedom; CFI = Comparative fit index; TLI = Tucker-Lewis index; RMSEA = Root mean square error of approximation; CI = Confidence interval; SRMR = Standardized root mean square residual.

n = 474 (Clinicians); n = 428 (Parents); n = 267 (Teachers).

*p < .001.

Figure 1

Factor models of externalizing spectrum disorders



Note. Graphical representation of the factor models considered in this study. Full unidirectional arrows represent the target factor loadings and dashed unidirectional arrows represent the cross-loadings. Item numbers are displayed in the boxes and residuals are not shown for clarity of presentation. CFA = confirmatory factor analysis; ESEM = exploratory structural equation modeling; IN = inattention (9 items: A01–A09); HI = hyperactivity–impulsivity (9 items: B01–B09); ODD-AD = oppositionality with chronic irritability/anger (10 items: A01–A08; D01–D02); CD = conduct disorder (5 items: B01–B05); CU = callous–unemotional (11 items: C01a–C04d). A short description of each item is provided in Table S2.

5 Identifying symptoms of ADHD and disruptive behavior disorders most strongly associated with functional impairment in children: A symptom-level approach

Thöne, A.-K., Dose, C., Junghänel, Hautmann, C., Jendreizik, L. T., Treier, A.-K., Vetter, P., von Wirth, E., Banaschewski, T., Becker, K., Brandeis, D., Dürrwächter, U., Geissler, J., Hebebrand, J., Hohmann, S., Holtmann, M., Huss, M., Jans, T., Kaiser, A., Ketter, J., Legenbauer, T., Millenet, S., Poustka, L., Renner, T., Romanos, M., Uebel-von Sandersleben, H., Schneider, P., Wenning, J., Ziegler, M., Görtz-Dorten, A., & Döpfner, M. (2023). Identifying symptoms of ADHD and disruptive behavior disorder most strongly associated with functional impairment in children: A symptom-level approach. *J. Psychopathol. Behav. Assess.*, *45*(2), 277–293. https://doi.org/10.1007/s10862-023-10025-z

Abstract

Objective: To enhance the understanding of how symptoms of attention-deficit/hyperactivity disorder (ADHD) and disruptive behavior disorders such as oppositional defiant disorder (ODD), conduct disorder (CD), including callous-unemotional (CU) traits, differentially relate to functional impairment (FI). *Methods:* Participants were 474 German school-age children (age: M = 8.90, SD = 1.49, 81% male) registered for participation in the ESCAschool trial (ESCAschool: *Evidence-based, Stepped Care of ADHD in school-aged children*). Clinicians assessed the severity of individual symptoms and five FI domains specifically associated with ADHD symptoms or ODD/CD/CU symptoms using a semi-structured clinical interview. We conducted two multiple linear regression analyses, combined with relative importance analyses, to determine the impact of individual symptoms on global FI associated with ADHD and ODD/CD/CU symptoms. Next, we estimated two networks and identified the strongest associations of ADHD symptoms or ODD/CD/CU symptoms with the five FI domains. *Results:* Symptoms varied substantially in their associations with global FI. The ADHD symptom *Easily Distracted* (15%) and ODD symptom *Argues with Adults* (10%) contributed most strongly to the total explained variance. FI related to academic performance, home life and family members, and psychological strain were most strongly associated with ADHD inattention symptoms, whereas FI related to relationships with adults and relationships with children and recreational activities were most strongly associated with hyperactivity-impulsivity symptoms. By comparison, the ODD/CD/CU symptoms most closely linked to FI domains originated from the ODD and CD dimensions. *Conclusions:* Our findings contribute to a growing body of literature on the importance of analyzing individual symptoms and highlight that symptom-based approaches can be clinically useful.

Keywords: functional impairment; attention-deficit/hyperactivity disorder; oppositional defiant disorder; conduct disorder; callous-unemotional traits; network analysis

Introduction

Attention-deficit/hyperactivity disorder (ADHD) and disruptive behavior disorders such as oppositional defiant disorder (ODD) and conduct disorder (CD) with its specifier for callous-unemotional (CU) traits are common and highly impairing mental disorders in childhood and adolescence (Herpers et al., 2012; Polanczyk et al., 2015). In the field of disruptive behavior disorders, symptoms of affective dysregulation have gained increasing attention in recent years. Yet, diagnostic classification systems still disagree in the assignment to diagnostic categories. While symptoms of affective dysregulation are conceptualized as core to the diagnosis of disruptive mood dysregulation disorder (DMDD) diagnosis in the 5th edition of the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 2013) and in the 10th edition of the International Classification of Diseases (ICD-10; World Health Organization, 1993), the ICD-11 (World Health Organization, 2019) confines to adding a specifier for chronic irritability/anger to the ODD diagnosis because of the difficulty in empirically distinguishing DMDD from other mental disorders, particularly ADHD and ODD (Evans et al., 2017; Lochman et al., 2015). According to both classification systems, the diagnosis of any of these mental disorders requires the presence not only of symptoms with a certain frequency and severity, but also of functional impairment (FI). Although symptoms and FI are often interrelated, they are not identical (Palermo et al., 2008). FI manifests across a wide range of domains and can be defined as *"the extent of restriction in a child's ability to perform important daily life activities including physical, social, and personal activities due to their health condition or to specific symptoms"* (Palermo et al., 2008, p. 984). For example, the presence and severity of symptoms may interfere with children's ability to attend school or to maintain friendships. In fact, many children and adolescents or their parents seek help because of FI rather than the presence of symptoms (Epstein & Weiss, 2012). Studies assessing the relationship between symptoms and FI have yielded mostly moderate associations (reviewed by Rapee et al., 2012). In line with this, not all individuals meeting the symptom criteria for an ADHD or ODD/CD diagnosis demonstrate FI as a result of their symptoms, while others with only subthreshold symptoms may show marked levels of impairment (Arildskov et al., 2022; DuPaul et al., 2014; A. Pickles et al., 2001). The assessment of FI is not only important for making diagnoses and verifying the need for treatment, but also for treatment planning, identifying maintaining factors of psychiatric symptoms and treatment targets, and evaluating treatment success (DuPaul, 2022; Haack & Gerdes, 2011; Winters et al., 2005). Therefore, measures of FI provide important additional information to measures of symptoms.

Relationship between symptom dimensions and functional impairment domains

Previous studies have demonstrated considerable variability in how symptom dimensions (e.g. ADHD inattention, hyperactivity-impulsivity) relate to various domains of FI. With regard to ADHD, the findings generally suggest that the inattention dimension is especially associated with academic impairment (Garner et al., 2013; Massetti et al., 2008; Willcutt et al., 2012; Zoromski et al., 2015). By comparison, the ADHD hyperactivity-impulsivity dimension generally shows the strongest associations with impaired social functioning, e.g. in the form of classroom disruption or peer exclusion (Garner et al., 2013; Willcutt et al., 2012). In terms of symptoms of disruptive behavior disorders, previous findings generally suggest that a greater severity of ODD problems is particularly related to impaired relationships with peers and family members (J. D. Burke et al., 2014; Dose et al., 2019; Kernder et al., 2019). Similarly, CD problems were found to predict workplace problems in young adults (J. D. Burke et al., 2014), and CU traits were found to predict impaired social functioning in children (Haas et al., 2018). Notably, most of these associations remained significant after controlling for demographic factors and comorbid symptoms (J. D. Burke et al., 2014; Garner et al., 2013; Haas et al., 2018; Massetti et al., 2008; Power et al., 2017; Willcutt et al., 2012).

Yet, despite the growing body of literature on FI associated with symptoms of ADHD and disruptive behavior disorders, there is still a need for clarification regarding the relationship between *individual* symptoms and different domains of FI. For example, it remains unclear whether certain symptoms are more impairing than others, and if so, what the relative importance of these symptoms is compared to other symptoms. This gap in the research is highly clinically relevant given the large variability in the symptoms experienced by patients (Mota & Schachar, 2000; Zoromski et al., 2015).

Relationship between individual symptoms and global functional impairment

To address this need for clarification, some studies have investigated the impact of individual symptoms, rather than symptom dimensions, on global FI. With respect to ADHD, Mota and Schachar (2000) suggested that some individual ADHD symptoms, including, for example, the teacher-rated *leaves seat* and the parent-rated *blurts out*, predict global FI in children better than other symptoms. In a more recent study, Zoromski et al. (2015) examined teacher ratings and found that certain symptoms of inattention (i.e. *does not listen* during early childhood, *does not follow through* during adolescence) and certain symptoms of hyperactivity-impulsivity (i.e. *on the go* during early childhood, *leaves seat* during middle childhood,

interrupts or intrudes during adolescence) showed the most robust relationships with classroom impairment. Regarding other externalizing behavior problems, research has shown that particularly ODD symptoms reflecting irritability (e.g. *anger* and *temper outbursts*) are associated with FI in children (Kolko & Pardini, 2010; Wesselhoeft et al., 2019) and that these symptoms predict greater FI than do other symptoms, even after controlling for other mental disorders (Dougherty et al., 2015). While this previous work has provided valuable insights into the relationship between symptoms and FI, additional exploration is needed. In particular, in previous studies, ratings for various domains of FI were combined into an overall, global score (Dougherty et al., 2013, 2015; Mota & Schachar, 2000) or the assessment of impairment was limited to classroom settings (Zoromski et al., 2015). A more nuanced approach, which considers the associations between individual symptoms and different FI domains in more detail while simultaneously accounting for ADHD as well as ODD, CD, and CU symptoms within a common approach, could help to enhance the understanding of these probably complex interrelations.

Network analysis

Network analysis offers such a more nuanced approach, which could potentially uncover individual symptom relations and associations with different FI domains (Borsboom & Cramer, 2013). From a network perspective, mental disorders are conceptualized as networks of mutually interacting symptoms (Borsboom & Cramer, 2013). Psychological networks may be depicted as structures comprising nodes (e.g. symptoms, FI domains), which are connected via edges (e.g. positive or negative associations). Some nodes can have a more dominant, central position in a network, for example, as expressed by stronger connections with other nodes (Borsboom & Cramer, 2013; Epskamp et al., 2018). Moreover, network analysis helps to uncover the unique strongest connections between nodes by relying on regularized partial correlations. This approach is somewhat similar to conducting multiple linear regressions at once, but with the advantage of reducing the risk of false positives (Epskamp & Fried, 2018). Hence, it is possible to determine which individual symptom shows the strongest association with a particular FI domain.

In recent years, many studies have applied network analysis to investigate how symptoms of different mental disorders are interrelated, particularly in the fields of anxiety and mood-related disorders (reviewed by Contreras et al., 2019). In the field of externalizing disorders, network analysis has been applied to symptoms of ADHD (Burns et al., 2022; Goh et al., 2020; Goh, Martel, et al., 2021; Martel et al., 2016, 2021; Preszler et al., 2020; Silk et al.,

2019), ODD (Smith et al., 2017), or ADHD and ODD together (Martel, Levinson, et al., 2017; Preszler & Burns, 2019), to CU traits (Bansal et al., 2020; Deng et al., 2021), and to CU traits in conjunction with ODD and CD (Bansal et al., 2021), in samples ranging from preschool age to adulthood. Furthermore, a recent study explored the relations between ADHD symptoms, executive functioning, and temperament traits, and found support for the primary role of effortful control as a potential risk marker for the characterization of ADHD across childhood and adolescence (Goh, Smith, et al., 2021).

To the best of our knowledge, so far, only three studies have focused on the associations between ADHD symptoms and different FI domains using network analysis (Burns et al., 2022; Goh et al., 2020; Goh, Martel, et al., 2021). Goh et al. (2020) explored individual ADHD and sluggish cognitive tempo symptoms and their relations with multiple domains of FI in a nationally representative sample of 1,742 children and adolescents (age: M = 11.51, SD = 3.36, range 6 – 17 years). The results revealed that in particular, the following eight symptoms were related to various FI domains and especially to the domains of academic and social impairment: *difficulties following through on instructions, inability to stay seated, acting without thinking, impatience, disinhibition, apathy/withdrawal, slowness*, and *lacking initiative*. Similarly, Goh, Martel, et al. (2021) explored individual ADHD symptoms (with an expanded impulsivity set) and their relations with multiple FI domains using network analysis and random forest regression in a nationally representative sample of 1,249 adults. The results from both techniques revealed that in particular, three inattention symptoms (*difficulty organizing, does not follow through, makes careless mistakes)* and one hyperactivity symptom (*difficulty engaging in leisure activities*) were strongly associated with global FI and FI, especially in the domains of social and interpersonal relationships, and difficulties maintaining structure in daily life. Finally, Burns et al. (2022) applied network and latent variable models to mother, father, and teacher ratings of ADHD inattention symptoms, sluggish cognitive tempo, and depressive symptoms in a sample of 2,142 Spanish children (age range 8–13 years; M = 10.30, SD = 1.21). Most interestingly, across all three sources, the same two ADHD inattention symptoms *difficulty keeping attention focused during tasks* and *avoids, dislikes or is reluctant to engage in tasks that require sustained mental efforts* showed unique relations with academic impairment. To summarize, each of these network studies uncovered important aspects of psychopathology, including the mutual associations among symptoms and the centrality of symptoms in a mental disorder network. However, the important role of FI has rarely been considered. To our knowledge, no study to date has identified which individual ADHD, ODD, CD, and CU symptoms show the strongest associations with particular domains of FI in a

clinical sample of school-age children, in whom the associations between symptoms and impairments are probably more pronounced and relevant than in a community sample (Borsboom, 2017).

The present study

The overall aim of this study was to enhance the understanding of how individual symptoms of ADHD and disruptive behavior disorders (i.e. ODD, CD, CU symptoms) may differentially relate to global FI and to FI in the five domains of psychological strain; home life and family members; relationships with adults; relationships with children and recreational activities; and academic performance. We extended previous research according to four important aspects: (a) We assessed FI domains specifically related to ADHD symptoms and, likewise, specifically related to ODD/CD/CU symptoms. (b) Consistent with current evidence, we included not only symptoms of ODD, but also affective dysregulation symptoms, as suggested in the ICD-11 for the subtype ODD with chronic irritability/anger (Evans et al., 2017; World Health Organization, 2019). (c) We provided another rater perspective (i.e. clinician ratings), since structured clinical interviews may be considered as the gold standard for diagnosing mental disorders (Rettew et al., 2009). (d) We extended previous findings by using a clinical sample of school-age children.

First, we determined the impact of individual ADHD symptoms or ODD/CD/CU symptoms, respectively, on global FI using linear regression, combined with relative importance analyses. In line with the results of previous studies, we assumed that the ADHD symptoms *leaves seat, blurts out*, *interrupts,* or *organizational skills* (Goh, Martel, et al., 2021; Mota & Schachar, 2000; Zoromski et al., 2015) and the ODD-related symptoms of irritability, such as *loses temper*, *touchy*, or *angry* (Dougherty et al., 2013; Wesselhoeft et al., 2019) would have the highest impact on global FI. Second, we estimated two psychological networks to identify the unique strongest associations between individual ADHD symptoms or ODD/CD/CU symptoms, respectively, and multiple FI domains. For the ADHD network, we expected that the symptoms most strongly associated with academic FI would originate from the inattention domain (Burns et al., 2022; Garner et al., 2013; Goh et al., 2020; Massetti et al., 2008; Willcutt et al., 2012; Zoromski et al., 2015), while symptoms most strongly associated with social impairment would originate from the hyperactivity-impulsivity domain (Garner et al., 2013; Goh, Martel, et al., 2021; Willcutt et al., 2012). For the ODD/CD/CU network, we expected the symptoms most strongly associated with academic FI to originate from the CD domain (J. D. Burke et al., 2014), the symptoms most strongly associated with impaired home

life and family members and with relationships with adults to originate from the ODD or CD domain (Dose et al., 2019; Kernder et al., 2019), and the symptoms most strongly associated with impaired relationships with children to originate from the ODD or CU domain (J. D. Burke et al., 2014; Dose et al., 2019; Haas et al., 2018; Kernder et al., 2019).

Methods

Participants and procedure

Data for the present analyses were collected within the ESCAschool study (ESCAschool: *Evidence-based, Stepped Care of ADHD in school-aged children*), which is part of the ESCAlife consortium and involves multiple study sites in Germany. The ESCAschool study was designed to investigate an evidence-based, individualized, stepwise- intensifying treatment program for children diagnosed with ADHD, which is based on behavioral and pharmacological interventions.

Children and their families were eligible for participation if the child met diagnostic criteria for ADHD according to the DSM-5, was aged between 6;0 and 11;11 years, and attended school. The following exclusion criteria were applied: child IQ < 80; a child diagnosis of a pervasive developmental disorder, schizophrenia, bipolar disorder, severe depressive episode, epilepsy, or heart disease; insufficient German language or reading skills of the parents; a current or planned behavior therapy for child ADHD or ODD occurring at least weekly; a known non-response of the child to all standard ADHD medication; and psychotropic medication of the child other than for the treatment of ADHD/antipsychotic medication other than for the treatment of disturbances of impulse control. Further details on the background and procedures are outlined in the study protocol (Döpfner et al., 2017).

The present study analyzed baseline data (i.e., data collected before any intervention) of 474 children (M = 8.90, SD = 1.49; 81% males). The screening to check whether the participants met the diagnostic criteria for ADHD relied on the DSM-5-based *Clinical Parent Interview for Externalizing Disorders in Children and Adolescents [Interview-Leitfaden für Externale Störungen]* (ILF-EXTERNAL; Görtz-Dorten et al., 2022; see Measures section). If children were receiving ADHD medication prior to the study, parents described their child's behavior with and without medication. For the present analyses, we investigated the children's symptomatology without medication.

Clinical diagnoses of ADHD, ODD, CD, and CU symptoms were based on the interview ILF-EXTERNAL according to the DSM-5. In addition, to assess comorbid symptoms, all clinicians applied a clinical diagnostic checklist (DCL-SCREEN) from the *Diagnostic System for Mental*

Disorders in Children and Adolescents based on the ICD-10 and DSM-5 [Diagnostik-System für psychische Störungen nach ICD-10 und DSM-5 für Kinder und Jugendliche - III] (DISYPS-III; Döpfner & Görtz-Dorten, 2017). The sample used in the present study also included children who did not fulfill DSM-5 criteria for an ADHD diagnosis. These screening negatives (n = 32, 72% males) were characterized by subclinical ADHD symptoms (Thöne et al., 2020). The ESCAschool study was registered at the German Clinical Trials Register (identifier: DRKS00008973).

Measures

Clinical Parent Interview for Externalizing Disorders in Children and Adolescents (ILF-EXTERNAL)

The ILF-EXTERNAL is part of the German semi-structured *Interview for Diagnosing Mental Disorders According to the DSM-5 in Children and Adolescents* [*Interview-Leitfäden zum Diagnostik-System für psychische Störungen nach DSM-5 für Kinder und Jugendliche*] (DISYPS-ILF; Görtz-Dorten et al., 2022) and comprises two interview sections. The first section assesses ADHD criteria and the second assesses criteria for ODD, CD, including CU traits according to the DSM-5.

In terms of ADHD criteria, the ILF-EXTERNAL measures inattention symptoms (nine items), hyperactivity-impulsivity symptoms (nine items), and FI associated with ADHD symptoms (five items) in the domains of home life and family members, relationships with adults, relationships with children and recreational activities, academic performance, and psychological strain.

In terms of ODD, CD, and CU criteria, the ILF-EXTERNAL assesses ODD symptoms (eight items), CD symptoms (15 items), disruptive mood dysregulation symptoms (five items in total; two items assessing affective dysregulation and three items assessing irritability/anger associated with ODD), CU symptoms (11 items) and, resembling the ADHD section, functioning and psychological strain associated with these symptoms (five items).

Following a semi-structured interview format, clinicians rate each item on a 4-point Likert scale ranging from 0 (age-typical / not at all), to 3 (very much), with higher scores reflecting higher symptom severity. Item scores of 2 and higher are interpreted as clinically relevant and considered to fulfill the DSM-5 symptom criteria. It should be noted that the ILF-EXTERNAL captures the full set of diagnostic criteria for a given mental disorder (i.e. each item explores a DSM-5 symptom criterion), thereby avoiding issues associated with skip outs in symptom networks (Hoffman et al., 2019). However, for the present analyses, we excluded

the CD items assessing aggressive and antisocial symptoms, which are recommended for use in participants aged 11 years or older, due to obvious floor effects.

Interview training. All interviewers involved in the recruitment of patients for the ESCAschool study were trained psychologists or educators with a master's degree, doctoral students, or in training to become a child and adolescent psychotherapist/psychiatrist. The interviewers received standardized training on how to administer and score the ILF-EXTERNAL, which included viewing a practice video. Finally, all interviewers were given guidance by supervisors if they experienced any difficulties in scoring the ILF EXTERNAL.

Psychometric properties of the ILF-EXTERNAL. Psychometric evaluations revealed a good to excellent interrater reliability [intraclass correlation (ICC) on the scale level: ICC(1,1) = .83 - .95; ICC(1,3) = .94 - .98] as well as convergent and divergent validity of the ILF-EXTERNAL scale scores with parent ratings of the respective constructs (Thöne et al., 2020). Furthermore, the basic factorial configuration of the ILF-EXTERNAL scale scores was confirmed (Thöne et al., 2021). In the current sample, internal consistencies (Cronbach's alpha: $.60 \leq \alpha \leq .87$) and item-total correlations ($.22 \leq r_{it} \leq .68$) of the corresponding scales were, for the most part, satisfactory to good (see Table S1). The internal consistencies of the ADHD FI scale ($\alpha = .62$) and the CD scale ($\alpha = .60$) were somewhat below the satisfactory range. The following items demonstrated item-total correlations below $r_{it} = .30$: A01 *Careless*, A06 *Concentration* (both ADHD symptoms), F05 *Interferes with Educational Activities* (FI related to ADHD), B03 *Cruel to Animals*, B05 *Steals Without Confrontation* (both CD items), C04c *Manipulates* (CU item). However, excluding any of these items did not noticeably change the Cronbach's alpha of the respective scales.

Multicollinearity and missing data checks. Since symptoms of ADHD and, likewise, ODD/CD/CU symptoms are usually correlated with each other, we performed multicollinearity checks using the variance inflation factor. The variance inflation factor did not exceed the cut-off value of 5 for any symptom (Table S2), indicating no considerable multicollinearity problems (Craney & Surles, 2002). The amount of missing data per item is reported in Table S2 and varied from 0% (ADHD symptoms) through 5% (ODD, CD symptoms) to 6% (CU symptoms) and 7% (FI related to ODD/CD/CU symptoms).

Data analytic plan

First, we performed two multiple linear regression analyses, combined with relative importance analyses, in order to estimate the impact of individual symptoms on global FI. Second, we estimated two networks, one comprising ADHD symptoms and FI and the other

comprising ODD/CD/CU symptoms and FI. Both networks included five FI domains (nodes F01 – F05), which were specifically associated with symptoms of ADHD or ODD/CD/CU, respectively. Descriptive statistics were calculated using SPSS version 27. The network and regression analyses were carried out using the R statistical software version 4.0.3 in RStudio 1.3.1093 for macOS.

Multiple linear regressions and relative importance analysis

To examine the impact of individual symptoms on global FI, we computed the average score across the five FI domains (F01 – F05) associated with ADHD or ODD/CD/CU symptoms, respectively. We used the 18 ADHD symptoms and the 26 ODD/CD/CU symptoms, respectively, as predictors of global FI in a multiple linear regression model, additionally controlling for age and gender as covariates. An *a priori* power analysis using G-Power 3.1 (Faul et al., 2009) yielded a required sample size of $n = 222$ for the ADHD model and $n = 253$ for die ODD/CD/CU model, respectively, assuming a moderate effect size ($f^2 = 0.15$), a power of .95 and a significance level of 5%. Next, we allocated unique R^2 shares (i.e., proportion of explained variance) to each regressor to determine how much unique variance each individual symptom shared with global FI. We used the *R* package RELAIMPO (Grömping, 2006) which provides several metrics for assessing relative importance in linear models. The recommended metric is lmg (like in Lindeman, Merenda, and Gold, 1980) which estimates the importance of each regressor by splitting the total R^2 into one non-negative R^2 share per regressor. These non-negative R^2 shares sum up to the total R^2. To estimate the importance of each regressor, the contribution of each predictor at all possible entry points into the model is calculated and the average of these contributions is taken (i.e., an estimate for each variable is obtained by computing as many regressions as there are possible orders of regressors, and then the average of the individual R^2 values across all models is taken). The relative importance estimates were then adjusted to add up to 100% to facilitate their interpretation.

Then, we compared two multiple linear regressions models for the ADHD and ODD/CD/CD symptoms, respectively (for a similar analysis, please see Fried & Nesse, 2014). In the first model (unconstrained model), all regression weights for symptoms were free to vary, whilst in the second model (constrained model), the symptom weights were constrained to be equal. While the unconstrained model allows for differential associations between functional impairment and symptoms, the constrained model hypothesizes that symptoms are equally associated with functional impairment. A chi-square difference test was used to compare the two nested models.

Network analysis

We constructed undirected Gaussian graphical models (GGMs) using the *R* package *bootnet* version 1.4.3 (Epskamp et al., 2018). In these network models, the edges connecting the nodes represent estimates of partial correlations. In undirected GGMs, the edges can be interpreted as conditional dependence relations among nodes. If the analysis reveals that two nodes are connected, they are dependent after controlling for all other symptoms in the network. If no edge emerges, two nodes are interpreted as conditionally independent (Epskamp & Fried, 2018). As stated above, we estimated two network models: Our first network comprised 18 ADHD symptoms and five FI domains and our second network comprised 26 ODD/CD/CU symptoms and five FI domains. We decided to estimate two networks given that we assessed FI specifically related to ADHD symptoms and FI specifically related to ODD/CD/CU symptoms. To prevent spurious edges and to estimate a more parsimonious network model, we adopted the graphical least absolute shrinkage and selection statistical regularization technique, coupled with the extended Bayesian information criterion (EBICglasso) model selection (tuning hyperparameter = 0.5). This regularization technique pulls small associations (edges) to zero, which results in their removal from the network as potentially false positive edges (Epskamp & Fried, 2018). We used Spearman correlations to account for our ordinal data structure and pairwise complete observations to handle missing data (Epskamp et al., 2018). The resulting networks were visualized using the Fruchterman–Reingold algorithm from the *R* package *qgraph* version 1.6.9, which places related nodes closer to each other.

Furthermore, we assessed the accuracy of the edge weights by randomly resampling participants (nonparametric bootstrapping; 2,500 iterations) and estimating the 95% bootstrapped confidence intervals (Epskamp et al., 2018). Then, we performed bootstrapped difference tests (2,500 iterations) to investigate whether two edge weights significantly differed from each other. It should be noted that a correction for multiple testing when carrying out bootstrapped difference tests in network analysis has not yet been developed (Epskamp et al., 2018).

In addition, we ran a simulation study on the performance of network estimation by varying sample size using the *netSimulator* function from the *bootnet* package (2,500 iterations; Spearman correlations; EBICglasso network). The simulation analysis yields plots which demonstrate the sensitivity and specificity of the network as well as correlations between the "true" and estimated edges given our network structure (Epskamp et al., 2018).

Associations of externalizing symptoms with functional impairment domains

For each FI domain (nodes F01 – F05), we identified the ADHD symptom and the ODD/CD/CU symptom which was most strongly associated with the respective domain by determining the largest edge weight (i.e., the strongest partial correlation). These partial correlations reflect the associations between individual symptoms and a particular FI domain after controlling for the influence of all other symptoms and FI domains in the network. In this context, it should be mentioned that previous research relied on bridge analysis to explore which symptoms may act as "middlemen" regarding the associations between symptoms and FI (e.g. Goh et al., 2020; Goh, Martel, et al., 2021). Our approach contributes to this research insofar as we were interested in identifying the unique strongest associations between different FI domains and individual symptoms. As the EBICglasso regularization technique pulls smaller, potentially spurious edges to zero, the remaining edges may be considered as sufficiently strong for inclusion in the network model (Epskamp & Fried, 2018). Notably, as weaker edges should be interpreted with some caution, our interpretation concentrates on the stronger, and thus likely more stable, edges in the network structures.

Results

Sample characteristics

The present sample (Table 1) comprised school-age children mainly diagnosed with ADHD (94%) according to the DSM-5. The majority of these children (98%) as well as their mothers (87%) and fathers (84%) were born in Germany. About 90% of the families spoke predominantly German at home. Most of these children met the diagnostic criteria for the ADHD combined type (44%), followed by the ADHD predominantly hyperactive-impulsive type (39%). Probably due to this high percentage of children diagnosed with ADHD, the mean ADHD-related scale scores were generally higher than the ODD/CD/CU-related scale scores (Table S1). The intercorrelations between the ADHD-related items (i.e. symptoms and FI domains) or the ODD/CD/CU-related items (i.e. symptoms and FI domains), respectively, were predominantly positive and low to moderate (see Figure S1).

Multiple linear regression analyses and relative importance analyses

In our multiple linear regression models, we used 18 ADHD symptoms and 26 ODD/CD/CU symptoms, respectively, as predictors of global FI, additionally controlling for age and gender as covariates.

In the ADHD model, a significant equation was found [F (20, 451) = 12.69, $p < .001$], with an R^2 of 36%. It also emerged that seven out of the 18 ADHD symptoms as well as gender significantly predicted global FI ($p < .05$; see Table S3). A subsequent relative importance analysis revealed that different ADHD symptoms had markedly different effects on global FI, with estimates of the R^2 contributions ranging from 1% (A07 *Loses Things*) to 15% (A08 *Easily Distracted*; Figure 1a). The relative importance estimates were similar when the inattention dimension and the hyperactivity-impulsivity dimension were considered separately (inattention: R^2 = 52%; hyperactivity-impulsivity: R^2 = 48%). When comparing the ADHD unconstrained model (all regression weights for symptoms were free to vary) and the ADHD constrained model (the regression weights were constrained to be equal), the unconstrained model fit the data significantly better than the constrained model ($\chi^2_{diff} = 50.07$, $df_{diff} = 19$, $p < .001$). Also, the R^2 = 36% of the ADHD unconstrained model was higher than the R^2 = 29% of the constrained model.

In the ODD/CD/CU model, a significant equation was found [F (28, 411) = 16.06, $p < .001$], with an R^2 of 52%. It also emerged that 10 out of the 26 ODD/CD/CU symptoms significantly predicted global FI ($p < .05$; see Table S4). A subsequent relative importance analysis revealed that different symptoms had markedly different effects on global FI, with estimates ranging from < 1% (C03a *Indifferent to Poor Performance*; C04a *Shallow, Deficient Affect*) to 10% (A04 *Argues with Adults*; Figure 1b). When adding up the relative importance estimates of the symptoms belonging to the ODD, CD, and CU dimensions, the results showed differential contributions of the dimensions to explaining the variance in global FI: ODD (55%) was most strongly associated with global FI, while CD (24%) and CU (20%) were less strongly associated with global FI. When comparing the ODD/CD/CU unconstrained model and the ODD/CD/CU constrained model, the unconstrained model fit the data significantly better than the constrained model ($\chi^2_{diff} = 83.35$, $df_{diff} = 27$, $p < .001$). Also, the R^2 = 52% of the ODD/CD/CU unconstrained model was higher than the R^2 = 42% of the constrained model.

For interested readers, we provide a supplementary multiple regression analysis with global FI averaged across ADHD and ODD/CD/CU as a response variable and all externalizing symptoms as regressors in the online supplement (Table S5). Noteworthy, all significant predictors from this supplementary analysis also emerged as significant predictors in our ADHD and ODD/CD/CU regression analyses, respectively (e.g., ADHD A08 *Easily Distracted* was identified as a highly significant predictor in both analyses).

Network analysis

Network structures for the ADHD and ODD/CD/CU symptoms and associated functional impairment domains

As visualized in Figure 2, there were stronger associations with items within the same dimension (e.g. Inattention) than across dimensions. The results of our simulation studies on the performance of the ADHD and ODD/CD/CU network estimations by varying sample size are depicted in Figure S2.

The strongest associations of functional impairment domains with symptoms

Regarding the ADHD network, the strongest partial correlations were found between FI related to psychological strain (F01) and A08 *Easily Distracted* ($\rho_{xy \bullet z}$ = .09), between FI related to home life and family members (F02) and A03 *Does not Listen* ($\rho_{xy \bullet z}$ = .09), between FI related to relationships with adults (F03) and B02 *Leaves Seat* ($\rho_{xy \bullet z}$ = .07), between FI related to relationships with children and recreational activities (F04) and B08 *Interrupts, Intrudes* ($\rho_{xy \bullet z}$ = .15), and between FI related to academic performance (F05) and A06 *Concentration* ($\rho_{xy \bullet z}$ = .10).

Regarding the ODD/CD/CU network, the strongest partial correlations were found between FI related to psychological strain (F01) and A01 *Loses Temper* ($\rho_{xy \bullet z}$ = .07), between FI related to home life and family members (F02) and A04 *Argues with Adults* ($\rho_{xy \bullet z}$ = .18), between FI related to relationships with adults (F03) and B05 *Steals Without Confrontation* ($\rho_{xy \bullet z}$ = .08), between FI related to relationships with children and recreational activities (F04) and B01 *Physical Fights* ($\rho_{xy \bullet z}$ = .11), and between FI related to academic performance (F05) and B02 *Bullies, Threatens, Intimidates* ($\rho_{xy \bullet z}$ = .06).

Following our edge-weight accuracy checks, results from non-parametric bootstrapping generally indicate accurate estimations, since the sample values lie within the bootstrapped confidence intervals and the bootstrap mean values are generally well aligned with the sample values (Figure S3). As the bootstrapped confidence intervals were relatively wide, we recommend some caution when interpreting the presence and strength of weaker edges. The aforementioned associations (i.e. edge weights) between each FI domain and symptoms of the ADHD and ODD/CD/CU networks were significantly stronger than most of the other edge weights in the respective networks (Figures S4, S5). Importantly, while the associations may appear to be weak, it should be kept in mind that these are partial correlations after regularization (shrinking). As supplementary analyses, we also report the strongest associations between different domains of functional impairment and all externalizing symptoms in a

combined network. The results are mostly consistent with the two separate ADHD and ODD/CD/CU networks (e.g., ADHD F01 *Psychological Strain* is most strongly associated with A08 *Easily Distracted*). However, the results should be interpreted with some caution, as the network may be rather unstable due to the high number of nodes in a relatively small sample (Figures S6, S7). Furthermore, we aimed to support the interpretations drawn from our network analyses with additional regression analyses. We therefore calculated separate ordinal logistic regression analyses with the different FI domains as response variables to account for the ordinal data structure (Tables S6, S7). Most intriguingly, we were able to establish method equivalence insofar as those symptoms from the network analyses that were most strongly associated with a particular FI domain also emerged as significant predictors in the ordinal regression analyses.

Discussion

The present study aimed to enhance the understanding of how individual ADHD and ODD/CD/CU symptoms differentially relate to global FI as well as FI in the five domains of home life and family members, relationships with adults, relationships with children and recreational activities, academic performance, and psychological strain. Our findings contribute to a growing body of literature on the importance of analyzing the associations between individual symptoms and different domains of FI and highlight that symptom-based approaches can be clinically useful.

Distraction (ADHD) and Arguments with Adults (ODD) explained a large proportion of the variance in functional impairment

Overall, individual symptoms had differential impacts on global FI. In particular, A08 *Easily Distracted*, B03 *Runs, Climbs*, and B08 *Interrupts, Intrudes* explained a large proportion of the variance in global FI related to ADHD symptoms. These findings are largely consistent with previous studies, which reported that particularly ADHD hyperactivity-impulsivity symptoms were linked to global FI in children and adolescents (Mota & Schachar, 2000; Zoromski et al., 2015). Furthermore, A04 *Argues with Adults*, A05 *Refuses to Comply*, and B01 *Physical Fights* explained a large proportion of the variance in FI related to ODD/CD/CU, whereas CU symptoms made few unique contributions to global FI. One possible explanation for these findings could be that the nature of "covert" CU-symptoms is assessed to be less functionally impairing relative to typical, "overt" ODD/CD symptoms (e.g., losing temper, arguing) in the clinical interview. Interestingly, however, our additional logistic regression

analyses showed that individual CU traits were particularly strongly associated with FI related to relationships with adults. Of note, these conclusions are rather preliminary and require replication before more precise conclusions can be drawn about the relationships between CU traits and FI.

Inattention strongly relates to academic, family, and psychological strain, while hyperactivity-impulsivity strongly relates to social impairment

The interpretation of the ADHD network structure suggests that symptoms most strongly associated with FI in the domains of academic performance, home life and family members as well as with psychological strain originated from the inattention dimension. These results are consistent with previous studies, which reported that ADHD inattention symptoms show strong associations with academic impairment (Garner et al., 2013; Massetti et al., 2008; Willcutt et al., 2012; Zoromski et al., 2015). Furthermore, our analyses add to previous findings by providing another rater perspective (i.e. clinician ratings), investigating a clinical sample of school-age children, and applying novel analysis techniques. Moreover, we found that the ADHD inattention symptom A08 *Easily Distracted* was the symptom most strongly associated with FI related to psychological strain. This finding is quite interesting from a clinical perspective given that inattention symptoms have previously been regarded as less impairing than hyperactivity-impulsivity symptoms (Willcutt et al., 2012). However, while the combined presentation of inattention and hyperactivity-impulsivity symptoms is generally associated with a high degree of FI (Willcutt et al., 2012), if inattention symptoms are present, we recommend paying particular attention to exploring psychological strain, both to ensure adequate and appropriate treatment and to inform future research in this area. By contrast, the symptoms most strongly associated with FI in the domains of relationships with children and recreational activities, as well as relationships with adults, originated from the hyperactivity-impulsivity dimension. These results are consistent with previous research indicating that ADHD hyperactivity-impulsivity symptoms are strongly associated with impaired social functioning (Garner et al., 2013; Goh, Martel, et al., 2021; Willcutt et al., 2012). From a clinical perspective, the present findings highlight the need for a thorough assessment of impairment in the domains of relationships with children and adults in the presence of hyperactivity-impulsivity symptoms (e.g. B02 *Leaves Seat*, B08 *Interrupts, Intrudes*) and the subsequent selection of appropriate interventions, for example interventions focusing on positive parent-child interactions and communication skills.

Symptoms of ODD and CD more strongly relate to impairment than do CU traits

The interpretation of the ODD/CD/CU network suggests that the symptoms most closely linked to the examined FI domains originated from the ODD and CD dimension, while FI was less strongly associated with CU traits. In particular, the symptoms most strongly associated with FI in the domains of academic performance, relationships with children and recreational activities as well as relationships with adults were CD symptoms, whereas the symptoms most strongly associated with FI related to home life and family members as well as with psychological strain were ODD symptoms. Interestingly, CU traits remained a rather self-contained dimension in the network structure, that is, they were neither strongly associated with any FI domain nor with ODD/CD symptoms. This finding is particularly interesting given that CU traits can be added as a specifier to the CD diagnosis in the DSM-5 (American Psychiatric Association, 2013) and as a specifier to the ODD or CD diagnosis in the ICD-11 (World Health Organization, 2019). Moreover, the finding is in line with previous research on the structure underlying externalizing behavior disorders in children (Castagna et al., 2021; Haas et al., 2018; Thöne et al., 2021).

Clinical implications

Several clinical implications of these findings with respect to clinical assessment and therapy can be derived. First, given the only low to moderate correlations between externalizing symptoms and related functional impairment, as well as the importance of single, specific symptoms for the prediction of FI in different domains, we agree with the growing consensus that in order to gain a comprehensive understanding of psychopathology, the assessment of FI is as important as the assessment of symptoms (Arildskov et al., 2022; DuPaul, 2022). To further clarify the relationship between symptoms and FI, a recent study by Arildskov et al. (2022) examined whether the relationship between ADHD symptom severity and FI was nonlinear (i.e. whether there was a symptom severity threshold linked to a marked increase in impairment). The authors found a gradual linear increase in impairment with higher symptom severity, suggesting that the current symptom severity threshold for an ADHD diagnosis may be arbitrarily defined with respect to the presence or absence of FI (Arildskov et al., 2022). In line with this, not all individuals meeting the symptom-based criteria for an ADHD or ODD/CD diagnosis suffer from FI as a result of their symptoms, while others with only subthreshold symptoms show marked levels of impairment (DuPaul et al., 2014; A. Pickles et al., 2001). Arildskov et al. (2022) thus highlight *"the continuing need for the clinical assessment and diagnosis of ADHD to be based on two independent decisions: one about the symptom threshold*

and one about functional impairment with both decisions having a certain degree of arbitrariness and social subjectivity" (Arildskov et al., 2022, p. 5). As discussed by DuPaul (2022), in the current diagnostic systems, the functional impairment criterion seems to be considered categorically (i.e., either present or absent), while lacking a precise operationalization and an explicit dimensional threshold. This - along with the influence of co-occurring symptoms on impairment ratings - illustrates the challenges clinicians face when making decisions on functional impairment (DuPaul, 2022).

Second, we emphasize the need for a more nuanced perspective on the association of specific symptoms with FI. For example, in our analyses, the ADHD symptom A08 *Easily Distracted* explained about five times as much variance in global FI as A01 *Careless*, even though both symptoms originate from the inattention dimension. In line with this, we found that constraining regression weights of ADHD and ODD/CD/CU symptoms, respectively, to be equal when predicting global FI led to a significantly reduced model fit and lower R^2 compared freely estimating symptom contributions. These findings suggest that symptoms have differential impacts on functional impairment and that differentially weighing symptoms does indeed lead to notable improvements in predicting impairment. Interestingly, similar findings of variable associations with impairment were also reported for individual depression symptoms (Fried & Nesse, 2014). In light of these differential symptom-impairment relations, we recommend that besides symptom dimensions, *individual* symptoms should be considered in the process of diagnostic assessment. In the long term, research on the associations between single symptoms and FI in different domains could inform future modifications of the diagnostic criteria. For instance, instead of simple criterion counts in which various DSM-5 criteria are given equal weight (e.g. 6/9 symptoms of ADHD inattention), symptoms that show particularly strong associations with FI could be weighted based on the strength of their relationship with impairment.

Third, the results of the present study may also have important implications for the development of assessment measures, particularly screening instruments. As also suggested by Zoromski et al. (2015), and provided that further research is conducted, results of network analyses on symptom-impairment relations might inform the development of more economical screening instruments, which might initially include only symptoms that have particularly strong associations with FI domains. Such screening instruments might be valuable for both clinical research and practice. In research, they could precede the use of more extensive diagnostic batteries (Zoromski et al., 2015) and, for example, help identify participants eligible for a particular study more efficiently. In practice, several possible applications are conceivable.

Such screening instruments might be used by clinical psychologists, but also other health care or educational professionals, to identify children in need for closer observation, more comprehensive diagnostics, or (early) intervention, if possible, before impairment manifests (Zoromski et al., 2015). Moreover, given that impairments rather than the presence of symptoms are often the reason for seeking treatment (Epstein & Weiss, 2012), and provided that the association between particular symptoms and FI domains is replicated, these screening instruments could assist clinicians in identifying target areas for treatment and selecting appropriate interventions accordingly (Zoromski et al., 2015). In addition, the sensitivity and specificity of such screening tools compared with standard screening procedures could be evaluated by comparing both approaches with more extensive current diagnostic practice (Zoromski et al., 2015). Finally, on a more global level, Zoromski et al. (2015) suggest to evaluate gating questions in existing structured interviews in terms of their correspondence to symptoms most strongly associated with FI domains and to further investigate whether having gating questions about symptoms most strongly linked to FI domains can facilitate diagnostics or treatment planning, e.g., by weighing symptoms according to the strength of their relationship with FI.

Limitations and future directions

Several limitations need to be considered when interpreting the results of the present study. First, we concede that a cross-sectional data design cannot uncover causal processes, for example in terms of the causal associations between FI and symptoms (e.g. symptoms and FI may potentially influence each other). Moreover, it remains to be investigated how the associations between FI and symptoms develop across age (e.g. whether certain symptoms become more impairing in adolescence). Ideally, future studies should address these gaps using a longitudinal data design. Second, in terms of the diversity of our sample, all children included in this study were registered for participation in a randomized control trial. As these children met stringent inclusion and exclusion criteria, our findings need to be replicated in clinically referred children under routine care conditions. In this context, it might be criticized that our FI items may not be able to differentiate between FI related to ADHD and FI related to ODD/CD/CU symptoms. However, the only moderate scale correlations between ADHD-related FI and ODD/CD/CU-related FI (r = .53) indicate that the clinicians (and the parents who were interviewed) were indeed able to differentiate between impairment related to the two symptom domains, and consequently that ADHD- and ODD/CD/CU-related impairment are perhaps overlapping but nevertheless generally distinct constructs. Third, we must

acknowledge that we were unable to calculate confidence intervals and associated *p*-values as part of our relative importance analyses due to overly intensive computational efforts. We assume that these computational problems are related to our large number of regressors. Nevertheless, we would like to point out that the statistical significance of individual regressors can also be obtained from the multiple linear regression tables. Fourth, it should be noted that ADHD symptoms explained only 33% and ODD/CD/CU symptoms only 49% of the variance in global FI, underlining the likelihood that there are many sources of influence that contribute to the severity of FI, including variables not assessed in this study (e.g. parenting style, socioeconomic status). Since the DSM-5 diagnoses require the presence of symptom-related impairment, we limited the current analyses to the associated variables. However, the consideration of additional variables influencing FI in future studies could yield valuable contributions to both theoretical reflections and the advancement of treatments.

Conclusion

The current findings demonstrate that symptoms of ADHD or ODD/CD/CU, respectively, vary substantially in their associations with global FI. The total amount of variance in global FI explained by ADHD symptoms ranged from 1% (A07 *Loses Things*) to 15% (A08 *Easily Distracted*) and the amount of variance in global FI explained by ODD/CD/CU symptoms ranged from < 1% (C03a *Indifferent to Poor Performance*; C04a *Shallow, Deficient Affect*) to 10% (A04 *Argues with Adults*). Moreover, ADHD symptoms most strongly associated with FI in the domains of academic performance, home life and family members as well as with psychological strain originated from the inattention dimension (A03 *Does not Listen*, A06 *Concentration*, A08 *Easily Distracted*), whereas the symptoms most strongly associated with FI in the domains of relationships with children/adolescents and recreational activities as well as with relationships with adults originated from the hyperactivity-impulsivity dimension (B02 *Leaves Seat*, B08 *Interrupts, Intrudes*) . By contrast, the symptoms most closely linked to the examined FI domains related to ODD/CD/CU symptoms originated from the ODD and CD dimension (A01 *Loses Temper*, A04 *Argues with Adults*, B01 *Physical Fights*, B02 *Bullies, Threatens, Intimidates*, B05 *Steals Without Confrontation*), while FI was less strongly associated with CU traits. Our study contributes to this growing body of research in that diagnosticians should equally pay attention to both the severity and frequency of symptoms and the degree to which these symptoms are related to academic or social impairment. In particular, the present study provides first evidence on the extent to which individual symptoms are related to specific domains of functional impairment. In clinical practice, this could be of

particular interest for both diagnostics and treatment planning purposes, as the presence of certain symptoms might lead clinicians to particularly assess functional impairment in specific domains, which could also be a target for treatment. Of note, the results of this study are preliminary and require replication in larger samples before more precise conclusions can be drawn about the relationships between symptoms and impairment.

Acknowledgements

We thank all staff members who participated in the ESCAschool study in nine different study recruiting centers and who were involved in recruitment, study organization and management, blind rating, assessment and therapy, and the Clinical Trials Unit Freiburg: Veronika Frommberger, Alexander Hellmer, Lydia Herbstritt, Maria Huber, Carolin Jenkner, Stefanie Lade, Patrick Müller, Barbara Schilling, Carla Schneider, and Ralf Tostmann. Further thanks to Sarah Mannion for English-language proofreading.

Author contributions

A-KTh conceptualized the study design, conducted the formal data analysis, interpreted and visualized the results, and developed the first draft of the manuscript. She was involved in the recruitment and data acquisition of the ESCAschool study site in Cologne and is co-author of the DISYPS-ILF on which the results on the present study were based. MJ contributed to the interpretation of the results and critically revised the first draft of the manuscript. CD contributed to the interpretation of the results, contributed to the management and organization of the study, heads the telephone-assisted self-help trial performed in one treatment arm of the ESCAschool study, and critically revised the first draft of the manuscript. CH was involved in the development of the ESCAschool research proposal and organization of the study. LJ and A-KTr coordinate ESCAschool and contribute to the management and organization of the study. PV is a scientific staff member at the Cologne study site and was involved in patient recruitment and data acquisition. EvW coordinates the ESCAschool study, contributed to the implementation of the study, and provided supervision for patient treatment. TB coordinates the ESCAlife consortium and is co-principal investigator in the ESCAschool study. He made substantial contributions to the conception and design of the ESCAschool research proposal and heads the Mannheim study site of ESCAschool. KB heads the Marburg study site of ESCAschool, is PI for the ESCApreschool study, is a member of the ESCAlife consortium, is a co-applicant of the ESCAlife research project, and made substantial contributions to the conception and design of the ESCAlife study. DB made substantial contributions to the conception and design of the ESCAschool research proposal and the neuropsychological research battery and heads the sub-project ESCAbrain. UD coordinates the Tübingen study site of ESCAschool, contributes to the management and organization of the study, and was involved in patient recruitment and data acquisition. JG coordinates the Würzburg study site and contributed to the management and organization of the ESCAlife projects. JH heads the Essen study site of ESCAschool and made substantial contributions to the conception and design of the ESCAschool study. SH coordinates the ESCAlife consortium and made substantial contributions to the conception of the neuropsychological research battery and the neurofeedback protocol. MAH heads the Hamm/Bochum study site of ESCAschool and was involved in the development of the ESCAschool research proposal. MIH heads the Mainz study site of ESCAschool and was involved in the implementation of the ESCAlife projects. TJ was involved in the planning of the ESCAlife research projects and application for funding, is co-PI for the ESCAadol trial and was involved in the implementation of the ESCAlife projects at the Würzburg study site. AK is a scientific staff member at the Mannheim study site,

contributed to the implementation of the study, was involved in data acquisition, contributed to the management and organization of the (especially, ESCAbrain) study, and critically revised the first draft of the manuscript. JK coordinates ESCAschool at the Marburg study site, contributes to the management and organization of the study, and was involved in patient recruitment and data acquisition. TL heads the research department of the Hamm/Bochum study site of ESCAschool, contributed to the implementation of the study in Hamm, and supervised the realization of the trial in Hamm. SM heads the Mannheim study site of ESCAschool, coordinates the ESCAlife consortium, and made substantial contributions to the conception of the neuropsychological research battery and the neurofeedback protocol. LP heads the Göttingen study site of ESCAschool and was involved in the implementation of the ESCAlife projects. TR heads the Tübingen study site of ESCAschool and contributed to the implementation of the ESCAschool study. MR was involved in the planning of the ESCAlife research projects and application for funding, is co-PI for the ESCAadol trial, and was involved in the implementation of the ESCAlife projects at the Würzburg study site. HUS coordinates ESCAschool at the Göttingen study site, contributed to the management and organization of the study, and was involved in patient recruitment and data acquisition. PS coordinates the Tübingen study site of ESCAschool, contributes to the management and organization of the study, and was involved in patient recruitment and data acquisition. JW coordinates ESCAschool at the Essen study site and contributed to the management and organization of the study. MZ is a scientific staff member at the Mannheim study site, contributed to the implementation of the study, and was involved in patient recruitment and data acquisition. AGD developed the DISYPS- III system, developed the DISYPS-ILF, contributed to the interpretation of the results, and critically revised the first draft of the manuscript. MD is principal investigator of the ESCAschool trial, developed the basic ESCAschool study design, is part of the ESCAlife consortium, is head of the Cologne recruiting center for the ESCA children trials (ESCApreschool, ESCAschool and ESCAadol), developed the DISYPS- III system, developed the DISYPS-ILF, conceptualized the study design, contributed to the interpretation of the results, and critically revised the first draft of the manuscript. All authors critically revised the manuscript, gave final approval of the last version of the manuscript, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding

Open Access funding enabled and organized by Projekt DEAL. The ESCAschool study was funded by the German Federal Ministry of Education and Research (BMBF) Grant 01EE1408B. This work was supported by the research consortium on ADHD, ESCA-Life, funded by the German Federal Ministry of Education and Research (FKZ 01EE1408A-E). Bundesministerium für Bildung und Forschung, FZK 01EE1408B, Manfred Döpfner

Data availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics approval and consent to participate

Ethical approval was obtained for the study center Cologne from the University of Cologne (ID 15–216), for the study center Essen from the University of Duisburg- Essen (ID 17–7404-BO), for the study center Hamm from the Ruhr- University Bochum (ID 15–5564), for the study center Göttingen from the University Medical Center Göttingen (ID 3/3/17), for the study center Mainz from the Federal Medical Association of Rhineland-Palatinate (ID 837.237.17 [11071]), for the study center Mannheim from the Ruprecht- Karls-University Heidelberg (ID 2015–646 N-MA), for the study center Marburg from the Philipps-University

Marburg (ID Studie 03/16), for the study center Tübingen from the Eberhard-Karls-University Tübingen (ID 791/2015BO2), and for the study center Würzburg from the Julius- Maximilians-University Würzburg (ID 332/15_z).

Conflicts of interest

A-KTh, AGD and MD were involved in the development of the DISYPS-ILF and will receive royalties from the publisher Hogrefe after publication of this instrument. AGD and MD are AKiP supervisors and lecturers and received income as heads of the School for Child and Adolescent Behavior Therapy at the University of Cologne and royalties from treatment manuals, books, and psychological tests published by Guilford, Hogrefe, Enke, Beltz, and Huber. MD received consulting income and research support from Medice, Shire, and eyelevel. AGD received consulting income and research support from Medice and eyelevel. TB served in an advisory or consultancy role for ADHS digital, Infectopharm, Lundbeck, Medice, Neurim Pharmaceuticals, Oberberg GmbH, Roche, and Takeda. He received conference support or speaker's fee by Medice and Takeda. He received royalities from Hogrefe, Kohlhammer, CIP Medien, Oxford University Press. KB receives or has received research grants from the German Research Foundation (DFG), German Federal Ministry for Education and Research (BMBF), Philipps-University Marburg, Federal Joint Committee (G-BA), German Ministry for Health, University Hospital Giessen and Marburg and Rhön Klinikum AG. Further she receives royalties and honorary from Georg Thieme Publisher. DB served as an unpaid scientific advisor for an EU-funded neurofeedback trial unrelated to the present work. JG has received a research grant from the Bavarian State Ministry of Family, Labor and Social Affairs. She receives royalties from Hogrefe for the publication of an ADHD treatment manual. MAH served in an advisory role for Shire and Medice and received conference attendance support or was paid for public speaking by Medice, Shire. and Neuroconn. He receives research support from the German Research Foundation and the German Ministry of Education and Research. He receives royalties as Editor-in-Chief of the German Journal for Child and Adolescent Psychiatry and for textbooks from Hogrefe. MIH has served as a member of the advisory boards of Eli Lilly and Co., Engelhardt Arzneimittel, Janssen-Cilag, Medice, Novartis, Shire, and Steiner Arzneimittel within the past five years; served as a consultant to Engelhardt Arzneimittel, Medice, and Steiner Arzneimittel; received honoraria from Eli Lilly and Co., Engelhardt Arzneimittel, Janssen- Cilag, Medice, Novartis, and Shire; and received unrestricted grants for investigator-initiated trials from Eli Lilly and Co., Medice, Engelhardt Arzneimittel, and Steiner Arzneimittel. LP served in an advisory or consultancy role for Shire, Roche, and Infectopharm. She has received speaker's fees from Shire and royalties from Hogrefe, Kohlhammer, and Schattauer. HUS served in an advisory or consultancy role for Medice. He has received speaker's fees from Shire and Medice. CD, MJ, CD, CH, LJ, A-KTr, PV, EvW, UD, JH, SH, TJ, AK, JK, TL, SM, TR, MR, PS, JW, MZ are not aware of any biases that might be perceived as affecting the objectivity of this manuscript and declare that they have no competing interests.

Experiment participants

Patients and parents or guardians were informed verbally and received information sheets about the scope and the relevance of the study as well as the study procedures. All participants were able to clarify questions about the study with research staff members. For study participation, patients and their parents or guardians were required to give their assent and informed consent.

Table 1

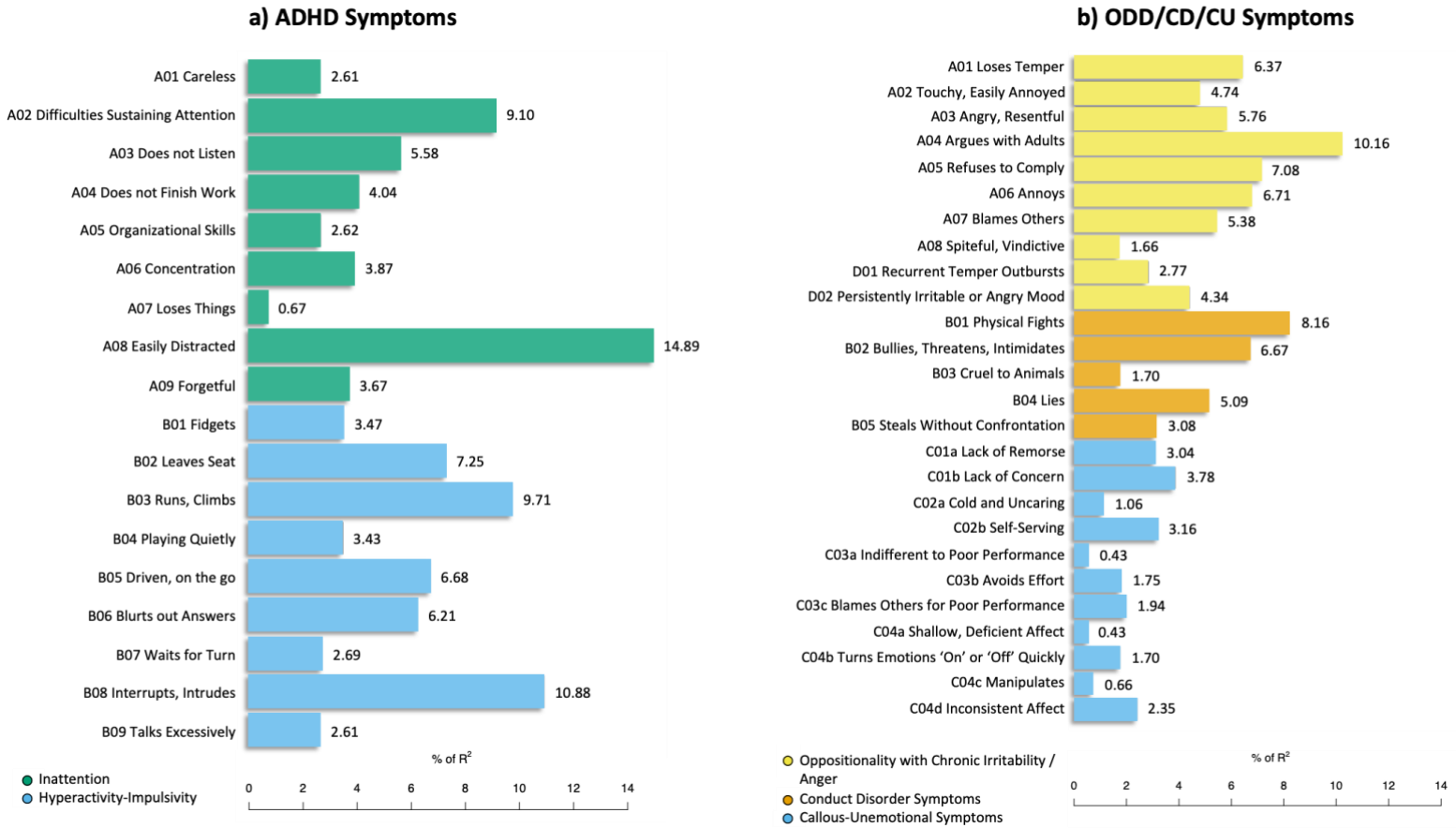
Sample characteristics

Variable	Total Sample (*N* = 474)
Age Mean (*SD*)	8.90 (1.49)
Male *n* (%)	382 (81)
Primary Diagnosis *n* (%)	
No ADHD diagnosis	32 (7)
ADHD – combined type	208 (44)
ADHD – predominantly inattentive type	184 (39)
ADHD – predominantly hyperactive-impulsive type	50 (11)
Comorbidities *n* (%)	*n* = 454 - 465
Internalizing disorders:	
- Anxiety	29 (6)
- Depression	15 (3)
Externalizing disorders:	
- Oppositional defiant disorder	166 (37)
- Conduct disorder	28 (6)
- Disruptive mood dysregulation disorder	40 (9)
Other disorders:	
- Obsessive-compulsive disorder	2 (< 1)
- Tic disorder	24 (5)
- Autism spectrum disorder	2 (< 1)
Medication *n* (%)	*n* = 462
ADHD medication	150 (33)
Parents' Primary Language *n* (%)	*n* = 458
German	429 (94)
Parents' Highest Educational Attainment *n* (%)	*n* = 455
Higher-track school	261 (57)
Vocational school	26 (6)
Medium-track school	123 (27)
Lower-track school	43 (10)

Note. Clinical diagnoses of ADHD and externalizing behavior disorders were based on the semi-structured *Clinical Parent Interview for Externalizing Disorders in Children and Adolescents* (ILF-EXTERNAL) conducted with the parents. Additional comorbid symptoms were evaluated using a clinical diagnostic checklist. ADHD = attention-deficit/hyperactivity disorder.

Figure 1

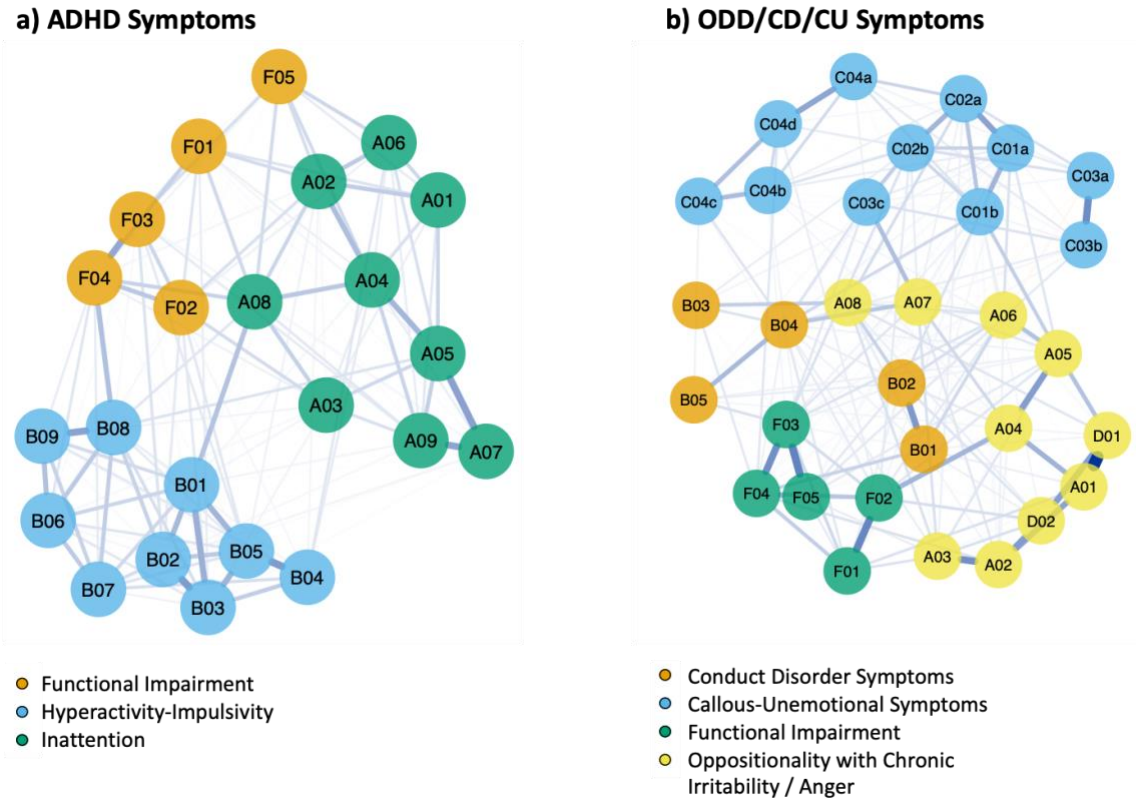
Relative importance estimates of ADHD and ODD/CD/CU symptoms on global functional impairment



Note. Depicted are the relative importance coefficients of ADHD (Figure 1a) and ODD/CD/CU (Figure 1b) symptoms on global functional impairment associated with these symptoms, respectively. Each value represents the unique shared variance (%) between a symptom and functional impairment. Estimates are adjusted to sum up to 100%.

Figure 2

Networks of ADHD and ODD/CD/CU symptoms with associated functional impairment domains



Note. The figure depicts the network structures of the ADHD and ODD/CD/CU symptoms and functional impairment related to psychological strain (F01) and in the domains of home life and family members (F02), relationships with adults (F03), relationships with children/adolescents and recreational activities (F04), and academic performance (F05). Item dimensions are differentiated by color. Blue edges represent positive partial correlations, and the thickness of an edge represents the strength of the partial correlation. A short description of each item is provided in Figure 1. ADHD = attention-deficit/hyperactivity disorder; ODD = oppositional defiant disorder; CD = conduct disorder; CU = callous-unemotional.

6 Discussion

Having developed the clinical parent interviews for diagnosing mental disorders based on the DSM-5 in children and adolescents (DISYPS-ILF; Görtz-Dorten et al., 2022), this doctoral dissertation presented a comprehensive psychometric evaluation of the clinical parent interview for diagnosing externalizing disorders of children and adolescents (ILF-EXTERNAL). This discussion section presents the main findings of each publication along with their overall theoretical and clinical implications, identifies limitations, and addresses challenges and future research directions.

6.1 Summary and clinical implications

Study 1 (Thöne et al., 2020) presented the reliability and validity of the ILF-EXTERNAL in a clinical sample of school-age children. Overall, results supported the convergent and divergent validity as indicated by moderate to high correlations ($.33 \leq r \leq .69$, $p < .001$) between the ILF-EXTERNAL scale scores with corresponding externalizing symptom scales and significantly lower correlations with divergent symptom scales from the CBCL/6-18R. These findings are largely in line with previous findings that have reported small to moderate correlations between the CBCL and clinical diagnoses obtained from other semi-structured interviews, such as the K-SADS (Birmaher et al., 2009; Brasil & Bordin, 2010; Y. L. Chen et al., 2017; Kim et al., 2004). Study 1 (Thöne et al., 2020) also addresses concerns related to clinical interviews that *"the dependence on respondents in these interviews is similar to the dependence on respondents completing a checklist on their own except for the potential error introduced by interviewer characteristics and interviewer–respondent exchanges"* (Boyle et al., 2017, p. 2). The present results, however, found moderate to substantial agreement between ratings from parents and clinicians ($.57 \leq r \leq .78$, $p < .001$). These results emphasize that both viewpoints are complementary and that both are vital for a comprehensive clinical assessment (cf. Döpfner & Petermann, 2012). One focus of Study 1 (Thöne et al., 2020) was to examine agreement between clinicians from a categorical and dimensional perspective. Regarding the DSM-5 categorical diagnoses, diagnostic agreement ranged from fair ($\kappa = .38$ ADHD hyperactive-impulsive type), to moderate ($\kappa = .55$ DMDD), substantial ($\kappa = .71$ ADHD combined type; $\kappa = .74$ ADHD inattentive type; $\kappa = .74$ any ADHD) to almost perfect agreement ($\kappa = .82$ ODD; $\kappa = .94$ CD; $\kappa = .82$ limited prosocial emotions). Regarding interrater reliability on a dimensional level, ICC coefficients ranged between 'very good' to 'excellent' [ICC(1,1) = .83−.95] for all scales of the ILF-EXTERNAL. As becomes clear, interrater reliability

estimates were generally higher at the dimensional level compared to categorical diagnoses. As outlined in Chapter 1.1, these findings are consistent with the meta-analyses by Markon and colleagues (2011) who demonstrated an expected 15% rise in reliability and a 37% rise in validity when dimensional instead of categorical measures of psychopathology were adopted. Accordingly, the degree to which an individual is below or above the threshold for a diagnosis is considered irrelevant to the diagnostic construct (Clark et al., 2017; Coghill & Sonuga-Barke, 2012; Lahey et al., 2022). Consider for example, the ODD diagnosis that consists of 8 symptoms, where the threshold is set at 4 symptoms. According to the categorical approach, values from 1-3 are converted to 'no diagnosis' and values from 4-8 are converted to 'diagnosis present'. This dichotomization leads to a substantial loss of information about fine-grained differences in symptom severity (Lahey et al., 2022; Markon et al., 2011). In general, this study contributes to empirical research demonstrating that mental disorders are not characterized by abrupt thresholds, but rather that mental disorders are better described as continuous phenomena (Haslam et al., 2012, 2020).

Study 2 (Thöne et al., 2021) evaluated the factorial structure of the ILF-EXTERNAL and systematically tested the underlying latent dimensions of externalizing symptoms through CFA and ESEM models. Overall, the *a priori* assumed DSM-5-based factorial structure of the ILF-EXTERNAL could be confirmed as demonstrated by the correlated factors models and hierarchical models. When comparing the solutions obtained from CFA and ESEM, the ESEM models generally outperformed the CFA models, since they exhibited significantly better model fit, greater interpretability, less biased factor loadings, and smaller correlations between the factors. These results are consistent with the growing body of evidence supporting the use of ESEM models, which provide a more realistic depiction of psychological symptoms (Arias et al., 2016; Gomez et al., 2021; Gomez & Stavropoulos, 2020; Rodenacker et al., 2017). Regarding goodness-of-fit indices, the traditional bifactor ESEM model showed the best model fit and successfully addressed issues observed in its CFA counterpart, such as negative variances or negative factor loadings. The presence of anomalous outcomes observed in traditional bifactor CFA models aligns with methodological concerns (Bonifay et al., 2017; Eid, 2020; Eid et al., 2017; Watts et al., 2019), and empirical studies that have also reported various anomalous findings when traditional bifactor models are applied to psychological symptoms (Arias et al., 2018; Burns et al., 2020a; Heinrich et al., 2020, 2021; Rodenacker et al., 2018; Thöne et al., 2022). However, one significant limitation of bifactor models is the inconsistent interpretation of the g-factor across studies, which subsequently affects the interpretation of the s-factors due to their definition based on the residual variance associated with the g-factor (Eid

et al., 2017; Reise, 2012). As argued by Watts and colleagues (2019, p. 1285): *"The general factor appeared to reflect a differentially weighted amalgam of psychopathology rather than a liability for psychopathology broadly construed".* This leads to significant problems in synthesizing research findings through replication studies, literature reviews or meta-analyses (Eid et al., 2017; Reise, 2012). In comparison, applying bifactor S-1 models with a general HI reference factor and four residual s-factors (IN, ODD-AD, CD, CU symptoms) resulted in a statistically sound factor structure and unambiguous and meaningful interpretation. These novel insights align well with other empirical studies which applied bifactor S-1 CFA models to symptoms of ADHD/ODD in clinical (Junghänel et al., 2020, 2022; Thöne et al., 2022) and community samples (Burns et al., 2020a; Thöne et al., 2022) of children and adolescents. Notably, Study 2 (Thöne et al., 2021) expands these findings, since it incorporated ESEM models and examining a broader scope of externalizing symptoms. Additionally, Study 2 (Thöne et al., 2021) supports the emerging notion that applying a bifactor S-1 model can address distinct research inquiries compared to traditional bifactor models, thereby offering additional value and enhancing our understanding of psychopathology (Burns et al., 2020a). The measurement invariance of the prevailing bifactor S-1 CFA/ESEM models was evaluated following the guidelines of Meredith (1993) and Vandenberg and Lance (2000) across different raters or informants (parents, clinicians, teachers) and assessment methods (interview, questionnaires). Simply put, configural invariance tests whether the basic model configuration (i.e., item-factor structure) remains consistent across groups. Metric invariance further assumes that item loadings are equivalent across groups. Scalar invariance goes a step further and assumes that item intercepts are equal across groups (Meredith, 1993; Vandenberg & Lance, 2000). The findings indicated that the prevailing models exhibited configural and metric invariance but did not demonstrate scalar invariance, implying that the severity of individual symptoms may vary across different contexts. These results hold significant importance since it has been suggested that cross-informant discrepancies reflect some kind of invalidity in terms of measurement error or rater bias (see Chapter 1.3). However, these findings contribute to the growing recognition that informant discrepancies often contain meaningful information (i.e., valid data), e.g., they shed light on how children's behavior may vary meaningfully across different settings, such as at school versus at home. (De Los Reyes et al., 2013, 2023). Moreover, Study 2 (Thöne et al., 2021) highlights that a more reasonable assumption would be that different informants will offer varying assessments of a child's symptoms and that multi-informant assessments hold significance as they encompass distinct viewpoints from each informant who contributes their report (De Los Reyes et al., 2013, 2023; Dirks et al., 2012;

Hunsley & Mash, 2007). In light of these findings, Study 2 (Thöne et al., 2021) raises doubts about the validity of traditional bifactor models as the 'winning models' for conceptualizing psychopathology. Additionally, Study 2 (Thöne et al., 2021) presents a set of model-based criteria (such as model fit indices, factor loadings, omega statistics, and model parsimony) that can assist researchers in evaluating structural models of psychopathology more comprehensively (see also Thöne et al., 2022).

Study 3 (Thöne et al., 2023) zoomed in on the symptom-level and focused on examining individual symptoms of ADHD and ODD/CD/CU and their differential associations with global FI and impairments in five specific domains. Multivariate regression analyses revealed significant variations in the associations between symptoms and global FI. The symptoms A08 *Easily Distracted* in ADHD (15%) and A04 *Argues with Adults* in ODD (10%) had the strongest contributions to the explained variance. These findings align with previous studies highlighting the diverse effects of individual symptoms on global FI, e.g., in the fields of ADHD (Mota & Schachar, 2000; Zoromski et al., 2015) or depression (Fried & Nesse, 2014). Network analyses further revealed that FI showed the strongest associations with ADHD inattention symptoms in domains such as academic performance, home life and family members, and psychological strain. On the other hand, hyperactivity-impulsivity symptoms were most strongly linked to FI in domains including relationships with adults as well as relationships with children and recreational activities. The results from Study 3 (Thöne et al., 2023) align with previous studies emphasizing the impact of ADHD inattention symptoms on academic impairment (Garner et al., 2013; Massetti et al., 2008; Willcutt et al., 2012; Zoromski et al., 2015) and ADHD hyperactivity-impulsivity symptoms on impaired social functioning (Garner et al., 2013; Goh, Martel, et al., 2021; Willcutt et al., 2012). By comparison, symptoms from the ODD and CD dimensions were closely related to FI domains, while CU traits showed weaker associations. This finding is particularly intriguing considering the inclusion of CU traits a specifier to the CD diagnosis in the DSM-5 (American Psychiatric Association, 2013) and as a specifier to the diagnoses of ODD or CD in the upcoming ICD-11 (World Health Organization, 2019). Study 3 (Thöne et al., 2023) offers valuable clinical implications for clinical assessment and therapy. First, it supports the growing consensus that assessing FI is as crucial as assessing symptoms for a comprehensive understanding of psychopathology (Arildskov et al., 2022; DuPaul, 2022). Arildskov et al. (2022) conducted a recent study to explore the associations between the severity of ADHD symptoms and FI, specifically investigating if there is a nonlinear association or a symptom severity threshold that significantly increases impairment. The authors observed a gradual linear rise in impairment proportionate to the increase in symptom severity, challenging

the current arbitrary definition of the symptom severity threshold for ADHD diagnosis based on the presence or absence of impairment (Arildskov et al., 2022). The authors therefore emphasize *"the continuing need for the clinical assessment and diagnosis of ADHD to be based on two independent decisions: one about the symptom threshold and one about functional impairment with both decisions having a certain degree of arbitrariness and social subjectivity"* (Arildskov et al., 2022, p. 5). Second, Study 3 (Thöne et al., 2023) emphasizes the importance of a nuanced perspective on the relationship between specific symptoms and FI. For instance, the ADHD symptom A08 *Easily Distracted* had a much greater impact on global FI compared to A01 *Careless*, despite both symptoms originating from the inattention dimension. These results indicate that symptoms have varying effects on FI, and giving differential weight to symptoms can significantly enhance the prediction of FI. In the future, studying the links between specific symptoms and FI across various domains could provide insights for potential revisions to the diagnostic criteria. Third, the findings of Study 3 (Thöne et al., 2023) hold significance for the advancement of assessment tools, particularly in the context of screening instruments. Building on the suggestions of Zoromski et al. (2015), network analyses examining the relationships between symptoms and impairments could guide the development of more cost-effective screening tools, which could initially focus on symptoms that exhibit strong associations with specific FI domains. These screening measures have the potential to benefit both clinical research and practice by providing efficient ways to identify individuals at risk and in need of further assessment (Zoromski et al., 2015).

6.2 Limitations

Some methodological and conceptual aspects that are recurrent throughout this doctoral dissertation warrant attention. First, it is important to note that the analyses in this dissertation are based on the ESCAschool sample (Döpfner et al., 2017), which predominantly consists of children diagnosed with ADHD (94%). Consequently, the sample used in this dissertation does not encompass the complete spectrum of externalizing disorders but rather consists of distinct ADHD subtypes exhibiting different levels of severity and related externalizing behavior symptoms (Thöne et al., 2021). Simultaneously, the ESCAschool sample provides an opportunity for diagnostic assessment: While field samples often consist of children who are symptom-free, the ESCAschool sample necessitates the differentiation between clinically significant and non-significant symptoms to ascertain whether an individual meets the criteria outlined in the ICD/DSM diagnostic systems. Furthermore, the ESCAschool sample offers an additional advantage as it primarily consists of patients with clinically significant symptoms.

Consequently, the analyses predominantly focus on describing psychometric criteria within the clinical spectrum of these instruments, which holds particular relevance for clinical studies. Regarding the sample diversity, all children included in this study were enrolled in the randomized control trial ESCAschool. As these children met stringent criteria for inclusion and exclusion, it would be valuable to replicate the findings of this study with clinically referred children in routine care settings to ensure the generalizability of the results (Bear et al., 2020; Weisz et al., 2005).

Second, it is important to emphasize that the analyses conducted in this dissertation utilized baseline data from the ESCAschool trial, which means that a cross-sectional research design was employed. Consequently, the cross-sectional nature of the data limits our ability to infer causal relationships, such as the causal interplay between FI and symptoms. It is plausible that symptoms and FI may reciprocally influence each other (Thöne et al., 2023). Furthermore, considering that the data utilized in this dissertation are confined to children of school age, future investigations should explore how the associations between FI and symptoms evolve across different age groups, including the examination of whether specific symptoms become more impairing during adolescence (e.g., Martel et al., 2021).

Third, it is important to clarify that the data analyzed in this study comprise behavioral data obtained through clinical interviews and questionnaires. As discussed in Study 2 (Thöne et al., 2021), it is worth noting that latent factor models solely based on behavioral symptoms have limitations in providing comprehensive insights into the etiology of mental disorders, considering the intricate nature of neurobiological vulnerabilities and neural circuits (Beauchaine & Hinshaw, 2020). However, achieving alignment with projections derived from etiological theories marks a notable progression in the formulation of a dimensional framework for mental disorders. To further advance our understanding, future research should incorporate multiple units of analysis, as exemplified by ongoing efforts such as the *Research Domain Criteria* project (RDoC; Beauchaine & Hinshaw, 2020; Insel et al., 2010). Such investigations would yield fresh perspectives on the hierarchical structure of mental disorders and have the potential to inform the development of more targeted psychotherapeutic interventions.

Lastly, one important limitation of the present work is that the ILF-EXTERNAL was conducted with the parents only. In other words, although there are two different rater perspectives (clinicians, parents), the assessment of the child's symptomatology is based on the same informants (parents). Accordingly, parents may only offer limited insights into their children’s covert internal states. This issue is of particular importance in studies dealing with children’s internalizing symptoms, in comparison to the present work, which is mainly

concerned with well-observed, externalizing behavior problems. In the present work, the children's self-reports were deliberately not included because the extent to which children can make reliable and valid statements about their inner mental states depends not only on the age of the children but also on various characteristics of the respective questionnaire, such as the response format, complexity of item formulation, or the degree of abstractness of items (reviewed by Taber, 2010). Consequently, it is of great clinical relevance that studies investigating psychological problems in adolescence include, whenever possible, self-reports in addition to parent-reports. While clinical interviews are widely regarded as the most reliable method for diagnosing mental disorders, it is crucial to acknowledge these consideration when incorporating reports from multiple informants as an essential element of best practices in evidence-based assessment (cf. Chapter 1.3; De Los Reyes, 2011; De Los Reyes et al., 2013, 2023; Dirks et al., 2012; Hunsley & Mash, 2007; Thöne et al., 2022).

6.3 Challenges and future directions

Addressing the categorical and dimensional aspects of mental disorders through latent variable and network approaches poses several challenges, which point towards directions for future research.

Latent factor and network approaches of psychopathology have often been portrayed as some kind of competitive field, with researchers from one camp challenging the views of the other (Borsboom et al., 2017, 2018, 2021, 2022; Cramer et al., 2010, 2012; Forbes et al., 2017a, 2017b, 2019; Fried, 2020; Neal et al., 2022; Steinley et al., 2017). With respect to the challenges facing the latent factor approach, several authors have expressed significant concerns about the application of latent factor models and, more specifically, traditional bifactor models. These concerns encompass various issues related to traditional bifactor models used in empirical studies (for a thorough discussion, see Study 2; Thöne et al., 2021). First, there is concern that bifactor models may exhibit superior fit due to overfitting, potentially compromising their validity (Bonifay et al., 2017). Moreover, the interpretation of the g-factor in bifactor models may not be directly comparable across different studies due to its strong dependence on specific item selections (Eid et al., 2017; Reise, 2012). Furthermore, traditional bifactor models can also yield problematic parameter estimates, such as negative variances or factor loadings, which hinder meaningful interpretation (Eid et al., 2017; Heinrich et al., 2020). Additionally, the commonly overlooked omega factor reliability coefficients reveal that s-factors in bifactor models may be weakly defined or even vanish empirically (Arias et al., 2018; Rodriguez et al., 2016). Although the bifactor S-1 model has been proposed to address these concerns, it faces

several challenges that still need to be addressed (J. D. Burke & Johnston, 2020; Willoughby, 2020). These challenges include the potential for overfitting similar to traditional bifactor models, the arbitrary selection of reference facets, interpretation of the g-factor, and the added value of the bifactor S-1 model. Additionally, challenges arise when applying a bifactor S-1 model to item-level data and when using bifactor models to identify a general psychopathology factor (J. D. Burke & Johnston, 2020; Burns et al., 2020a; Willoughby, 2020). As Burns and colleagues (2020b) discussed in detail, future research on latent factor analysis must address these challenges before bifactor S-1 models eventually find their way into psychopathology research. This includes a focus on testing the advantages of the bifactor S-1 model, particularly in terms of its practical relevance for individual clinical assessment (Eid, 2020) as well as establishing external and internal validity, which *"are critical toward explaining rather than describing psychopathology"* (Burns et al., 2020b, p. 921).

One of the fundamental challenges within the field of psychopathological network research concerns the reliability and replicability of network results (for a review, see Neal et al., 2022). Critics of the network approach expressed concerns that *"psychopathology networks have limited replicability"* (Forbes et al., 2017b, p. 1011) and that *"popular network analysis methods produce unreliable results"* (Forbes et al., 2017b, p. 1011). Forbes and colleagues (2017a) provided explanations for the observed unreliability of network parameters, including violations of statistical and theoretical assumptions. Steinley and colleagues went even further and claimed that *"the problem is likely worse (or at least more complicated) than they initially indicated"* (p. 1000). However, Borsboom and colleagues (2017) re-analyzed the same data using advanced network analyses techniques (e.g., the Network Comparison Test), found that *"network models replicate very well"* (p. 997), and attributed their contradictory results to *"statistical inaccuracies"* (p. 998) and *"debatable metrics to assess replicability"* (p. 989). In an effort to provide best practices for future research, psychopathology network advocates have provided a roadmap for network replication studies (Borsboom et al., 2017) and tutorial articles on network analysis (Burger et al., 2022; Epskamp et al., 2018; Epskamp & Fried, 2018) that can help future researchers obtain more objective and trustworthy results. This debate was revisited a few years later in the journals *World Psychiatry* (Borsboom et al., 2018; Forbes et al., 2019), *Multivariate Behavioral Research* (Forbes, Wright, et al., 2021b, 2021a; Fried et al., 2021; Jones et al., 2021; Steinley, 2021; Waldorp & Marsman, 2022) and *Nature Reviews Method Primers* (Borsboom et al., 2021, 2022; Neal et al., 2022). Following Borsboom et al.'s (2018) reaffirmation that *"the general picture which emerges is that network structures replicate and generalize well"* (p. 143), critics of the network approach countered that this

debate was far from over, since Borsboom et al. (2021) *"omitted or glossed over core critiques of [network] models and methods"* (Neal et al., 2022, p. 1) and provided directions for future research, including (a) additional value of network models when alternative multivariate methods already exist, (b) the degree to which inferences about causality or within-processes can be drawn when using cross-sectional data, (c) best practices for network estimation reliability, e.g., through simulation studies, and (d) establishing clear guidelines when interpreting network metrics (Neal et al., 2022). Even though it may seem as if two completely opposite perspectives are meeting here, they nevertheless pursue a similar goal, namely, the quest for scientific knowledge. Therefore, it is imperative for future research to advance our understanding of psychopathological networks and to recognize the obstacles that lie ahead.

While both approaches lead to a different understanding of the psychological construct of interest, recent studies have demonstrated the mathematical equivalence between latent factor models and network models (e.g., van Bork et al., 2021) and therefore can complement – rather than exclude – each other regarding the research question being addressed. For example, a network model with positive associations between variables can result in a well-fitting latent factor model and, conversely, a unidimensional model with positive factor loadings is consistent with a fully connected network model with positive edge weights (van Bork et al., 2021; Van Der Maas et al., 2006). In a comparative analysis between factor analysis and network analysis, Christensen and Golino (2021) discovered that the cumulative connections of a node (referred to as strength centrality) were approximately comparable to CFA factor loadings. Consequently, they deduced that *"network loadings can provide similar information to factor loadings when the data are generated from a factor model and therefore can be used in a similar way"* (Christensen & Golino, 2021, p. 1563). Ironically, this leads to the consequence that if psychopathological networks have poor replicability, then factor structures would also have poor replicability (Borsboom et al., 2017). In broader terms, the challenges of achieving replicability and generalizability of findings have become an increasing issue not only in psychology (e.g., Pashler & Wagenmakers, 2012) but also across the scientific domain as a whole (e.g., Baker, 2016).

As becomes clear, these are exciting times for classification research, offering several directions for future research (see also Carragher et al., 2015; Clark et al., 2017; Eaton, 2015). First, the emergence of advanced latent variable models provides novel avenues for evaluating the categorical-dimensional nature of psychopathology. For example, factor mixture modeling allows to combine categorical and dimensional components within a hybrid framework (B. Muthén, 2006). In brief, hybrid mixture models are both categorical (i.e., they assign

individuals to liability classes) and dimensional (i.e., they model differences in severity between classes through continuous latent variables), thereby facilitating meaningful distinctions between homogeneous groups while accounting for different levels of severity (B. Muthén, 2006). Besides, hybrid modeling with longitudinal data seems to be particularly powerful in elucidating different paths of problematic development (Carragher et al., 2015; B. Muthén, 2006). Second, recent integrations of latent variable and network modeling into one framework, namely, latent variable network modeling (Epskamp et al., 2017), have opened up exciting directions for future research. Latent variable modeling offers to compensate for the shortcomings of the other approach, respectively (Epskamp et al., 2017). First attempts to apply this new statistical method to psychological data have been made in the fields of depression (Belvederi Murri et al., 2022), posttraumatic stress disorder (G. Li et al., 2020), and eating disorder (Forbush et al., 2022), with overall versatile and promising results that complement the spectrum of psychometric models.

Although it remains to be seen whether dimensional classification systems such as the HiTOP will actually replace our existing categorical systems as they intend to (DeYoung et al., 2022, but see Haeffel, Jeronimus, Fisher, et al., 2022; Haeffel, Jeronimus, Kaiser, et al., 2022), dimensional models are increasingly finding their way into the conceptualization of psychopathology. One prominent example comprises conceptualization of personality disorders in the ICD-11 (Hopwood et al., 2018; World Health Organization, 2019). Following a categorical-dimensional hybrid approach, clinicians evaluate individuals on a dimensional scale and make categorical judgments by applying diagnostic thresholds (World Health Organization, 2019). In addition, the subthreshold ‘personality difficulty’ and three levels of severity of personality disorder can be amplified by six trait specifiers (World Health Organization, 2019). There is hope that, after some time of adjustment, clinicians may prefer the dimensional view over categorical thinking (Clark et al., 2017; Lahey et al., 2022). Nevertheless, such proposals are unlikely to gain acceptance and find broad implementation in clinical practice unless they offer clinically relevant information during patient interactions that justifies the investment of time and effort needed to adopt the new systems (Clark et al., 2017; Zimmerman, 2021). Beneficial directions for future research therefore include (a) identifying clinically relevant dimensions and appropriate cut-off thresholds for aspects requiring categorical decisions or treatment decisions (Clark et al., 2017), and (b) identifying useful dimensions for different aspects of care, e.g., types of treatment or interventions and their respective durations (Clark et al., 2017). Recognizing that such a shift will be challenging at best, Lahey et al. (2021) emphasize the importance of demonstrating that dimensional

assessments yield superior patient outcomes when compared to established categorical systems. For this purpose, Zimmerman (2021) proposes conducting randomized controlled treatment trials in which patients will be randomized and evaluated under either categorical or dimensional approaches. At this point, only the future can tell whether the treatment of most patients would remain the same or actually improve, depending on which diagnostic approach is chosen (Zimmerman, 2021). As Zimmerman (2021) states, *"(…) before I put forth the time and effort to learn and use HiTOP, or a HiTOP-like system, I will need to see data demonstrating that this will improve the care I provide to my patients. Specifically, I would need to see studies showing that more of my patients are likely to get better"* (p. 71). Overall, the assessment of mental disorders is moving toward more dimensional conceptualizations of psychopathology, but clinicians and researchers must be on board for this important shift to finally take hold.

6.4 Conclusion

This doctoral dissertation provided a comprehensive psychometric evaluation of the clinical parent interview for diagnosing externalizing disorders in children and adolescents (ILF-EXTERNAL) from the DISYPS-ILF (Görtz-Dorten et al., 2022) and, more generally, revisited contemporary aspects to the debate between categorical and dimensional concepts of diagnostic classification. The utilization of contemporary analytic methods, particularly the comparison of latent factor and network models, has revealed that resolving this debate empirically does not have a straightforward or all-encompassing solution. Each categorical and dimensional approach offers its own advantages, and integrating both approaches could potentially mitigate their respective limitations. The findings presented in this dissertation can serve as a roadmap for future investigations and contribute to establishing a robust diagnostic framework for mental disorders in children and adolescents, characterized by high-quality standards. Specifically, this dissertation has shown that (a) diagnostic agreement between clinicians was generally higher at the dimensional level compared to categorical diagnoses, (b) discrepancies among informants frequently hold valuable insights, and the significance of employing multiple informants in assessments lies in capturing the distinct viewpoints offered by each informant providing their respective reports, and (c) the network approach provides a more detailed, though intricate, understanding of the connections between individual symptoms and impairments. These findings, whilst intriguing, represent only the beginning of the journey and there is still a need to investigate categorical and dimensional concepts as well as recently proposed hybrid systems of psychopathology in greater detail. Moving forward, it is crucial to

place additional emphasis on the practicality of empirically derived dimensions, examining their clinical significance in terms of treatment selection and adaptive interventions. Furthermore, systematic investigation into patient outcomes when employing dimensional measures of psychopathology, as opposed to categorical measures, is warranted.

The beginning of this doctoral dissertation featured two quotations, each symbolizing a core message of this work (cf. Clark et al., 2017). Edmund Burke's quote serves as a reminder that even though day and night lack distinct boundaries, we are still able to differentiate and classify them. Thomas S. Eliot's quote suggests that as our understanding of mental disorders deepens, we don't encounter completely unrecognizable insights, but rather a fresh and lucid perspective on what we previously grasped vaguely. After all, clinicians and scientists need to collaborate, rather than work against each other as they often have in the past, to address and further advance the debate between categorical and dimensional concepts of mental disorders. That day may seem a long way off, but it is the goal to strive for.

6.5 References

Achenbach, T. M. (1966). The classification of children's psychiatric symptoms: A factor-analytic study. *Psychol. Monogr., 80*(7), 1–37.

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4–18 and 1991 profile.* University of Vermont, Department of Psychiatry.

Achenbach, T. M. (2020). Bottom-up and top-down paradigms for psychopathology: A half-century odyssey. *Annu. Rev. Clin. Psychol., 16*, 1–24. https://doi.org/10.1146/annurev-clinpsy-071119-115831

Achenbach, T. M. (2021). Hierarchical dimensional models of psychopathology: Yes, but…. *World Psychiatry*, *20*(1), 64–65. https://doi.org/10.1002/wps.20810

Achenbach, T. M., Becker, A., Döpfner, M., Heiervang, E., Roessner, V., Steinhausen, H. C., & Rothenberger, A. (2008). Multicultural assessment of child and adolescent psychopathology with ASEBA and SDQ instruments: Research findings, applications, and future directions. *J. Child Psychol. Psychiatry Allied Discip., 49*(3), 251–275. https://doi.org/10.1111/j.1469-7610.2007.01867.x

Achenbach, T. M., Ivanova, M. Y., Rescorla, L. A., Turner, L. V., & Althoff, R. R. (2016). Internalizing/externalizing problems: Review and recommendations for clinical and research applications. *J. Am. Acad. Child Adolesc. Psychiatry, 55*(8), 647–656. https://doi.org/10.1016/j.jaac.2016.05.012

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychol. Bull.*, *101*(2), 213–232. https://doi.org/10.1037/0033-2909.101.2.213

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms and profiles*. ASEBA.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles: An integrated system of multi-informant assessment.* ASEBA.

Achenbach, T. M., & Rescorla, L. A. (2003). *Manual for the ASEBA Adult Forms & Profiles.* ASEBA.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Association. https://doi.org/https://doi.org/10.1176/appi.books.9780890425596

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., t). American Psychiatric Association.

Andersson, A., Tuvblad, C., Chen, Q., Du Rietz, E., Cortese, S., Kuja-Halkola, R., & Larsson, H. (2020). Research Review: The strength of the genetic overlap between ADHD and other psychiatric symptoms – a systematic review and meta-analysis. *J. Child Psychol. Psychiatry*, *61*(11), 1173–1183. https://doi.org/10.1111/jcpp.13233

Angold, A., & Costello, E. J. (1995). A test-retest reliability study of child reported psychiatric symptoms and diagnoses using the Child and Adolescent Psychiatric Assessment (CAPA-C). *Psychol. Med.*, *25*(4), 755–762. https://doi.org/10.1017/S0033291700034991

Angold, A., & Costello, E. J. (2000). The Child and Adolescent Psychiatric Assessment (CAPA). *J. Am. Acad. Child Adolesc. Psychiatry*, *39*(1), 39–48. https://doi.org/10.1097/00004583-200001000-00015

Angold, A., Costello, E. J., & Erkanli, A. (1999). Comorbidity. *J. Child Psychol. Psychiatry Allied Discip.*, *40*(1), 57–87.

Arbeitsgruppe Deutsche Child Behavior Checklist. (1998). *Elternfragebogen über das Verhalten von Kindern und Jugendlichen; deutsche Bearbeitung der Child Behavior Checklist (CBCL/4-18). Einführung und Anleitung zur Handauswertung. 2. Auflage mit deutschen Normen, bearbeitet von Döpfner M, Plück J, Bölte S, Lenz*. Arbeitsgruppe

Kinder-, Jugend- und Familiendiagnostik (KJFD).
Arias, V. B., Ponce, F. P., Martínez-Molina, A., Arias, B., & Núñez, D. (2016). General and specific attention-deficit/hyperactivity disorder factors of children 4 to 6 years of age: An exploratory structural equation modeling approach to assessing symptom multidimensionality. *J. Abnorm. Psychol*, *125*(1), 125–137. https://doi.org/10.1037/abn0000115
Arias, V. B., Ponce, F. P., & Núñez, D. E. (2018). Bifactor models of attention-deficit/hyperactivity disorder (ADHD): An evaluation of three necessary but underused psychometric indexes. *Assessment*, *25*(7), 885–897. https://doi.org/10.1177/1073191116679260
Arildskov, T. W., Sonuga-Barke, E. J. S., Thomsen, P. H., Virring, A., & Østergaard, S. D. (2022). How much impairment is required for ADHD? No evidence of a discrete threshold. *J. Child Psychol. Psychiatry Allied Discip.*, *63*(2), 229–237. https://doi.org/10.1111/jcpp.13440
Association of Scientific Medical Societies in Germany AWMF. (2018). *Interdisciplinary Evidence- and Consensus-based (S3) Guideline "Attention Deficit/Hyperactivity Disorder (ADHD) in Children, Young People and Adults" [in German]*. 1–198. https://www.awmf.org/uploads/tx_szleitlinien/028-045l_S3_ADHS_2018-06.pdf
Baker, M. (2016). Is there a reproducibility crisis? *Nature*, *533*(26), 452–454. https://doi.org/10.1038/d41586-019-00067-3
Balázs, J., & Keresztény, Á. (2014). Subthreshold attention deficit hyperactivity in children and adolescents: A systematic review. *Eur. Child Adolesc. Psychiatry*, *23*(6), 393–408. https://doi.org/10.1007/s00787-013-0514-7
Balázs, J., Miklõsi, M., Keresztény, Á., Hoven, C. W., Carli, V., Wasserman, C., Apter, A., Bobes, J., Brunner, R., Cosman, D., Cotter, P., Haring, C., Iosue, M., Kaess, M., Kahn, J. P., Keeley, H., Marusic, D., Postuvan, V., Resch, F., … Wasserman, D. (2013). Adolescent subthreshold-depression and anxiety: Psychopathology, functional impairment and increased suicide risk. *J. Child Psychol. Psychiatry Allied Discip.*, *54*(6), 670–677. https://doi.org/10.1111/jcpp.12016
Bansal, P. S., Goh, P. K., Eng, A. G., Elkins, A. R., Thaxton, M., Smith, T. E., & Martel, M. M. (2021). Identifying the inter-domain relations among ODD, CD, and CU traits in preschool children using network analysis. *Res. Child Adolesc. Psychopathol.*, *49*(10), 1289–1301. https://doi.org/10.1007/s10802-021-00836-7
Bansal, P. S., Goh, P. K., Lee, C. A., & Martel, M. M. (2020). Conceptualizing callous-unemotional traits in preschool through confirmatory factor and network analysis. *J. Abnorm. Child Psychol.*, *48*(4), 539–550. https://doi.org/10.1007/s10802-019-00611-9
Bear, H. A., Edbrooke-Childs, J., Norton, S., Krause, K. R., & Wolpert, M. (2020). Systematic review and meta-analysis: Outcomes of routine specialist mental health care for young people with depression and/or anxiety. *J. Am. Acad. Child Adolesc. Psychiatry*, *59*(7), 810–841. https://doi.org/10.1016/j.jaac.2019.12.002
Beauchaine, T. P., & Hinshaw, S. P. (2020). The Research Domain Criteria and psychopathology among youth: Misplaced assumptions and an agenda for future research. *J. Clin. Child Adolesc. Psychol.*, *49*(3), 322–340. https://doi.org/10.1080/15374416.2020.1750022
Beauchaine, T. P., & McNulty, T. (2013). Comorbidities and continuities as ontogenic processes: Toward a developmental spectrum model of externalizing psychopathology. *Dev. Psychopathol.*, *25*(4pt2), 1505–1528. https://doi.org/10.1017/S0954579413000746
Beauchaine, T. P., Zisner, A. R., & Sauder, C. L. (2017). Trait impulsivity and the externalizing spectrum. *Annu. Rev. Clin. Psychol.*, *13*(1), 343–368. https://doi.org/10.1146/annurev-clinpsy-021815-093253
Becker, K., Banaschewski, T., Brandeis, D., Dose, C., Hautmann, C., Holtmann, M., Jans, T.,

Jendreizik, L., Jenkner, C., John, K., Ketter, J., Millenet, S., Pauli-pott, U., Renner, T., Romanos, M., Treier, A., Wirth, E. Von, Wermter, A., & Döpfner, M. (2020). Individualised stepwise adaptive treatment for 3 – 6-year-old preschool children impaired by attention-deficit / hyperactivity disorder (ESCApreschool): Study protocol of an adaptive intervention study including two randomised controlled trials within th. *Trials*, *21*, 1–19. https://doi.org/https://doi.org/10.1186/s13063-019-3872-8

Belvederi Murri, M., Grassi, L., Caruso, R., Nanni, M. G., Zerbinati, L., Andreas, S., Ausín, B., Canuto, A., Härter, M., Lopez, M. M., Weber, K., Wittchen, H. U., Volkert, J., & Alexopoulos, G. S. (2022). Depressive symptom complexes of community-dwelling older adults: A latent network model. *Mol. Psychiatry*, *27*(2), 1075–1082. https://doi.org/10.1038/s41380-021-01310-y

Bennett, D. A. (2001). How can I deal with missing data in my study? *Aust. N. Z. J. Public Health*, *25*(5), 464–469.

Birmaher, B., Ehmann, M., Axelson, D. A., Goldstein, B. I., Monk, K., Kalas, C., Kupfer, D., Gill, M. K., Leibenluft, E., Bridge, J., Guyer, A., Egger, H. L., & Brent, D. A. (2009). Schedule for affective disorders and schizophrenia for school-age children (K-SADS-PL) for the assessment of preschool children - A preliminary psychometric study. *J. Psychiatr. Res.*, *43*(7), 680–686. https://doi.org/10.1016/j.jpsychires.2008.10.003

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, *327*(8476), 307–310.

Bonifay, W., Lane, S. P., & Reise, S. P. (2017). Three concerns with applying a bifactor model as a structure of psychopathology. *Clin. Psychol. Sci.*, *5*(1), 184–186. https://doi.org/10.1177/2167702616657069

Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *J. Clin. Psychol.*, *64*(9), 1089–1108. https://doi.org/10.1002/jclp.20503

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, *16*(1), 5–13. https://doi.org/10.1002/wps.20375

Borsboom, D., & Cramer, A. O. J. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annu. Rev. Clin. Psychol.*, *9*, 91–121. https://doi.org/10.1146/annurev-clinpsy-050212-185608

Borsboom, D., Cramer, A. O. J., Schmittmann, V. D., Epskamp, S., & Waldorp, L. J. (2011). The small world of psychopathology. *PLoS ONE*, *6*(11). https://doi.org/10.1371/journal.pone.0027407

Borsboom, D., Deserno, M. K., Rhemtulla, M., Epskamp, S., Fried, E. I., McNally, R. J., Robinaugh, D. J., Perugini, M., Dalege, J., Costantini, G., Isvoranu, A. M., Wysocki, A. C., van Borkulo, C. D., van Bork, R., & Waldorp, L. J. (2021). Network analysis of multivariate data in psychological science. *Nat. Rev. Methods Primers*, *1*(58), 1–18. https://doi.org/10.1038/s43586-021-00055-w

Borsboom, D., Deserno, M. K., Rhemtulla, M., Epskamp, S., Fried, E. I., McNally, R. J., Robinaugh, D. J., Perugini, M., Dalege, J., Costantini, G., Isvoranu, A. M., Wysocki, A. C., van Borkulo, C. D., van Bork, R., & Waldorp, L. J. (2022). Reply to 'Critiques of network analysis of multivariate data in psychological science.' *Nat. Rev. Methods Primers*, *2*(1), 1–2. https://doi.org/10.1038/s43586-022-00178-8

Borsboom, D., Fried, E. I., Epskamp, S., Waldorp, L. J., van Borkulo, C. D., van der Maas, H. L. J., & Cramer, A. O. J. (2017). False alarm? A comprehensive reanalysis of "Evidence that psychopathology symptom networks have limited replicability" by Forbes, Wright, Markon, and Krueger (2017). *J. Abnorm. Psychol.*, *126*(7), 989–999. https://doi.org/10.1037/abn0000306

Borsboom, D., Robinaugh, D. J., Rhemtulla, M., & Cramer, A. O. J. (2018). Robustness and replicability of psychopathology networks. *World Psychiatry*, *17*(2), 143–144. https://doi.org/10.1002/wps.20515

Boyd, J. H., Burke, J. D., Gruenberg, E., Holzer, C. E., Rae, D. S., George, L. K., Karno, M., Stoltzman, R., McEvoy, L., & Nestadt, G. (1984). Exclusion criteria of DSM-III: A study of co-occurrence of hierarchy-free syndromes. *Arch. Gen. Psychiatry*, *41*(10), 983–989. https://doi.org/10.1001/archpsyc.1984.01790210065008

Boyle, M. H., Duncan, L., Georgiades, K., Bennett, K., Gonzalez, A., Van Lieshout, R. J., Szatmari, P., MacMillan, H. L., Kata, A., Ferro, M. A., Lipman, E. L., & Janus, M. (2017). Classifying child and adolescent psychiatric disorder by problem checklists and standardized interviews. *Int. J. Methods Psychiatr. Res.*, *26*(4), 1–9. https://doi.org/10.1002/mpr.1544

Brasil, H. H. A., & Bordin, I. A. (2010). Convergent validity of K-SADS-PL by comparison with CBCL in a Portuguese speaking outpatient population. *BMC Psychiatry*, *10*(83), 1–11. https://doi.org/10.1186/1471-244X-10-83

Brown, T. A., & Barlow, D. H. (2005). Dimensional versus categorical classification of mental disorders in the fifth edition of the diagnostic and statistical manual of mental disorders and beyond: Comment on the special section. *J. Abnorm. Psychol.*, *114*(4), 551–556. https://doi.org/10.1037/0021-843X.114.4.551

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *J. Pers.*, *80*(4), 796–846. https://doi.org/10.1111/j.1467-6494.2011.00749.x

Burger, J., Isvoranu, A.-M., Lunansky, G., Haslbeck, J. M. B., Epskamp, S., Hoekstra, R. H. A., Fried, E. I., Borsboom, D., & Blanken, T. F. (2022). Reporting standards for psychological network analyses in cross-sectional data. *Psychol. Methods*. https://doi.org/10.31234/osf.io/4y9nz

Burke, E. (1963). Thoughts on the cause of present discontents (1770). In P. J. Stanlis (Ed.), *Edmund Burke - Selected writings and speeches* (1st ed., pp. 121–175). Routledge. https://doi.org/10.4324/9781351312325

Burke, J. D., & Johnston, O. G. (2020). The bifactor S-1 Model: A psychometrically sounder alternative to test the structure of ADHD and ODD? *J. Abnorm. Child Psychol.*, *48*, 911–915. https://doi.org/10.1007/s10802-020-00645-4

Burke, J. D., Rowe, R., & Boylan, K. (2014). Functional outcomes of child and adolescent ODD symptoms in young adult men. *J Child Psychol Psychiatry*, *55*(3), 264–272. https://doi.org/10.1111/jcpp.12150

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociol. Methods Res*, *33*(2), 261–304. https://doi.org/10.1177/0049124104268644

Burns, G. L., De Moura, M. A., Beauchaine, T. P., & McBurnett, K. (2014). Bifactor latent structure of ADHD/ODD symptoms: Predictions of dual-pathway/trait-impulsivity etiological models of ADHD. *J. Child Psychol. Psychiatry*, *55*(4), 393–401. https://doi.org/10.1111/jcpp.12165

Burns, G. L., Geiser, C., Servera, M., Becker, S. P., & Beauchaine, T. P. (2020a). Application of the bifactor S – 1 model to multisource ratings of ADHD/ODD Symptoms: An appropriate bifactor model for symptom ratings. *J. Abnorm. Child Psychol.*, *48*(7), 881–894. https://doi.org/10.1007/s10802-019-00608-4

Burns, G. L., Geiser, C., Servera, M., Becker, S. P., & Beauchaine, T. P. (2020b). Promises and pitfalls of latent variable approaches to understanding promises and pitfalls of latent variable approaches to understanding psychopathology: Reply to Burke and Johnston , Eid , Junghänel and Colleagues, and Willoughby. *J. Abnorm. Child Psychol.*, *48*, 917–922. https://doi.org/10.1007/s10802-020-00656-1

Burns, G. L., Preszler, J., Ahnach, A., Servera, M., & Becker, S. P. (2022). Multisource network and latent variable models of sluggish cognitive tempo, ADHD-Iinattentive, and depressive symptoms with Spanish children: Equivalent findings and recommendations. *Res. Child Adolesc. Psychopathol.*, *50*(7), 881–894. https://doi.org/10.1007/s10802-021-

00890-1
Burns, G. L., Walsh, J. A., Servera, M., Lorenzo-Seva, U., Cardo, E., & Rodríguez-Fornells, A. (2013). Construct validity of ADHD/ODD rating scales: Recommendations for the evaluation of forthcoming DSM-V ADHD/ODD scales. *J. Abnorm. Child Psychol.*, *41*(1), 15–26. https://doi.org/10.1007/s10802-012-9660-5
Byrne, B. M. (2005). Factor analytic models: Viewing the structure of an assessment instrument from three perspectives. *J. Pers. Assess.*, *85*(1), 17–32. https://doi.org/10.1207/s15327752jpa8501_02
Caron, C., & Rutter, M. (1991). Comorbidity in child psychopathology: Concepts, issues and research strategies. *J. Child Psychol. Psychiatry*, *32*(7), 1063–1080. https://doi.org/10.1111/j.1469-7610.1991.tb00350.x
Carragher, N., Krueger, R. F., Eaton, N. R., & Slade, T. (2015). Disorders without borders: Current and future directions in the meta-structure of mental disorders. *Soc. Psychiatry Psychiatr. Epidemiol.*, *50*(3), 339–350. https://doi.org/10.1007/s00127-014-1004-z
Carver, C. S., Johnson, S. L., & Timpano, K. R. (2017). Toward a functional view of the p factor in psychopathology. *Clin. Psychol. Sci.*, *5*(5), 880–889. https://doi.org/10.1177/2167702617710037
Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., Meier, M. H., Ramrakha, S., Shalev, I., Poulton, R., & Moffitt, T. E. (2014). The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clin. Psychol. Sci.*, *2*(2), 119–137. https://doi.org/10.1177/2167702613497473
Caspi, A., & Moffitt, T. E. (2018). All for one and one for all: Mental disorders in one dimension. *Am. J. Psychiatry*, *175*(9), 831–844. https://doi.org/10.1176/appi.ajp.2018.17121383
Castagna, P. J., Babinski, D. E., Waxmonsky, J. G., & Waschbusch, D. A. (2021). The significance of limited prosocial emotions among externalizing disorders in children. *Eur. Child Adolesc. Psychiatry*, 1–12. https://doi.org/10.1007/s00787-020-01696-0
Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct. Equ. Modeling*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834
Chen, Y. L., Shen, L. J., & Gau, S. S. F. (2017). The Mandarin version of the Kiddie-Schedule for Affective Disorders and Schizophrenia-Epidemiological version for DSM–5 – A psychometric study. *J. Formos. Med. Assoc.*, *116*(9), 671–678. https://doi.org/10.1016/j.jfma.2017.06.013
Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Struct. Equ. Modeling*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902
Chmielewski, M., Clark, L. A., Bagby, M. R., & Watson, D. (2015). Method matters: Understanding diagnostic reliability in DSM-IV and DSM-5. *J. Abnorm. Child Psychol.*, *124*(3), 764–769. https://doi.org/10.1037/abn0000069.Method
Christensen, A. P., & Golino, H. (2021). On the equivalency of factor and network loadings. *Behav. Res. Methods*, *53*(4), 1563–1580. https://doi.org/10.3758/s13428-020-01500-6
Cicchetti, D. V. (1994). Guidlines, criteria, and rules of thumb for evalauting normed and standardized assessment instruments in psychology. *Psychol. Assess.*, *6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284
Clark, L. A., Cuthbert, B., Lewis-Fernández, R., Narrow, W. E., & Reed, G. M. (2017). Three approaches to understanding and classifying mental disorder: ICD-11, DSM-5, and the National Institute of Mental Health's Research Domain Criteria (RDoC). *Psychol. Sci. Public Interest*, *18*(2), 72–145. https://doi.org/10.1177/1529100617727266
Coghill, D., & Sonuga-Barke, E. J. S. (2012). Annual research review: Categories versus dimensions in the classification and conceptualisation of child and adolescent mental disorders - Implications of recent empirical study. *J. Child Psychol. Psychiatry Allied*

Discip., *53*(5), 469–489. https://doi.org/10.1111/j.1469-7610.2011.02511.x
Contreras, A., Nieto, I., Valiente, C., Espinosa, R., & Vazquez, C. (2019). The study of psychopathology from the network analysis perspective: A systematic review. *Psychother. Psychosom.*, *88*(2), 71–83. https://doi.org/10.1159/000497425
Conway, C. C., Forbes, M. K., Forbush, K. T., Fried, E. I., Hallquist, M. N., Kotov, R., Mullins-Sweatt, S. N., Shackman, A. J., Skodol, A. E., South, S. C., Sunderland, M., Waszczuk, M. A., Zald, D. H., Afzali, M. H., Bornovalova, M. A., Carragher, N., Docherty, A. R., Jonas, K. G., Krueger, R. F., … Eaton, N. R. (2019). A hierarchical taxonomy of psychopathology can transform mental health research. *Perspect. Psychol. Sci.*, *14*(3), 419–436. https://doi.org/10.1177/1745691618810696
Conway, C. C., Krueger, R. F., Cicero, D. C., DeYoung, C. G., Eaton, N. R., Forbes, M. K., Hallquist, M. N., Kotov, R., Latzman, R. D., Ruggero, C. J., Simms, L. J., Waldman, I. D., Waszczuk, M. A., Watson, D., Widiger, T. A., & Wright, A. G. C. (2021). Rethinking the diagnosis of mental disorders: Data-driven psychological dimensions, not categories, as a framework for mental-health research, treatment, and training. *Curr. Dir. Psychol. Sci.*, *30*(2), 151–158. https://doi.org/10.1177/0963721421990353
Costello, E. J., Mustillo, S., Erkanli, A., Keeler, G., & Angold, A. (2003). Prevalence and development of psychiatric disorders in childhood and adolescence. *Arch. Gen. Psychiatry*, *60*(8), 837–844. https://doi.org/10.1001/archpsyc.60.8.837
Cramer, A. O. J., van der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., Kendler, K. S., & Borsboom, D. (2012). Dimensions of normal personality as networks in search of equilibrium: You can't like parties if you don't like people. *Eur. J. Pers.*, *26*(4), 414–431. https://doi.org/10.1002/per.1866
Cramer, A. O. J., Waldorp, L. J., Van Der Maas, H. L. J., & Borsboom, D. (2010). Comorbidity: A network perspective. *Behav. Brain Sci.*, *33*(2–3), 137–150. https://doi.org/10.1017/S0140525X09991567
Craney, T. A., & Surles, J. G. (2002). Model-dependent variance inflation factor cutoff values. *Quality Engineering*, *14*(3), 391–403. https://doi.org/10.1081/QEN-120001878
de la Peña, F. R., Villavicencio, L. R., Palacio, J. D., Félix, F. J., Larraguibel, M., Viola, L., Ortiz, S., Rosetti, M., Abadi, A., Montiel, C., Mayer, P. A., Fernández, S., Jaimes, A., Feria, M., Sosa, L., Rodríguez, A., Zavaleta, P., Uribe, D., Galicia, F., … Ulloa, R. E. (2018). Validity and reliability of the kiddie schedule for affective disorders and schizophrenia present and lifetime version DSM-5 (K-SADS-PL-5) Spanish version. *BMC Psychiatry*, *18*(1), 1–7. https://doi.org/10.1186/s12888-018-1773-0
De Los Reyes, A. (2011). Introduction to the special section: More than measurement error: Discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. *J. Clin. Child Adolesc. Psychol.*, *40*(1), 1–9. https://doi.org/10.1080/15374416.2011.533405
De Los Reyes, A., Asmundson, G. J. G., Augenstein, T. M., & Becker, K. D. (2023). Editorial statement about JCCAP's 2023 special issue on informant discrepancies in youth mental health assessments: Observations, guidelines, and future directions grounded in 60 years of research. *J. Clin. Child Adolesc. Psychol.*, *52*(1), 147–158. https://doi.org/10.1080/15374416.2022.2158842
De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychol. Bull.*, *141*(4), 858–900. https://doi.org/10.1037/a0038498.The
De Los Reyes, A., Lerner, M. D., Keeley, L. M., Weber, R. J., Drabick, D. A. G., Rabinowitz, J., & Goodman, K. L. (2019). Improving interpretability of subjective assessments about psychological phenomena: A review and cross-cultural meta-analysis. *Rev. Gen. Psychol.*, *23*(3), 293–319. https://doi.org/10.1177/1089268019837645

De Los Reyes, A., Thomas, S. A., Goodman, K. L., & Kundey, S. M. A. (2013). Principles underlying the use of multiple informants' reports. *Annu. Rev. Clin. Psychol., 9*, 123–149. https://doi.org/10.1146/annurev-clinpsy-050212-185617.Principles

Deng, J., Wang, M. C., Shou, Y., & Gao, Y. (2021). Core features of callous–unemotional traits: Network analysis of the inventory of callous–unemotional traits in offender and community samples. *J. Clin. Psychol.*, *77*(6), 1487–1498. https://doi.org/10.1002/jclp.23090

DeYoung, C. G., Kotov, R., Krueger, R. F., Cicero, D. C., Conway, C. C., Eaton, N. R., Forbes, M. K., Hallquist, M. N., Jonas, K. G., Latzman, R. D., Rodriguez-Seijas, C., Ruggero, C. J., Simms, L. J., Waldman, I. D., Waszczuk, M. A., Widiger, T. A., & Wright, A. G. C. (2022). Answering questions about the Hierarchical Taxonomy of Psychopathology (HiTOP): Analogies to whales and sharks miss the boat. *Clin. Psychol. Sci.*, *10*(2), 279–284. https://doi.org/10.1177/21677026211049390

Diedenhofen, B., & Musch, J. (2015). Cocor: A comprehensive solution for the statistical comparison of correlations. *PLoS ONE*, *10*(4), 1–12. https://doi.org/10.1371/journal.pone.0121945

Dirks, M. A., De Los Reyes, A., Briggs-Gowan, M., Cella, D., & Wakschlag, L. S. (2012). Annual research review: Embracing not erasing contextual variability in children's behavior - Theory and utility in the selection and use of methods and informants in developmental psychopathology. *J. Child Psychol. Psychiatry*, *53*(5), 558–574. https://doi.org/10.1111/j.1469-7610.2012.02537.x

Döpfner, M. (2022). Kategoriale versus dimensionale Klassifikation psychischer Störungen in der Kinder- und Jugendpsychiatrie und -psychotherapie. In J. M. Fegert, F. Resch, P. P., M. Kaess, M. Döpfner, K. Konrad, & T. Legenbauer (Eds.), *Psychiatrie und Psychotherapie des Kindes- und Jugendalters.* Springer Reference Medizin. https://doi.org/10.1007/978-3-662-49289-5_64-1

Döpfner, M., Breuer, D., Wille, N., Erhart, M., & Ravens-Sieberer, U. (2008). How often do children meet ICD-10/DSM-IV criteria of attention deficit-/hyperactivity disorder and hyperkinetic disorder? Parent-based prevalence rates in a national sample - Results of the BELLA study. *Eur. Child Adolesc. Psychiatry*, *17*(Suppl. 1), 59–70. https://doi.org/10.1007/s00787-008-1007-y

Döpfner, M., & Görtz-Dorten, A. (2017). *Diagnostik-System für psychische Störungen nach ICD-10 und DSM-5 für Kinder und Jugendliche – III [Diagnostic System of Mental Disorders in Children and Adolescents based on the ICD-10 and DSM-5] (DISYPS–III)*. Hogrefe.

Döpfner, M., Hautmann, C., Dose, C., Banaschewski, T., Becker, K., Brandeis, D., Holtmann, M., Jans, T., Jenkner, C., Millenet, S., Renner, T., Romanos, M., & von Wirth, E. (2017). ESCAschool study: Trial protocol of an adaptive treatment approach for school-age children with ADHD including two randomised trials. *BMC Psychiatry*, *17*(1), 1–14. https://doi.org/10.1186/s12888-017-1433-9

Döpfner, M., & Lehmkuhl, G. (1997). Von der kategorialen zur dimensionalen Diagnostik. *Prax. Kinderpsychol. Kinderpsychiatr.*, *46*, 519–547.

Döpfner, M., & Petermann, F. (2012). *Diagnostik psychischer Störungen im Kindes-und Jugendalter. Leitfaden Kinder- und Jugendpsychotherapie* (3rd. Ed.). Hogrefe Verlag.

Döpfner, M., Plück, J., Kinnen, C., & Arbeitsgruppe Deutsche Child Behavior Checklist. (2014). *CBCL Handbuch-Schulalter. Manual zum Elternfragebogen über das Verhalten von Kindern und Jugendlichen, (CBCL/6-18R), zum Lehrerfragebogen über das Verhalten von Kindern und Jugendlichen (TRF/6-18R) und zum Fragebogen für Jugendliche (YSR/11-18R)*. Hogrefe.

Dose, C., Hautmann, C., & Doepfner, M. (2019). Functional impairment in children with externalizing behavior disorders: Psychometric properties of the Weiss Functional

Impairment Rating Scale–parent report in a German clinical sample. *J. Atten. Disord.*, *23*(13), 1546–1556. https://doi.org/10.1177/1087054716661234
Dougherty, L. R., Smith, V. C., Bufferd, S. J., Kessel, E., Carlson, G. A., & Klein, D. N. (2015). Preschool irritability predicts child psychopathology, functional impairment, and service use at age nine. *J. Child Psychol. Psychiatry Allied Discip.*, *56*(9), 999–1007. https://doi.org/10.1111/jcpp.12403
Dougherty, L. R., Smith, V. C., Bufferd, S. J., Stringaris, A., Leibenluft, E., Carlson, G. A., & klein. (2013). Preschool Irritability: Longitudinal Associations With Psychiatric Disorders at Age 6 and Parental Psychopathology. *J Am Acad Child Adolesc Psychiatry*, *52*(12), 1304–1313. https://doi.org/10.1016/j.jaac.2013.09.007.Preschool
Dunn, O. J., & Clark, V. (1969). Correlation coefficients measured on the same individuals. *J. Am. Stat. Assoc.*, *64*(325), 366–377.
DuPaul, G. J. (2022). The conceptual and diagnostic importance of ADHD-related impairment: A commentary on Arildskov et al. (2021). *J. Child Psychol. Psychiatry Allied Discip.*, *63*(3), 238–240. https://doi.org/10.1111/jcpp.13467
DuPaul, G. J., Reid, R., Anastopoulos, A. D., & Power, T. J. (2014). Assessing ADHD symptomatic behaviors and functional impairment in school settings: Impact of student and teacher characteristics. *Sch. Psychol. Q.*, *29*(4), 409–421. https://doi.org/10.1037/spq0000095
Eaton, N. R. (2015). Latent variable and network models of comorbidity: Toward an empirically derived nosology. *Soc. Psychiatry Psychiatr. Epidemiol.*, *50*(6), 845–849. https://doi.org/10.1007/s00127-015-1012-7
Eid, M. (2020). Multi-faceted constructs in abnormal psychology: Implications of the bifactor S - 1 model for individual clinical assessment. *J. Abnorm. Child Psychol.*, *48*, 895–900. https://doi.org/10.1007/s10802-020-00624-9
Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychol. Methods*, *22*(3), 541–562. https://doi.org/10.1037/met0000083
Eliot, T. S. (1942). *Little Gidding*. http://www.columbia.edu/itc/history/winter/w3206/edit/tseliotlittlegidding.html
Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behav. Res. Methods*, *50*(1), 195–212. https://doi.org/10.3758/s13428-017-0862-1
Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychol. Methods*, *23*(4), 617–634. https://doi.org/10.1037/met0000167
Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika*, *82*(4), 904–927. https://doi.org/10.1007/s11336-017-9557-x
Epstein, J. N., & Weiss, M. D. (2012). Assessing treatment outcomes in attention-deficit/hyperactivity disorder: A narrative review. *The Primary Care Companion for CNS Disorders*, *14*(6). https://doi.org/10.4088/PCC.11r01336
Erhart, M., Döpfner, M., Ravens-Sieberer, U., & the BELLA study group. (2008). Psychometric properties of two ADHD questionnaires: Comparing the Conners' scale and the FBB-HKS in the general population of German children and adolescents - Results of the BELLA study. *Eur. Child Adolesc. Psychiatry*, *17*(SUPPL. 1), 106–115. https://doi.org/10.1007/s00787-008-1012-1
Evans, S. C., Burke, J. D., Roberts, M. C., Fite, P. J., Lochman, J. E., de la Peña, F. R., & Reed, G. M. (2017). Irritability in child and adolescent psychopathology: An integrative review for ICD-11. *Clin. Psychol. Rev.*, *53*, 29–45. https://doi.org/10.1016/j.cpr.2017.01.004
Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using

G*Power 3.1: Tests for correlation and regression analyses. *Behav. Res. Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149
Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics* (5th ed.). SAGE.
Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychol. Bull.*, *76*(5), 378–382. https://doi.org/10.1037/h0031619
Forbes, M. K., Sunderland, M., Rapee, R. M., Batterham, P. J., Calear, A. L., Carragher, N., Ruggero, C., Zimmerman, M., Baillie, A. J., Lynch, S. J., Mewton, L., Slade, T., & Krueger, R. F. (2021). A detailed hierarchical model of psychopathology: From individual symptoms up to the general factor of psychopathology. *Clin. Psychol. Sci.*, *9*(2), 139–168. https://doi.org/10.1177/2167702620954799
Forbes, M. K., Wright, A. G. C., Markon, K. E., & Krueger, R. F. (2017a). Evidence that psychopathology symptom networks have limited replicability. *J. Abnorm. Psychol.*, *126*(7), 969–988. https://doi.org/10.1037/abn0000276
Forbes, M. K., Wright, A. G. C., Markon, K. E., & Krueger, R. F. (2017b). Further evidence that psychopathology networks have limited replicability and utility: Response to Borsboom et al. (2017) and Steinley et al. (2017). *J. Abnorm. Psychol.*, *126*(7), 1011–1016. https://doi.org/10.1037/abn0000313
Forbes, M. K., Wright, A. G. C., Markon, K. E., & Krueger, R. F. (2019). The network approach to psychopathology: Promise versus reality. *World Psychiatry*, *18*(3), 272–273. https://doi.org/10.1002/wps.20659
Forbes, M. K., Wright, A. G. C., Markon, K. E., & Krueger, R. F. (2021a). On unreplicable inferences in psychopathology symptom networks and the importance of unreliable parameter estimates. *Multivar. Behav. Res.*, *56*(2), 368–376. https://doi.org/10.1080/00273171.2021.1886897
Forbes, M. K., Wright, A. G. C., Markon, K. E., & Krueger, R. F. (2021b). Quantifying the reliability and replicability of psychopathology network characteristics. *Multivar. Behav. Res.*, *56*(2), 224–242. https://doi.org/10.1080/00273171.2019.1616526
Forbush, K. T., Swanson, T. J., Chen, Y., Siew, C. S. Q., Hagan, K. E., Chapa, D. A. N., Tregarthen, J., Wildes, J. E., & Christensen, K. A. (2022). Generalized network psychometrics of eating-disorder psychopathology. *Int. J. Eat. Disord.*, *55*(11), 1603–1613. https://doi.org/10.1002/eat.23801
Frick, P. J., Barry, C. T., & Kamphaus, R. W. (2010). *Clinical assessment of child and adolescent personality and behavior* (3rd Ed.). Springer Science+Business Media. https://doi.org/10.1007/978-1-4419-0641-0_11
Frick, P. J., Ray, J. V., Thornton, L. C., & Kahn, R. E. (2014a). Annual research review: A developmental psychopathology approach to understanding callous-unemotional traits in children and adolescents with serious conduct problems. *J. Child Psychol. Psychiatry*, *55*(6), 532–548. https://doi.org/10.1111/jcpp.12152
Frick, P. J., Ray, J. V., Thornton, L. C., & Kahn, R. E. (2014b). Can callous-unemotional traits enhance the understanding, diagnosis, and treatment of serious conduct problems in children and adolescents? A comprehensive review. *Psychol. Bull.*, *140*(1), 1–57. https://doi.org/10.1037/a0033076
Fried, E. I. (2020). Lack of theory building and testing impedes progress in the factor and network literature. *Psychol. Inq.*, *31*(4), 271–288. https://doi.org/10.1080/1047840X.2020.1853461
Fried, E. I., & Nesse, R. M. (2014). The impact of individual depressive symptoms on impairment of psychosocial functioning. *PLoS ONE*, *9*(2). https://doi.org/10.1371/journal.pone.0090311
Fried, E. I., van Borkulo, C. D., & Epskamp, S. (2021). On the importance of estimating parameter uncertainty in network psychometrics: A response to Forbes et al. (2019). *Multivar. Behav. Res.*, *56*(2), 243–248. https://doi.org/10.1080/00273171.2020.1746903

Galanter, C. A., & Patel, V. L. (2005). Medical decision making: A selective review for child psychiatrists and psychologists. *J. Child Psychol. Psychiatry Allied Discip.*, *46*(7), 675–689. https://doi.org/10.1111/j.1469-7610.2005.01452.x

Garner, A. A., O'Connor, B. C., Narad, M. E., Tamm, L., Simon, J., & Epstein, J. N. (2013). The relationship between ADHD symptom dimensions, clinical correlates, and functional impairments. *J. Dev. Behav. Pediatr.*, *34*(7), 469–477. https://doi.org/10.1097/DBP.0b013e3182a39890

Geissler, J., Jans, T., Banaschewski, T., Becker, K., Renner, T., Brandeis, D., Döpfner, M., Dose, C., Hautmann, C., Holtmann, M., Jenkner, C., Millenet, S., & Romanos, M. (2018). Individualised short-term therapy for adolescents impaired by attention-deficit/hyperactivity disorder despite previous routine care treatment (ESCAadol)-Study protocol of a randomised controlled trial within the consortium ESCAlife. *Trials*, *19*(1), 1–16. https://doi.org/10.1186/s13063-018-2635-2

Ghanizadeh, A., Mohammadi, M. R., & Yazdanshenas, A. (2006). Psychometric properties of the Farsi translation of the kiddie schedule for affective disorders and schizophrenia-present and lifetime version. *BMC Psychiatry*, *6*, 10. https://doi.org/10.1186/1471-244X-6-10

Gignac, G. E. (2016). The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intelligence*, *55*, 57–68. https://doi.org/10.1016/j.intell.2016.01.006

Gisev, N., Bell, J. S., & Chen, T. F. (2013). Interrater agreement and interrater reliability: Key concepts, approaches, and applications. *Res. Social Adm. Pharm.*, *9*(3), 330–338. https://doi.org/10.1016/j.sapharm.2012.04.004

Goh, P. K., Martel, M. M., & Barkley, R. A. (2020). Clarifying ADHD and sluggish cognitive tempo item relations with impairment: A network analysis. *J. Abnorm. Child Psychol.*, *48*(8), 1047–1061. https://doi.org/10.1007/s10802-020-00655-2

Goh, P. K., Martel, M. M., Jones, P. J., Bansal, P. S., Eng, A. G., Elkins, A. R., Thaxton, M. H., & Barkley, R. A. (2021). Clarifying relations between ADHD and functional impairment in adulthood: Utilization of network and machine learning approaches. *Assessment*. https://doi.org/10.1177/10731911211050921

Goh, P. K., Smith, T. E., Lee, C. A., Bansal, P. S., Eng, A. G., & Martel, M. M. (2021). Etiological networks of attention-deficit/hyperactivity disorder during childhood and adolescence. *J. Clin. Child Adolesc. Psychol.*, *00*(00), 1–14. https://doi.org/10.1080/15374416.2021.1946820

Gomez, R., & Stavropoulos, V. (2020). Confirmatory factor analysis and exploratory structural equation modeling of the structure of attention-deficit/hyperactivity disorder symptoms in adults. *Assessment*. https://doi.org/10.1177/1073191120905892

Gomez, R., Stavropoulos, V., Zarate, D., & Griffiths, M. (2021). ADHD symptoms, the current symptom scale, and exploratory structural equation modeling: A psychometric study. *Res. Dev. Disabil.*, *111*, 103850. https://doi.org/10.1016/j.ridd.2020.103850

Görtz-Dorten, A., Ise, E., Hautmann, C., Walter, D., & Döpfner, M. (2014). Psychometric properties of a German parent rating scale for oppositional defiant and conduct disorder (FBB-SSV) in clinical and community samples. *Child Psychiatry Hum. Dev.*, *45*(4), 388–397. https://doi.org/10.1007/s10578-013-0409-3

Görtz-Dorten, A., Thöne, A.-K., & Döpfner, M. (2022). *DISYPS-ILF Interview-Leitfäden zum Diagnostik-System für psychische Störungen nach DSM-5 für Kinder und Jugendliche [Clinical Parent Interviews for Diagnosing Mental Disorders Based on the DSM-5 in Children and Adolescents]*. Hogrefe.

Grömping, U. (2006). Relative importance for linear regression in R: The package relaimpo. *J. Stat. Softw.*, *17*(1), 1–27. https://doi.org/10.18637/jss.v017.i01

Haack, L. M., & Gerdes, A. C. (2011). Functional impairment in Latino children with ADHD:

Implications for culturally appropriate conceptualization and measurement. *Clin. Child Fam. Psychol. Rev.*, *14*(3), 318–328. https://doi.org/10.1007/s10567-011-0098-z
Haas, S. M., Becker, S. P., Epstein, J. N., & Frick, P. J. (2018). Callous-unemotional traits are uniquely associated with poorer peer functioning in school-aged children. *J. Abnorm. Child Psychol.*, *46*(4), 781–793. https://doi.org/10.1007/s10802-017-0330-5
Haeffel, G. J., Jeronimus, B. F., Fisher, A. J., Kaiser, B. N., Weaver, L. J., Vargas, I., Goodson, J. T., Soyster, P. D., & Lu, W. (2022). The Hierarchical Taxonomy of Psychopathology (HiTOP) is not an improvement over the DSM. *Clin. Psychol. Sci.*, *10*(2), 285–290. https://doi.org/10.1177/21677026211068873
Haeffel, G. J., Jeronimus, B. F., Kaiser, B. N., Weaver, L. J., Soyster, P. D., Fisher, A. J., Vargas, I., Goodson, J. T., & Lu, W. (2022). Folk classification and factor rotations: Whales, sharks, and the problems with the Hierarchical Taxonomy of Psychopathology (HiTOP). *Clin. Psychol. Sci.*, *10*(2), 259–278. https://doi.org/10.1177/21677026211002500
Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutor. Quant. Methods Psychol.*, *8*(1), 23–34. https://doi.org/10.20982/tqmp.08.1.p023
Hartman, C. A. (2021). The important gain is that we are lumpers and splitters now; it is the splitting that needs our hard work. *World Psychiatry*, *20*(1), 72–73. https://doi.org/10.1002/wps.20816
Haslam, N., Holland, E., & Kuppens, P. (2012). Categories versus dimensions in personality and psychopathology: A quantitative review of taxometric research. *Psychol. Med.*, *42*(5), 903–920. https://doi.org/10.1017/S0033291711001966
Haslam, N., McGrath, M. J., Viechtbauer, W., & Kuppens, P. (2020). Dimensions over categories: A meta-analysis of taxometric research. *Psychol. Med.*, *50*(9), 1418–1432. https://doi.org/10.1017/S003329172000183X
Heinrich, M., Geiser, C., Zagorscak, P., Leonard Burns, G., Bohn, J., Becker, S. P., Eid, M., Beauchaine, T. P., & Knaevelsrud, C. (2021). On the meaning of the general factor of psychopathology ("P-Factor") in symmetrical bifactor models: Recommendations for future research from the bifactor-(S−1) perspective. *Assessment*, 1–12. https://doi.org/10.1177/10731911211060298
Heinrich, M., Zagorscak, P., Eid, M., & Knaevelsrud, C. (2020). Giving G a meaning: An application of the bifactor-(S-1) approach to realize a more symptom-oriented modeling of the Beck Depression Inventory–II. *Assessment*, *27*(7), 1429–1447. https://doi.org/10.1177/1073191118803738
Herpers, P. C. M., Rommelse, N. N. J., Bons, D. M. A., Buitelaar, J. K., & Scheepers, F. E. (2012). Callous-unemotional traits as a cross-disorders construct. *Soc. Psychiatry Psychiatr. Epidemiol.*, *47*(12), 2045–2064. https://doi.org/10.1007/s00127-012-0513-x
Hoffman, M., Steinley, D., Trull, T. J., & Sher, K. J. (2019). Estimating transdiagnostic symptom networks: The problem of "skip outs" in diagnostic interviews. *Psychol. Assess.*, *31*(1), 73–81. https://doi.org/10.1037/pas0000644
Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electron. J. Bus. Res. Methods*, *6*(1), 53–60. https://doi.org/10.21427/D79B73
Hopwood, C. J., Kotov, R., Krueger, R. F., Watson, D., Widiger, T. A., Althoff, R. R., Ansell, E. B., Bach, B., Michael Bagby, R., Blais, M. A., Bornovalova, M. A., Chmielewski, M., Cicero, D. C., Conway, C., De Clercq, B., De Fruyt, F., Docherty, A. R., Eaton, N. R., Edens, J. F., … Zimmermann, J. (2018). The time has come for dimensional personality disorder diagnosis. *Pers. Ment. Health*, *12*(1), 82–86. https://doi.org/10.1002/pmh.1408
Hoyer, J., & Knappe, S. (2012). Psychotherapie braucht strukturierte Diagnostik! *Psychotherapie Im Dialog*, *13*(01), 2–5. https://doi.org/10.1055/s-0031-1298922

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct. Equ. Modeling*, *6*(1), 1–55.

Hunsley, J., & Mash, E. J. (2007). Evidence-based assessment. *Annu. Rev. Clin. Psychol.*, *3*, 29–51. https://doi.org/10.1146/annurev.clinpsy.3.022806.091419

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D., Quinn, K., Sanislow, C., & Wang, P. (2010). Research Domain Criteria (RDoC): Toward a new classification framework for research of mental disorders. *Am. J. Psychiatry*, *167*(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

Ivanova, M. Y., Achenbach, T. M., Dumenci, L., Rescorla, L. A., Almqvist, F., Weintraub, S., Bilenberg, N., Bird, H., Chen, W. J., Dobrean, A., Döpfner, M., Erol, N., Fombonne, E., Fonseca, A. C., Frigerio, A., Grietens, H., Hannesdóttir, H., Kanbayashi, Y., Lambert, M., … Verhulst, F. C. (2007). Testing the 8-syndrome structure of the child behavior checklist in 30 societies. *J. Clin. Child Adolesc. Psychol.*, *36*(3), 405–417. https://doi.org/10.1080/15374410701444363

Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Dias, P., Doepfner, M., Jusiene, R., Gudmundsson, H. S., Lobel, S. P., Rojas, P. O., Rodriguez, J. T., Schmeck, K., Silva, J. R., & Verhulst, F. C. (2011). Syndromes of preschool psychopathology reported by teachers and caregivers in 14 societies using the Caregiver-Teacher Report Form (C-TRF). *JECIP*, *7*, 99–116.

Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Dumenci, L., Almqvist, F., Bathiche, M., Bilenberg, N., Bird, H., Domuta, A., Erol, N., Fombonne, E., Fonseca, A., Frigerio, A., Kanbayashi, Y., Lambert, M. C., Leung, P., Liu, X., Minaei, A., Roussos, A., … Verhulst, F. C. (2007). Testing the teacher's report form syndromes in 20 societies. *School Psych. Rev.*, *36*(3), 468–483. https://doi.org/10.1080/02796015.2007.12087934

Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Dumenci, L., Almqvist, F., Bilenberg, N., Bird, H., Broberg, A. G., Dobrean, A., Döpfner, M., Erol, N., Forns, M., Hannesdottir, H., Kanbayashi, Y., Lambert, M. C., Leung, P., Minaei, A., Mulatu, M. S., Novik, T., … Verhulst, F. C. (2007). The generalizability of the youth self-report syndrome structure in 23 societies. *J. Consult. Clin. Psychol.*, *75*(5), 729–738. https://doi.org/10.1037/0022-006X.75.5.729

Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Guo, J., Althoff, R. R., Kan, K. J., Almqvist, F., Begovac, I., Broberg, A. G., Chahed, M., da Rocha, M. M., Dobrean, A., Döepfner, M., Erol, N., Fombonne, E., Fonseca, A. C., Forns, M., Frigerio, A., Grietens, H., … Verhulst, F. C. (2019). Testing syndromes of psychopathology in parent and youth ratings across societies. *J. Clin. Child Adolesc. Psychol.*, *48*(4), 596–609. https://doi.org/10.1080/15374416.2017.1405352

Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Harder, V. S., Ang, R. P., Bilenberg, N., Bjarnadottir, G., Capron, C., De Pauw, S. S. W., Dias, P., Dobrean, A., Doepfner, M., Duyme, M., Eapen, V., Erol, N., Esmaeili, E. M., Ezpeleta, L., Frigerio, A., Gonalves, M. M., … Verhulst, F. C. (2010). Preschool psychopathology reported by parents in 23 societies: Testing the seven-syndrome model of the child behavior checklist for ages 1.55. *J. Am. Acad. Child Adolesc. Psychiatry*, *49*(12), 1215–1224. https://doi.org/10.1016/j.jaac.2010.08.019

Jans, T., Weyers, P., Schneider, M., Hohage, A., Werner, M., Pauli, P., & Warnke, A. (2009). The Kiddie-SADS allows a dimensional assessment of externalizing symptoms in ADHD children and adolescents. *Attent. Deficit Hyperact. Disord.*, *1*(2), 215–222. https://doi.org/10.1007/s12402-009-0013-3

Jones, P. J., Williams, D. R., & McNally, R. J. (2021). Sampling variability is not nonreplication: A Bayesian reanalysis of Forbes, Wright, Markon, and Krueger. *Multivar. Behav. Res.*, *56*(2), 249–255. https://doi.org/10.1080/00273171.2020.1797460

Junghänel, M., Rodenacker, K., Dose, C., & Döpfner, M. (2020). Applying the bifactor S-1 model to ratings of ADHD/ODD symptoms: A commentary on Burns et al. (2019) and a re-analysis. *J. Abnorm. Child Psychol., 48*, 901–905. https://doi.org/10.1007/s10802-020-00637-4

Junghänel, M., Thöne, A.-K., Dose, C., Breuer, D., Görtz-Dorten, A., & Döpfner, M. (2023). Conceptualizing anxiety and depression in children and adolescents: A latent factor and network analysis. *Curr. Psychol.*, 1–16. https://doi.org/10.1007/s12144-023-04321-w

Junghänel, M., Thöne, A.-K., Ginsberg, C., Görtz-Dorten, A., Frenk, F., Mücke, K., Treier, A. K., Labarga, S. Z., Banaschewski, T., Millenet, S., Fegert, J. M., Bernheim, D., Hanisch, C., Kölch, M., Schüller, A., Ravens-Sieberer, U., Kaman, A., Roessner, V., Hinz, J., & Döpfner, M. (2022). Irritability and emotional impulsivity as core feature of ADHD and ODD in children. *J. Psychopathol. Behav. Assess.*, *44*, 679–697. https://doi.org/10.1007/s10862-022-09974-8

Kariuki, S. M., Newton, C. R. J. C., Abubakar, A., Bitta, M. A., Odhiambo, R., & Phillips Owen, J. (2018). Evaluation of psychometric properties and factorial structure of ADHD module of K-SADS-PL in children from rural Kenya. *J. Atten. Disord.*, 1–8. https://doi.org/10.1177/1087054717753064

Kasius, M. C. (1997). Interviewing children: Development of the Dutch version of the Semistructured Clinical Interview for Children and Adolescents (SCICA) and testing of the psychometric properties. In *Universidad de Rotterdam*. https://repub.eur.nl/pub/18144/970521_KASIUS, Marianne Christine.pdf

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., Williamson, D., & Ryan, N. (1997). Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): Initial reliability and validity data. *J. Am. Acad. Child Adolesc. Psychiatry*, *36*(7), 980–988. https://doi.org/10.1097/00004583-199707000-00021

Kernder, T., Doepfner, M., Dose, C., & Goertz-Dorten, A. (2019). Psychometric properties of a modified version of the Weiss Functional Impairment Rating Scale-Parent Report (WFIRS-P) in a clinical sample of children with aggressive behavior. *Qual. Life Res.*, *28*(1), 241–251. https://doi.org/10.1007/s11136-018-2015-0

Kessler, R. C., Chiu, W. T., Demler, O., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch. Gen. Psychiatry*, *62*(6), 617–627. http://www.ncbi.nlm.nih.gov/pubmed/17207632

Kessler, R. C., McGonagle, K., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., Wittchen, H.-U., & Kendler, K. S. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: Results from the National Comorbidity Survey. *Arch. Gen. Psychiatry*, *51*(1), 8–19.

Keyes, K. M., Eaton, N. R., Krueger, R. F., Skodol, A. E., Wall, M. M., Grant, B., Siever, L. J., & Hasin, D. S. (2013). Thought disorder in the meta-structure of psychopathology. *Psychol. Med.*, *43*(8), 1673–1683. https://doi.org/10.1017/S0033291712002292

Kim, Y. S., Cheon, K. A., Kim, B. N., Chang, S. A., Yoo, H. J., Kim, J. W., Cho, S. C., Seo, D. H., Bae, M. O., So, Y. K., Noh, J. S., Koh, Y. J., McBurnett, K., & Leventhal, B. (2004). The reliability and validity of Kiddie-Schedule for Affective Disorders and Schizophrenia-Present and Lifetime Version-Korean Version (K-SADS-PL-K). *Yonsei Med. J.*, *45*(1), 81–89. https://doi.org/10.3349/ymj.2004.45.1.81

King, K. M., Luk, J. W., Witkiewitz, K., Racz, S., McMahon, R. J., & Wu, J. (2018). Externalizing behavior across childhood as reported by parents and teachers: A partial measurement invariance model. *Assessment*, *25*(6), 744–758. https://doi.org/10.1177/1073191116660381

Kolko, D. J., & Pardini, D. A. (2010). ODD dimensions, ADHD, and callous-unemotional

traits as predictors of treatment response in children with disruptive behavior disorders. *J. Abnorm. Psychol., 119*(4), 713–725. https://doi.org/10.1037/a0020910

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med., 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kotov, R., Cicero, D. C., Conway, C. C., Deyoung, C. G., Dombrovski, A., Eaton, N. R., First, M. B., Forbes, M. K., Hyman, S. E., Jonas, K. G., Krueger, R. F., Latzman, R. D., Li, J. J., Nelson, B. D., Regier, D. A., Rodriguez-Seijas, C., Ruggero, C. J., Simms, L. J., Skodol, A. E., … Wright, A. G. C. (2022). The Hierarchical Taxonomy of Psychopathology (HiTOP) in psychiatric practice and research. *Psychol. Med., 52*(9), 1666–1678. https://doi.org/10.1017/S0033291722001301

Kotov, R., Krueger, R. F., Watson, D., Cicero, D. C., Conway, C. C., Deyoung, C. G., Eaton, N. R., Forbes, M. K., Hallquist, M. N., Latzman, R. D., Mullins-Sweatt, S. N., Ruggero, C. J., Simms, L. J., Waldman, I. D., Waszczuk, M. A., & Wright, A. G. C. (2021). The Hierarchical Taxonomy of Psychopathology (HiTOP): A quantitative nosology based on consensus of evidence. *Annu. Rev. Clin. Psychol., 17*, 83–108. https://doi.org/10.1146/annurev-clinpsy-081219-093304

Kotov, R., Waszczuk, M. A., Krueger, R. F., Forbes, M. K., Watson, D., Clark, L. A., Achenbach, T. M., Althoff, R. R., Ivanova, M. Y., Michael Bagby, R., Brown, T. A., Carpenter, W. T., Caspi, A., Moffitt, T. E., Eaton, N. R., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., … Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *J. Abnorm. Psychol., 126*(4), 454–477. https://doi.org/10.1037/abn0000258

Kottner, J., Audigé, L., Brorson, S., Donner, A., Gajewski, B. J., Hróbjartsson, A., Roberts, C., Shoukri, M., & Streiner, D. L. (2011). Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J. Clin. Epidemiol., 64*(1), 96–106. https://doi.org/10.1016/j.jclinepi.2010.03.002

Kraemer, H. C., Noda, A., & O'Hara, R. (2004). Categorical versus dimensional approaches to diagnosis: Methodological challenges. *J. Psychiatr. Res., 38*(1), 17–25. https://doi.org/10.1016/S0022-3956(03)00097-9

Krueger, R. F. (1999). The structure of common mental disorders. *Arch. Gen. Psychiatry, 56*(10), 921–926. https://doi.org/10.1001/archpsyc.56.10.921

Krueger, R. F., Caspi, A., Moffitt, T. E., & Silva, P. A. (1998). The structure and stability of common mental disorders (DSM-III-R): A longitudinal-epidemiological study. *J. Abnorm. Psychol., 107*(2), 216–227. https://doi.org/10.1001/archpsyc.58.6.597

Krueger, R. F., & Eaton, N. R. (2015). Transdiagnostic factors of mental disorders. *World Psychiatry, 14*(1), 27–29. https://doi.org/10.1002/wps.20175

Krueger, R. F., Kotov, R., Watson, D., Forbes, M. K., Eaton, N. R., Ruggero, C. J., Simms, L. J., Widiger, T. A., Achenbach, T. M., Bach, B., Bagby, R. M., Bornovalova, M. A., Carpenter, W. T., Chmielewski, M., Cicero, D. C., Clark, L. A., Conway, C., DeClercq, B., DeYoung, C. G., … Zimmermann, J. (2018). Progress in achieving quantitative classification of psychopathology. *World Psychiatry, 17*(3), 282–293. https://doi.org/10.1002/wps.20566

Krueger, R. F., & Markon, K. E. (2006). Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopathology. *Annu. Rev. Clin. Psychol., 2*, 111–133. https://doi.org/10.1146/annurev.clinpsy.2.022305.095213

Krueger, R. F., Markon, K. E., Patrick, C. J., Benning, S. D., & Kramer, M. D. (2007). Linking antisocial behavior, substance use, and personality: An integrative quantitative model of the adult externalizing spectrum. *J. Abnorm. Psychol, 116*(4), 645–666. https://doi.org/10.1037/0021-843X.116.4.645

Lahey, B. B., Applegate, B., Hakes, J. K., Zald, D. H., Hariri, A. R., & Rathouz, P. J. (2012).

Is there a general factor of prevalent psychopathology during adulthood? *J. Abnorm. Psychol.*, *121*(4), 971–977. https://doi.org/10.1037/a0028355

Lahey, B. B., Krueger, R. F., Rathouz, P. J., Waldman, I. D., & Zald, D. H. (2017). A hierarchical causal taxonomy of psychopathology across the life span. *Psychol. Bull.*, *143*(2), 142–186. https://doi.org/10.1037/bul0000069.supp

Lahey, B. B., Loeber, R., Stouthamer-Loeber, M., Christ, M. A. G., Green, S., Russo, M. F., Frick, P. J., & Dulcan, M. K. (1990). Comparison of DSM-III and DSM-III-R diagnoses for prepubertal children: Changes in prevalence and validity. *J. Am. Acad. Child Adolesc. Psychiatry*, *29*(4), 620–626. https://doi.org/10.1097/00004583-199007000-00017

Lahey, B. B., Moore, T. M., Kaczkurkin, A. N., & Zald, D. H. (2021). Hierarchical models of psychopathology: Empirical support, implications, and remaining issues. *World Psychiatry*, *20*(1), 57–63. https://doi.org/10.1002/wps.20824

Lahey, B. B., Tiemeier, H., & Krueger, R. F. (2022). Seven reasons why binary diagnostic categories should be replaced with empirically sounder and less stigmatizing dimensions. *JCPP Advances*, *2*(4), 1–8. https://doi.org/10.1002/jcv2.12108

Lahey, B. B., Zald, D. H., Perkins, S. F., Villalta-Gil, V., Werts, K. B., Van Hulle, C. A., Rathouz, P. J., Applegate, B., Class, Q. A., Poore, H. E., Watts, A. L., & Waldman, I. D. (2018). Measuring the hierarchical general factor model of psychopathology in young adults. *Int. J. Methods Psychiatr. Res.*, *27*(1), 1–9. https://doi.org/10.1002/mpr.1593

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.1109/ICDMA.2010.328

LeBreton, James, M., & Senter, Jenell, L. (2008). Answers to 20 Questions about interrater reliability and interrater agreement. *Organ. Res. Methods*, *11*(4), 815–852. https://doi.org/10.1177/1094428106296642

Lee, S., Burns, G. L., Beauchaine, T. P., & Becker, S. P. (2016). Bifactor latent structure of attention-deficit/hyperactivity disorder (ADHD)/oppositional defiant disorder (ODD) symptoms and first-order latent structure of sluggish cognitive tempo symptoms. *Psychol. Assess.*, *28*(8), 917–928. https://doi.org/10.1037/pas0000232

Leffler, J. M., Riebel, J., & Hughes, H. M. (2015). A review of child and adolescent diagnostic interviews for clinical practitioners. *Assessment*, *22*(6), 690–703. https://doi.org/10.1177/1073191114561253

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behav. Res. Methods*, *48*(3), 936–949. https://doi.org/10.3758/s13428-015-0619-7

Li, G., Wang, L., Cao, C., Fang, R., Bi, Y., Liu, P., Luo, S., Hall, B. J., & Elhai, J. D. (2020). An exploration of the DSM-5 posttraumatic stress disorder symptom latent variable network. *Eur. J. Psychotraumatol.*, *11*(1), 1–11. https://doi.org/10.1080/20008198.2020.1759279

Lindeman, R. H., Merenda, P. F., & Gold, R. Z. (1980). *Introduction to bivariate and multivariate analysis*. Scott, Foresman and Company.

Lochman, J. E., Evans, S. C., Burke, J. D., Roberts, M. C., Fite, P. J., Reed, G. M., De La Peña, F. R., Matthys, W., Ezpeleta, L., Siddiqui, S., & Elena Garralda, M. (2015). An empirically based alternative to DSM-5's disruptive mood dysregulation disorder for ICD-11. *World Psychiatry*, *14*(1), 30–33. https://doi.org/10.1002/wps.20176

Markon, K. E., Chmielewski, M., & Miller, C. J. (2011). The reliability and validity of discrete and continuous measures of psychopathology: A quantitative review. *Psychol. Bull.*, *137*(5), 856–879. https://doi.org/10.1037/a0023678

Marsh, H. W., Lüdtke, O., Muthén, B. O., Asparouhov, T., Morin, A. J., Trautwein, U., & Nagengast, B. (2010). A New Look at the Big Five Factor Structure Through Exploratory Structural Equation Modeling. *Psychol. Assess*, *22*(3), 471–491.

https://doi.org/10.1037/a0019227
Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annu. Rev. Clin. Psychol.*, *10*, 85–110. https://doi.org/10.1146/annurev-clinpsy-032813-153700
Martel, M. M., Goh, P. K., Lee, C. A., Karalunas, S. L., & Nigg, J. T. (2021). Longitudinal attention-deficit/hyperactivity disorder symptom networks in childhood and adolescence: Key symptoms, stability, and predictive validity. *J. Abnorm. Psychol.*, *130*(5), 562–574. https://doi.org/10.1037/abn0000661
Martel, M. M., Gremillion, M., Roberts, B., von Eye, A., & Nigg, J. T. (2010). The structure of childhood disruptive behaviors. *Psychol. Assess*, *22*(4), 816–826. https://doi.org/10.1037/a0020975
Martel, M. M., Levinson, C. A., Langer, J. K., & Nigg, J. T. (2016). A network analysis of developmental change in ADHD symptom structure from preschool to adulthood. *Clin. Psychol. Sci.*, *4*(6), 988–1001. https://doi.org/10.1177/2167702615618664
Martel, M. M., Levinson, C. A., Lee, C. A., & Smith, T. E. (2017). Impulsivity symptoms as core to the developmental externalizing spectrum. *J. Abnorm. Child Psychol.*, *45*(1), 83–90. https://doi.org/10.1007/s10802-016-0148-6
Martel, M. M., Markon, K., & Smith, G. T. (2017). Research review: Multi-informant integration in child and adolescent psychopathology diagnosis. *J. Child Psychol. Psychiatry Allied Discip.*, *58*(2), 116–128. https://doi.org/10.1111/jcpp.12611
Martel, M. M., Pan, P. M., Hoffmann, M. S., Gadelha, A., do Rosário, M. C., Mari, J. J., Manfro, G. G., Miguel, E. C., Paus, T., Bressan, R. A., Rohde, L. A., & Salum, G. A. (2017). A general psychopathology factor (P Factor) in children: Structural model analysis and external validation through familial risk and child global executive function. *J. Abnorm. Psychol.*, *126*(1), 137–148. https://doi.org/10.1037/abn0000205
Massetti, G. M., Lahey, B. B., Pelham, W. E., Loney, J., Ehrhardt, A., Lee, S. S., & Kipp, H. (2008). Academic achievement over eight years among children who met modified criteria for attention-deficit/hyperactivity disorder at 4–6 years of age. *J. Abnorm. Child Psychol.*, *36*(3), 399–410. https://doi.org/10.1007/s10802-007-9186-4.Academic
McConaughy, S. H., & Achenbach, T. M. (2001). *Manual for the Semistructured Interview for Children and Adolescents* (2nd Ed.). University of Vermont, Research Center for Children, Youth, & Families.
McDonald, R. P. (1999). *Test theory: A unified approach*. Erlbaum.
McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychol. Methods*, *1*(1), 30–46. https://doi.org/10.1037/1082-989X.1.1.30
McNally, R. J. (2021). Network analysis of psychopathology: Controversies and challenges. *Annu. Rev. Clin. Psychol.*, *17*, 31–53. https://doi.org/10.1146/annurev-clinpsy-081219-092850
Meier, M. A., & Meier, M. H. (2018). Clinical implications of a general psychopathology factor: A cognitive-behavioral transdiagnostic group treatment for community mental health. *J. Psychother. Integr.*, *28*(3), 253–268. https://doi.org/https://doi.org/10.1037/int0000095
Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543.
Morin, A. J., & Asparouhov, T. (2018). *Estimation of a hierarchical Exploratory Structural Equation Model (ESEM) using ESEM-within-CFA*. Substantive Methodological Synergy Research Laboratory.
Morin, A. J., Katrin Arens, A., & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Struct. Equ. Modeling*, *23*(1), 116–139.

https://doi.org/10.1080/10705511.2014.961800
Mota, V. L., & Schachar, R. J. (2000). Reformulating attention-deficit/hyperactivity disorder according to signal detection theory. *J. Am. Acad. Child Adolesc. Psychiatry*, *39*(9), 1144–1151. https://doi.org/10.1097/00004583-200009000-00014
Muthén, B. (2006). Should substance use disorders be considered as categorical or dimensional? *Addiction*, *101*, 6–16. https://doi.org/10.1111/j.1360-0443.2006.01583.x
Muthén, L. K., & Muthén, B. O. (2017). *Mplus User's Guide.* (8th ed.). Muthén & Muthén.
Neal, Z. P., Forbes, M. K., Neal, J. W., Brusco, M. J., Krueger, R., Markon, K., Steinley, D., Wasserman, S., & Wright, A. G. C. (2022). Critiques of network analysis of multivariate data in psychological science. *Nat. Rev. Methods Primers*, *2*(90), 1–2. https://doi.org/10.1038/s43586-022-00177-9
Newby, J. M., McKinnon, A., Kuyken, W., Gilbody, S., & Dalgleish, T. (2015). Systematic review and meta-analysis of transdiagnostic psychological treatments for anxiety and depressive disorders in adulthood. *Clin. Psychol. Rev.*, *40*, 91–110. https://doi.org/10.1016/j.cpr.2015.06.002
Nikolas, M. A., & Burt, S. A. (2010). Genetic and environmental influences on ADHD symptom dimensions of inattention and hyperactivity: A meta-analysis. *J. Abnorm. Psychol.*, *119*(1), 1–17. https://doi.org/10.1037/a0018010
Nishiyama, T., Sumi, S., Watanabe, H., Suzuki, F., Kuru, Y., Shiino, T., Kimura, T., Wang, C., Lin, Y., Ichiyanagi, M., & Hirai, K. (2020). The Kiddie Schedule for Affective Disorders and Schizophrenia Present and Lifetime Version (K-SADS-PL) for DSM-5: A validation for neurodevelopmental disorders in Japanese outpatients. *Compr. Psychiatry*, *96*, 152148. https://doi.org/10.1016/j.comppsych.2019.152148
Nordgaard, J., Sass, L. A., & Parnas, J. (2013). The psychiatric interview: Validity, structure, and subjectivity. *Eur. Arch. Psychiatry Clin. Neurosci.*, *263*(4), 353–364. https://doi.org/10.1007/s00406-012-0366-z
Nunnally, J. C. (1978). *Psychometric Theory* (2nd ed.). McGraw-Hill.
Palermo, T. M., Long, A. C., Lewandowski, A. S., Drotar, D., Quittner, A. L., & Walker, L. S. (2008). Evidence-based assessment of health-related quality of life and functional impairment in pediatric psychology. *J. Pediatr. Psychol.*, *33*(9), 983–996. https://doi.org/10.1093/jpepsy/jsn038
Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspect. Psychol. Sci.*, *7*(6), 528–530. https://doi.org/10.1177/1745691612465253
Pettersson, E., Larsson, H., & Lichtenstein, P. (2021). Psychometrics, interpretation and clinical implications of hierarchical models of psychopathology. *World Psychiatry*, *20*(1), 68–69. https://doi.org/10.1002/wps.20813
Pickles, A., Rowe, R., Simonoff, E., Foley, D., Rutter, M., & Silberg, J. (2001). Child psychiatric symptoms and psychosocial impairment: Relationship and prognostic significance. *Brit. J. Psychiatry*, *179*, 230–235. https://doi.org/10.1192/bjp.179.3.230
Pickles, Andrew, & Angold, A. (2003). Natural categories or fundamental dimensions: On carving nature at the joints and the rearticulation of psychopathology. *Dev. Psychopathol.*, *15*(3), 529–551. https://doi.org/10.1017/S0954579403000282
Plück, J., Scholz, K., Döpfner, M., & für die Arbeitsgruppe Deutsche Child Behavior Checklist. (2020). *Manual deutsche Vorschulalter-Formen der Child Behavior Checklist von Thomas M. Achenbach. Elternfragebogen für Klein- und Vorschulkinder (CBCL 11/2-5) und Fragebogen für ErzieherInnen von Klein- und Vorschulkindern (C-TRF 11/2-5)*. Hogrefe.
Polanczyk, G. V, Salum, G. A., Sugaya, L. S., Caye, A., & Rohde, L. A. (2015). Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *J. Child Psychol. Psychiatry Allied Discip.*, *56*(3), 345–365.

https://doi.org/10.1111/jcpp.12381
Power, T. J., Costigan, T. E., Eiraldi, R. B., & Leff, S. S. (2004). Variations in anxiety and depression as a function of ADHD subtypes defined by DSM-IV: Do subtype differences exist or not? *J. Abnorm. Child Psychol.*, *32*(1), 27–37. https://doi.org/10.1023/B:JACP.0000007578.30863.93
Power, T. J., Watkins, M. W., Anastopoulos, A. D., Reid, R., Lambert, M. C., & DuPaul, G. J. (2017). Multi-informant assessment of ADHD symptom-related impairments among children and adolescents. *J. Clin. Child Adolesc. Psychol.*, *46*(5), 661–674. https://doi.org/10.1080/15374416.2015.1079781
Preszler, J., & Burns, G. L. (2019). Network analysis of ADHD and ODD symptoms: Novel insights or redundant findings with the latent variable model? *J. Abnorm. Child Psychol.*, *47*(10), 1599–1610. https://doi.org/10.1007/s10802-019-00549-y
Preszler, J., Burns, G. L., Becker, S. P., & Servera, M. (2020). Multisource longitudinal network and latent variable model analyses of ADHD symptoms in children. *J. Clin. Child Adolesc. Psychol.*, *51*(2), 211–218. https://doi.org/10.1080/15374416.2020.1756297
Ranshoff, D. F., & Feinstein, A. R. (1978). Problems of spectrum bias in evaluating the efficacy of diagnostic tests. *N. Engl. J. Med.*, *299*(17), 926–930.
Rapee, R. M., Bögels, S. M., Van Der Sluis, C. M., Craske, M. G., & Ollendick, T. (2012). Annual research review: Conceptualising functional impairment in children and adolescents. *J. Child Psychol. Psychiatry Allied Discip.*, *53*(5), 454–468. https://doi.org/10.1111/j.1469-7610.2011.02479.x
Regier, D. A., Narrow, W. E., Clarke, D. E., Kraemer, H. C., Kuramoto, S. J., Kuhl, E. A., & Kupfer, D. J. (2013). DSM-5 field trials in the United States and Canada, part II: Test-retest reliability of selected categorical diagnoses. *Am. J. Psychiatry*, *170*(1), 59–70. https://doi.org/10.1176/appi.ajp.2012.12070999
Reich, W. (2000). Diagnostic Interview for Children and Adolescents (DICA). *J. Am. Acad. Child Adolesc. Psychiatry*, *39*(1), 59–66. https://doi.org/10.1097/00004583-200001000-00017
Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivar. Behav. Res.*, *47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555
Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *J. Pers. Assess.*, *95*(2), 129–140. https://doi.org/10.1080/00223891.2012.725437
Rescorla, L., Achenbach, T. M., Ivanova, M. Y., Dumenci, L., Almqvist, F., Bilenberg, N., Bird, H., Chen, W., Dobrean, A., Döpfner, M., Erol, N., Fombonne, E., Fonseca, A., Frigerio, A., Grietens, H., Hannesdottir, H., Kanbayashi, Y., Lambert, M., Larsson, B., … Verhulst, F. (2007). Behavioral and emotional problems reported by parents of children ages 6 to 16 in 31 societies. *J. Emot. Behav. Disord.*, *15*(3), 130–142. https://doi.org/10.1177/10634266070150030101
Rescorla, L., Ivanova, M. Y., Achenbach, T. M., Begovac, I., Chahed, M., Drugli, M. B., Emerich, D. R., Fung, D. S. S., Haider, M., Hansson, K., Hewitt, N., Jaimes, S., Larsson, B., Maggiolini, A., Marković, J., Mitrović, D., Moreira, P., Oliveira, J. T., Olsson, M., … Zhang, E. Y. (2012). International epidemiology of child and adolescent psychopathology II: Integration and applications of dimensional findings from 44 societies. *J. Am. Acad. Child Adolesc. Psychiatry*, *51*(12). https://doi.org/10.1016/j.jaac.2012.09.012
Rettew, D. C., Lynch, A. D., Achenbach, T. M., Dumenci, L., & Ivanova, M. Y. (2009). Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. *Int. J. Methods Psychiatr. Res.*, *18*(3), 169–184. https://doi.org/10.1002/mpr

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, *74*(1), 145–154.

Ringwald, W. R., Forbes, M. K., & Wright, A. G. C. (2023). Meta-analysis of structural evidence for the Hierarchical Taxonomy of Psychopathology (HiTOP) model. *Psychol. Med.*, *53*(2), 533–546. https://doi.org/10.1017/S0033291721001902

Roberts, B. W., & Caspi, A. (2001). Personality development and the person-situation debate: It's deja vu all over again. *Psychol. Inq.*, *12*(2), 104–109. https://doi.org/https://doi.org/10.1207/S15327965PLI1202_04

Rodenacker, K., Hautmann, C., Görtz-Dorten, A., & Döpfner, M. (2016). Bifactor models show a superior model fit: Examination of the factorial validity of parent-reported and self-reported symptoms of attention-deficit/hyperactivity disorders in children and adolescents. *Psychopathology*, *49*(1), 31–39. https://doi.org/10.1159/000442295

Rodenacker, K., Hautmann, C., Görtz-Dorten, A., & Döpfner, M. (2017). The factor structure of ADHD – Different models, analyses and informants in a bifactor framework. *J. Psychopathol. Behav. Assess.*, *39*(1), 92–102. https://doi.org/10.1007/s10862-016-9565-7

Rodenacker, K., Hautmann, C., Görtz-Dorten, A., & Döpfner, M. (2018). Evidence for the trait-impulsivity etiological model in a clinical sample: Bifactor structure and its relation to impairment and environmental risk. *J. Abnorm. Child Psychol.*, *46*(4), 659–669. https://doi.org/10.1007/s10802-017-0329-y

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *J. Pers. Assess.*, *98*(3), 223–237. https://doi.org/10.1080/00223891.2015.1089249

Rouquette, A., Pingault, J. B., Fried, E. I., Orri, M., Falissard, B., Kossakowski, J. J., Vitaro, F., Tremblay, R., Cote, S. M., & Borsboom, D. (2018). Emotional and behavioral symptom network structure in elementary school girls and association with anxiety disorders and depression in adolescence and early adulthood: A network analysis. *JAMA Psychiatry*, *75*(11), 1173–1181. https://doi.org/10.1001/jamapsychiatry.2018.2119

Ruggero, C. J., Kotov, R., Hopwood, C. J., First, M., Clark, L. A., Skodol, A. E., Mullins-Sweatt, S. N., Patrick, C. J., Bach, B., Cicero, D. C., Docherty, A., Simms, L. J., Michael Bagby, R., Krueger, R. F., Callahan, J. L., Chmielewski, M., Conway, C. C., Clercq, B. De, Dornbach-Bender, A., … Zimmermann, J. (2019). Integrating the Hierarchical Taxonomy of Psychopathology (HiTOP) into clinical practice. *J. Consult. Clin. Psychol.*, *87*(12), 1069–1084. https://doi.org/10.1037/ccp0000452

Segal, D. L., & Williams, K. N. (2014). Structured and semistructured interviews for differential diagnosis: Fundamental issues, applications, and features. In D. C. Beidel, B. C. Frueh, & M. Hersen (Eds.), *Adult Psychopathology and Diagnosis* (7th Ed., pp. 103–129). Wiley.

Shaffer, D., Fisher, P., Lucas, C. P., Dulcan, M. K., & Schwab-Stone, M. E. (2000). NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC- IV): Description, differences from previous versions, and reliability of some common diagnoses. *J. Am. Acad. Child Adolesc. Psychiatry*, *39*(1), 28–38. https://doi.org/10.1097/00004583-200001000-00014

Sheehan, D. V., Sheehan, K. H., Shytle, R. D., Janavs, J., Bannon, Y., Rogers, J. E., Milo, K. M., Stock, S. L., & Wilkinson, B. (2010). Reliability and validity of the Mini International Neuropsychiatric Interview for Children and Adolescents (MINI-KID). *J. Clin. Psychiatry*, *71*(3), 313–326. https://doi.org/10.4088/JCP.09m05305whi

Shen, C., Luo, Q., Jia, T., Zhao, Q., Desrivières, S., Quinlan, E. B., Banaschewski, T., Millenet, S., Bokde, A. L. W., Büchel, C., Flor, H., Frouin, V., Garavan, H., Gowland, P., Heinz, A., Ittermann, B., Martinot, J.-L., Artiges, E., Paillère-Martinot, M.-L., … for the IMAGEN consortium. (2020). Neural correlates of the dual-pathway model for

ADHD in adolescents. *Am. J. Psychiatry*, *177*(9), 844–854. https://doi.org/10.1176/appi.ajp.2020.19020183
Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychol. Bull.*, *86*(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420
Silk, T. J., Malpas, C. B., Beare, R., Efron, D., Anderson, V., Hazell, P., Jongeling, B., Nicholson, J. M., & Sciberras, E. (2019). A network analysis approach to ADHD symptoms: More than the sum of its parts. *PLoS ONE*, *14*(1), 1–17. https://doi.org/10.1371/journal.pone.0211053
Smith, T. E., Lee, C. A., Martel, M. M., & Axelrad, M. E. (2017). ODD symptom network during preschool. *J. Abnorm. Child Psychol.*, *45*(4), 743–748. https://doi.org/10.1007/s10802-016-0196-y
Spitzer, R. L., & Wakefield, J. C. (1999). DSM-IV diagnostic criterion for clinical significance: Does it help solve the false positives problem? *Am. J. Psychiatry*, *156*(12), 1856–1864. https://doi.org/10.1176/ajp.156.12.1856
Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychol. Bull.*, *87*(2), 245–251. https://doi.org/10.1037//0033-2909.87.2.245
Steinley, D. (2021). Recent advances in (graphical) network models. *Multivar. Behav. Res.*, *56*(2), 171–174. https://doi.org/10.1080/00273171.2021.1911777
Steinley, D., Hoffman, M., Brusco, M. J., & Sher, K. J. (2017). A method for making inferences in network analysis: Comment on Forbes, Wright, Markon, and Krueger (2017). *J. Abnorm. Psychol.*, *126*(7), 1000–1010. https://doi.org/10.1037/abn0000308
Taber, S. M. (2010). The veridicality of children's reports of parenting: A review of factors contributing to parent-child discrepancies. *Clin. Psychol. Rev.*, *30*(8), 999–1010. https://doi.org/10.1016/j.cpr.2010.06.014
Thöne, A.-K., Dose, C., Junghänel, M., Hautmann, C., Jendreizik, L. T., Treier, A. K., Vetter, P., von Wirth, E., Banaschewski, T., Becker, K., Brandeis, D., Dürrwächter, U., Geissler, J., Hebebrand, J., Hohmann, S., Holtmann, M., Huss, M., Jans, T., Kaiser, A., … Döpfner, M. (2023). Identifying symptoms of ADHD and disruptive behavior disorders most strongly associated with functional impairment in children: A symptom-level approach. *J. Psychopathol. Behav. Assess.*, *45*(2), 277–293. https://doi.org/10.1007/s10862-023-10025-z
Thöne, A.-K., Görtz-Dorten, A., Altenberger, P., Dose, C., Geldermann, N., Hautmann, C., Jendreizik, L. T., Treier, A.-K., von Wirth, E., Banaschewski, T., Brandeis, D., Millenet, S., Hohmann, S., Becker, K., Ketter, J., Hebebrand, J., Wenning, J., Holtmann, M., Legenbauer, T., … Döpfner, M. (2020). Toward a dimensional assessment of externalizing disorders in children: Reliability and validity of a semi-structured parent interview. *Front. Psychol.*, *11*, 1840. https://doi.org/10.3389/fpsyg.2020.01840
Thöne, A.-K., Junghänel, M., Görtz-Dorten, A., Breuer, D., del Giudice, T., Hanisch, C., Hennemann, T., & Döpfner, M. (2022). Empirically based dimensions of externalizing symptoms in children and adolescents: A multitrait-multisource approach. *J. Psychopathol. Behav. Assess.*, *44*(3), 844–861. https://doi.org/10.1007/s10862-022-09983-7
Thöne, A.-K., Junghänel, M., Görtz-Dorten, A., Dose, C., Hautmann, C., Jendreizik, L. T., Treier, A.-K., Vetter, P., von Wirth, E., Banaschewski, T., Becker, K., Brandeis, D., Dürrwächter, U., Geissler, J., Hebebrand, J., Hohmann, S., Holtmann, M., Huss, M., Jans, T., … Döpfner, M. (2021). Disentangling symptoms of externalizing disorders in children using multiple measures and informants. *Psychol. Assess.*, *33*(11), 1065–1079. https://doi.org/10.1037/pas0001053
Ulloa, R. E., Ortiz, S., Higuera, F., Nogales, I., Fresán, A., Apiquian, R., Cortés, J., Arechavaleta, B., Foulliux, C., Martínez, P., Hernández, L., Domínguez, E., & De La Peña, F. (2006). Interrater reliability of the Spanish version of Schedule for Affective

Disorders and Schizophrenia for School-Age Children-Present and Lifetime version (K-SADS-PL). *Actas Esp. Psiquiatr.*, *34*(1), 36–40.
van Bork, R., Epskamp, S., Rhemtulla, M., Borsboom, D., & van der Maas, H. L. J. (2017). What is the p-factor of psychopathology? Some risks of general factor modeling. *Theory Psychol.*, *27*(6), 759–773. https://doi.org/10.1177/0959354317737185
van Bork, R., Rhemtulla, M., Waldorp, L. J., Kruis, J., Rezvanifar, S., & Borsboom, D. (2021). Latent variable models and networks: Statistical equivalence and testability. *Multivar. Behav. Res.*, *56*(2), 175–198. https://doi.org/10.1080/00273171.2019.1672515
Van Der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychol. Rev.*, *113*(4), 842–861. https://doi.org/10.1037/0033-295X.113.4.842
Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organ. Res. Methods*, *3*(1), 4–69. https://doi.org/10.1177/109442810031002
Vitoratou, S., Garcia-Rosales, A., Banaschewski, T., Sonuga-Barke, E., Buitelaar, J., Oades, R. D., Rothenberger, A., Steinhausen, H. C., Taylor, E., Faraone, S. V., & Chen, W. (2019). Is the endorsement of the attention deficit hyperactivity disorder symptom criteria ratings influenced by informant assessment, gender, age, and co-occurring disorders? A measurement invariance study. *Int. J. Methods Psychiatr.*, *28*(4), 1–14. https://doi.org/10.1002/mpr.1794
Wakefield, J. C. (2007). The concept of mental disorder: Diagnostic implications of the harmful dysfunction analysis. *World Psychiatry*, *6*(3), 149–156.
Waldorp, L., & Marsman, M. (2022). Relations between networks, regression, partial correlation, and the latent variable model. *Multivar. Behav. Res.*, *57*(6), 994–1006. https://doi.org/10.1080/00273171.2021.1938959
Watts, A. L., Lane, S. P., Bonifay, W., Steinley, D., & Meyer, F. A. C. (2020). Building theories on top of, and not independent of, statistical models: The case of the the p-factor. *Psychol. Inq.*, *31*(4), 310–320. https://doi.org/https://doi.org/10.1080/1047840X.2020.1853476
Watts, A. L., Poore, H. E., & Waldman, I. D. (2019). Riskier tests of the validity of the bifactor model of psychopathology. *Clin. Psychol. Sci.*, *7*(6), 1285–1303. https://doi.org/10.1177/2167702619855035
Weisz, J. R., Doss, A. J., & Hawley, K. M. (2005). Youth psychotherapy outcome research: A review and critique of the evidence base. *Annu. Rev. Psychol.*, *56*, 337–363. https://doi.org/10.1146/annurev.psych.55.090902.141449
Weller, E. B., Weller, R. A., Rooney, M. T., & Fristad, M. A. (1999a). *Children's Interview for Psychiatric Syndromes –Parent version (P-ChIPS)*. American Psychiatric Press, Inc.
Weller, E. B., Weller, R. A., Rooney, M. T., & Fristad, M. A. (1999b). *Children's Interview for Psychiatric Syndromes (ChIPS)*. American Psychiatric Press, Inc.
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. *Eur. Child Adolesc. Psychiatry*, *28*(3), 351–365. https://doi.org/10.1007/s00787-018-1199-8
Willcutt, E. G., Nigg, J. T., Pennington, B. F., Solanto, M. V., Rohde, L. A., Tannock, R., Loo, S. K., Carlson, C. L., McBurnett, K., & Lahey, B. B. (2012). Validity of DSM-IV attention deficit/hyperactivity disorder symptom dimensions and subtypes. *J. Abnorm. Psychol.*, *121*(4), 991–1010. https://doi.org/10.1037/a0027347
Willoughby, M. T. (2020). Commentary on application of the bifactor S-1 model to multisource ratings of ADHD/ODD symptoms: An appropriate bifactor model for symptom ratings. *J. Abnorm. Child Psychol.*, *48*, 901–904.

https://doi.org/10.1007/s10802-020-00629-4
Winters, N. C., Collett, B. R., & Myers, K. M. (2005). Ten-year review of rating scales, VII: Scales assessing functional impairment. *J. Am. Acad. Child Adolesc. Psychiatry*, *44*(4), 309–338. https://doi.org/10.1097/01.chi.0000153230.57344.cd
Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität. [Inter-rater agreement and inter-rater reliability]*. Hogrefe.
World Health Organization. (1993). *The ICD-10 classification of mental and behavioural disorders*. World Health Organization.
World Health Organization. (2019). *International statistical classification of diseases and related health problems ICD-11*. World Health Organization.
Wright, A. G. C., Krueger, R. F., Hobbs, M. J., Markon, K. E., Eaton, N. R., & Slade, T. (2013). The structure of psychopathology: Toward an expanded quantitative empirical model. *J. Abnorm. Psychol.*, *122*(1), 281–294. https://doi.org/10.1037/a0030133.The
Zimmerman, M. (2021). Why hierarchical dimensional approaches to classification will fail to transform diagnosis in psychiatry. *World Psychiatry*, *20*(1), 70–71. https://doi.org/10.1002/wps.20815
Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's, α Revelle's β and McDonald's ω H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, *70*(1), 123–133. https://doi.org/10.1007/s11336-003-0974-7
Zoromski, A. K., Owens, J. S., Evans, S. W., & Brady, C. E. (2015). Identifying ADHD symptoms most associated with impairment in early childhood, middle childhood, and adolescence using teacher report. *J. Abnorm. Child Psychol.*, *43*(7), 1243–1255. https://doi.org/10.1007/s10802-015-0017-8
Zou, G. Y. (2012). Sample size formulas for estimating intraclass correlation coefficients with precision and assurance. *Stat. Med.*, *31*(29), 3972–3981. https://doi.org/10.1002/sim.5466

7 Appendix

7.1 Supplementary material for:

Toward a dimensional assessment of externalizing disorders in children: Reliability and validity of a semi-structured parent interview

Supplemental Table 1

Scale composition and interrater reliability of the ILF-EXTERNAL on the single item level

No.	Item	ICC(1,1)	95% CI	ICC(1,3)	95% CI	Pairwise percent agreement
ADHD Symptoms (A01-A09; B01-B09)		**.91**	**.87 - .95**	**.97**	**.95 - .98**	**88.1**
- Inattention (A01-A09)		**.83**	**.74 - .90**	**.94**	**.89 - .96**	**85.2**
A01	Careless	.73	.60 - .83	.89	.82 - .94	76.3
A02	Sustaining attention	.48	.30 - .65	.73	.57 - .85	68.9
A03	Does not listen	.80	.70 - .88	.92	.87 - .96	71.1
A04	Does not finish work	.69	.55 - .80	.87	.78 - .92	72.6
A05	Organizational skills	.71	.58 - .82	.88	.81 - .93	68.9
A06	Concentration	.74	.61 - .83	.89	.83 - .94	68.2
A07	Loses things	.84	.76 - .90	.94	.90 - .97	69.6
A08	Easily distracted	.64	.49 - .77	.84	.74 - .91	73.3
A09	Forgetful	.76	.64 - .85	.90	.84 - .94	62.2
- Hyperactivity-Impulsivity (B01-B09)		**.95**	**.91 - .97**	**.98**	**.97 - .99**	**82.2**
B01	Fidgets	.88	.82 - .93	.96	.93 - .98	82.2
B02	Leaves seat	.84	.75 - .90	.94	.90 - .97	74.1
B03	Runs / climbs	.79	.68 - .87	.92	.86 - .95	69.6
B04	Playing quietly	.77	.66 - .86	.91	.85 - .95	65.9
B05	Driven / on the go	.90	.85 - .94	.96	.74 - .98	75.6
B06	Blurts out answers	.86	.79 - .92	.95	.92 - .97	70.4
B07	Awaiting turn	.87	.80 - .92	.95	.92 - .97	71.9
B08	Interrupts / intrudes	.80	.70 - .88	.93	.88 - .86	73.3
B09	Talks excessively	.86	.78 - .91	.95	.91 - .97	77.8
ADHD Functional Impairment (F01-F05) (n = 39)		**.89**	**.82 - .94**	**.96**	**.93 - .98**	**80.7**
F01	Psychological strain (n = 43)	.85	.77 - .91	.95	.91 - .97	76.7
F02	Interferes with home life and family members (n = 42)	.77	.65 - .86	.91	.85 - .95	74.8
F03	Interferes with adult interaction (n = 42)	.82	.72 - .89	.93	.88 - .96	75.6
F04	Interferes with child interactions; interferes with recreational activities (n = 42)	.84	.75 - .90	.94	.90	72.4
F05	Interferes with educational activities (n = 39)	.77	.66 - .87	.91	.81 - .95	71.5
ODD/CD Symptoms - short version (A01-A08; B01-B05)		**.94**	**.90 - .96**	**.98**	**.97 - .99**	**91.1**
A01	Loses temper[1]	.84	.76 - .90	.94	.90 - .97	70.4
A02	Touchy / easily annoyed[1]	.89	.83 - .93	.96	.94 - .98	80.0

No.	Item	ICC(1,1)	95% CI	ICC(1,3)	95% CI	Pairwise percent agreement
A03	Angry / resentful[1]	.87	.80 - .92	.95	.93 - .97	78.5
A04	Argues with adults	.83	.74 - .90	.94	.90 - .96	76.3
A05	Complies with requests	.80	.70 - .88	.92	.88 - .96	69.6
A06	Annoys	.89	.83 - .93	.96	.94 - .98	78.5
A07	Blames others	.88	.81 - .92	.95	.93 - .98	78.5
A08	Spiteful / vindictive	.90	.85 - .94	.97	.94 - .98	91.1
- CD Symptoms - short version (B01-B05)		**.90**	**.85 - .94**	**.97**	**.94 - .98**	**88.2**
B01	Physical fights	.88	.81 - .93	.96	.93 - .98	88.9
B02	Bullies, threatens, or intimidates (n = 44)	.85	.77 - .91	.95	.91 - .97	79.5
B03	Cruel to animals	.90	.85 - .94	.97	.94 - .98	94.1
B04	Lies	.77	.65 - .85	.91	.85 - .95	65.9
B05	Steals without confrontation	.78	.67 - .86	.91	.86 - .95	86.7
B06	Uses weapon in fight[2]		-	-	-	-
B07	Cruel to people[2]	-	-	-	-	-
B08	Steals with confrontation[2]	-	-	-	-	-
B09	Sexual assault[2]	-	-	-	-	-
B10	Fire setting[2]	-	-	-	-	-
B11	Vandalism[2]	-	-	-	-	-
B12	Breaking in[2]	-	-	-	-	-
B13	Stays out at night[2]	-	-	-	-	-
B14	Runs away from home overnight[2]	-	-	-	-	-
B15	Truancy[2]	-	-	-	-	-
Disruptive Mood Dysregulation (D01-D02; A01-A03)		**.90**	**.85 - .94**	**.97**	**.94 - .98**	**83.7**
D01	Recurrent temper outbursts	.74	.61 - .83	.89	.83 - .94	69.6
D02	Persistently irritable or angry mood	.70	.57 - .81	.88	.80 - .93	65.2
Limited Prosocial Emotions (C01a-C04d) (n = 41)		**.93**	**.89 - .96**	**.98**	**.96 - .99**	**86.7**
C01a	Lack of remorse / guilt (n = 44)	.92	.88 - .96	.97	.96 - .98	92.4
C01b	Lack of concern (n = 44)	.89	.82 - .93	.96	.93 - .98	80.3
C02a	Cold and uncaring (n = 44)	.89	.83 - .94	.96	.94 - .98	85.6
C02b	Self-serving (n = 43)	.92	.88 - .96	.97	.96 - .99	87.8
C03a	Indifferent of poor performance (n = 42)	.86	.78 - .92	.95	.91 - .97	84.1
C03b	Avoids effort (n = 42)	.85	.76 - .91	.94	.90 - .97	76.2
C03c	Blames others for poor performance (n = 42)	.87	.80 - .93	.95	.92 - .97	76.2
C04a	Shallow / deficient affect (n = 42)	.69	.56 - .81	.87	.79 - .93	78.6
C04b	Turns emotions ‘on’ or ‘off’ quickly (n = 40)	.81	.71 - .89	.93	.88 - .96	80.8
C04c	Manipulates (n = 41)	.83	.73 - .90	.94	.89 - .96	83.7
C04d	Inconsistent affect (n = 41)	.87	.80 - .93	.95	.92 - .97	93.5
ODD/CD Functional Impairment (F01-F05) (n = 31)		**.92**	**.86 - .96**	**.97**	**.95 - .99**	**85.2**
F01	Psychological strain (n = 33)	.79	.66 - .88	.92	.85 - .96	72.3
F02	Interferes with home life and family members (n = 32)	.80	.67 - .89	.92	.86 - .96	75.2

No.	Item	ICC(1,1)	95% CI	ICC(1,3)	95% CI	Pairwise percent agreement
F03	Interferes with adult interaction (n = 32)	.81	.69 - .89	.93	.87 - .96	68.0
F04	Interferes with child interactions; interferes with recreational activities (n = 31)	.89	.81 - .94	.96	.93 - .98	80.8
F05	Interferes with educational activities (n = 31)	.88	.80 - .94	.96	.92 - .98	81.9

Note. ADHD = attention-deficit/hyperactivity disorder; CD = conduct disorder; CI = confidence interval; ICC = Intraclass correlation; ICC(1,1) = one-way random-effects, absolute agreement model for single rater/measurements; ICC(1,3) = one-way random-effects, absolute agreement model based on a mean-rating; ODD = oppositional defiant disorder; n = 45 (if not otherwise specified).

[1] Items also included in the Disruptive Mood Dysregulation scale.

[2] The items B06 to B15 assessing aggressive and antisocial symptoms from the age of 11 were excluded from further analyses due to an obvious floor effect.

Supplemental Table 2

Interrater reliability of the ILF-EXTERNAL scales between two independent raters

Scale	ICC(2,1)	95% CI	ICC(2,2)	95% CI	Pairwise percent agreement	*n*
ADHD Symptoms	.95	.83 - .98	.97	.91 - .99	86.7	45
- Inattention	.84	.71 - .91	.91	.83 - .95	82.2	45
- Hyperactivity-Impulsivity	.96	.91 - .98	.98	.95 - .99	75.6	45
ADHD Functional Impairment	.90	.82 - .95	.95	.90 - .97	75.6	39
ODD/CD Symptoms - short version	.96	.92 - .98	.98	.96 - .99	86.7	45
- ODD Symptoms	.95	.90 - .97	.97	.95 - .99	84.4	45
- CD Symptoms - short version	.91	.84 - .95	.95	.91 - .97	86.7	44
Disruptive Mood Dysregulation	.91	.85 - .95	.96	.92 - .98	80.0	45
Limited Prosocial Emotions	.97	.94 - .98	.98	.97 - .99	82.2	41
ODD/CD Functional Impairment	.92	.84 - .96	.96	.91 - .98	77.8	31

Note. ADHD = attention-deficit/hyperactivity disorder; CD = conduct disorder; CI = confidence interval; ICC = Intraclass correlation; ICC(2,1) = two-way random-effects, absolute agreement model for single rater/measurements; ICC(2,2) = two-way random-effects, absolute agreement model based on a mean-rating; ODD = oppositional defiant disorder.

7.2 Supplementary material for:

Disentangling symptoms of externalizing disorders in children using multiple measures and informants

Table S1

Descriptive information for all dimensions of externalizing disorders

Dimension	Number of items	Cronbach's α	Item-Total Correlation	Mean (*SD*)	*N*
Clinical Interview					
IN	9	.71	.29 - .49	1.95 (0.48)	474
HI	9	.87	50 - .68	1.64 (0.73)	474
ODD-AD	10	.86	.43 - .66	1.05 (0.62)	452
CD	5	.60	.22 - .47	0.40 (0.43)	450
CU	11	.77	.23 - .58	0.50 (0.42)	446
Parent Ratings					
IN	9	.86	.50 - .68	2.03 (0.57)	421
HI	9	.90	.61 - .70	1.70 (0.74)	413
ODD-AD	10	.90	.56 - .76	1.22 (0.66)	425
CD	5	.64	.29 - .55	0.37 (0.42)	422
CU	11	.86	.48 - .68	0.68 (0.54)	423
Teacher Ratings					
IN	9	.87	.56 - .72	1.91 (0.64)	246
HI	9	.93	.68 - .80	1.37 (0.89)	251
ODD-AD	10	.92	.49 - .80	0.91 (0.74)	262
CD	5	.67	.19 - .70	0.37 (0.43)	234
CU	11	.88	.47 - .72	0.68 (0.62)	256

Note. Clinician ratings were based on the *Clinical Parent Interview for Externalizing Disorders in Children and Adolescents* (ILF-EXTERNAL). Parents and teachers completed symptom checklists for the assessment of ADHD symptoms and symptoms of disruptive behavior disorders (FBB-ADHS; FBB-SSV). IN = Inattention; HI = Hyperactivity-impulsivity; ODD-AD = Oppositionality with chronic irritability/anger; CD = Conduct disorder; CU = Callous-unemotional

Table S2

Standardized factor loadings (standard error) for the unidimensional CFA model, first-order CFA model, higher-order CFA model, and ESEM solutions for clinician ratings

Symptoms	Unidimensional CFA	First-Order CFA	First-Order ESEM					Higher-Order CFA	Higher-Order ESEM				
			IN	HI	ODD-AD	CD	CU		IN	HI	ODD-AD	CD	CU
A01 Careless	.24 (.05)***	.45 (.05)***	**.50 (.05)*** **	-.02 (.05)	-.03 (.05)	-.21 (.06)***	.14 (.05)**	.44 (.06)***	**.51 (.05)*** **	0.02 (.06)	-.07 (.06)	-.24 (.07)**	.17 (.07)**
A02 Sustaining attention	.30 (.04)***	.53 (.05)***	**.47 (.06)*** **	.12 (.05)*	.06 (.05)	-.20 (.06)***	.01 (.06)	.51 (.05)***	**.49 (.05)*** **	0.17 (.06)**	.03 (.06)	-.22 (.06)***	.02 (.06)
A03 Does not listen	.28 (.05)***	.46 (.06)***	**.41 (.06)*** **	.08 (.05)	.09 (.05)	-.09 (.06)	-.06 (.06)	.45 (.06)***	**.41 (.06)*** **	0.12 (.06)*	.07 (.06)	-.13 (.06)*	-.03 (.06)
A04 Does not finish work	.38 (.04)***	.64 (.05)***	**.67 (.04)*** **	-.02 (.05)	.05 (.05)	.02 (.06)	-.04 (.06)	.65 (.05)***	**.70 (.04)*** **	-0.02 (.00)***	.05 (.00)***	.02 (.00)***	-.04 (.00)***
A05 Organizational skills	.36 (.04)***	.60 (.05)***	**.66 (.05)*** **	-.09 (.05)*	.04 (.05)	.15 (.06)*	-.05 (.05)	.61 (.05)***	**.63 (.05)*** **	-0.06 (.06)	.01 (.07)	.10 (.07)	.02 (.07)
A06 Concentration	.28 (.04)***	.43 (.05)***	**.37 (.06)*** **	-.09 (.05)	.12 (.06)+	-.13 (.06)*	.16 (.06)**	.45 (.05)***	**.37 (.05)*** **	-0.07 (.06)	.09 (.07)	-.13 (.07)	.18 (.06)**
A07 Loses things	.31 (.04)***	.50 (.05)***	**.46 (.06)*** **	.03 (.05)	-.08 (.06)	.15 (.06)*	.06 (.06)	.51 (.05)***	**.46 (.06)*** **	0.04 (.06)	-.09 (.06)	.14 (.07)	.10 (.07)
A08 Easily distracted	.38 (.05)***	.61 (.05)***	**.48 (.06)*** **	.31 (.06)***	.03 (.06)	-.06 (.06)	-.19 (.06)**	.58 (.06)***	**.49 (.06)** **	0.35 (.06)***	.01 (.06)	-.10 (.07)	-.17 (.06)**
A09 Forgetful	.36 (.04)***	.60 (.05)***	**.58 (.05)*** **	-0.040 (.05)	-.02 (.05)	.06 (.05)	.11 (.05)*	.61 (.05)***	**.57 (.05)*** **	-0.02 (.07)	-.04 (.07)	.03 (.07)	.16 (.07)*
B01 Fidgets	.63 (.03)***	.77 (.03)***	.15 (.04)**	**.78 (.03)*** **	-.01 (.04)	-.00 (.05)	-.09 (.05)	.77 (.03)***	.15 **(.01)*** **	**.80 (.03)*** **	-.01 (.00)***	-.00 (.00)***	-.09 (.00)***
B02 Leaves seat	.60 (.03)***	.73 (.03)***	.03 (.04)	**.71 (.04)*** **	.03 (.05)	.11 (.05)*	-.03 (.05)	.73 (.03)***	.07 **(.05)**	**.70 (.04)*** **	.05 (.06)	.09 (.06)	-.06 (.06)
B03 Runs / climbs	.66 (.03)***	.79 (.03)***	.07 (.04)	**.76 (.03)*** **	.01 (.04)	.09 (.04)*	-.02 (.05)	.79 (.03)***	.12 **(.06)***	**.75 (.04)*** **	.04 (.05)	.06 (.06)	-.05 (.06)
B04 Playing quietly	.52 (.04)***	.63 (.04)***	.09 (.05)	**.49 (.05)*** **	.04 (.04)	.13 (.05)*	.07 (.06)	.63 (.04)***	.12 **(.06)***	**.48 (.05)*** **	.06 (.06)	.12 (.06)	.05 (.06)
B05 Driven / on the go	.65 (.03)***	.78 (.03)***	.08 (.04)	**.74 (.03)*** **	.01 (.05)	.08 (.04)	.00 (.04)	.78 (.03)***	.12 **(.06)***	**.73 (.04)*** **	.04 (.05)	.06 (.06)	-.03 (.06)
B06 Blurts out answers	.52 (.04)***	.64 (.03)***	.04 (.05)	**.62 (.04)*** **	.04 (.05)	-.03 (.05)	.03 (.05)	.64 (.03)***	.09 **(.06)**	**.61 (.04)*** **	.06 (.06)	-.06 (.06)	.10 (.06)
B07 Awaiting turn	.57 (.03)***	.69 (.03)***	.06 (.05)	**.58 (.04)*** **	.07 (.05)	.02 (.05)	.09 (.05)	.69 (.03)***	.11 **(.06)**	**.57 (.04)*** **	.09 (.06)	-.00 (.06)	.07 (.06)
B08 Interrupts / intrudes	.64 (.03)***	.76 (.03)***	.12 (.05)**	**.63 (.04)*** **	.10 (.04)*	.04 (.05)	.02 (.05)	.76 (.03)***	.16 **(.06)** **	**.63 (.04)*** **	.12 (.05)*	.01 (.06)	.00 (.06)
B09 Talks excessively	.49 (.04)***	.61 (.04)***	.07 (.05)	**.58 (.04)*** **	.00 (.05)	-.12 (.05)*	.10 (.05)	.60 (.04)***	.12 **(.06)***	**.58 (.04)*** **	.02 (.06)	-.15 (.06)*	.08 (.06)

Symptoms	Unidimensional CFA	First-Order CFA	First-Order ESEM					Higher-Order CFA	Higher-Order ESEM				
			IN	HI	ODD-AD	CD	CU		IN	HI	ODD-AD	CD	CU
A01 Loses temper	.74 (.02)***	.77 (.03)***	.00 (.04)	-.04 (.04)	**1.01 (.03)*** **	-.26 (.04)***	-.05 (.04)	.82 (.02)***	.00 (.00)***	-.04 (.00)***	**1.01 (.03)*** **	-.26 (.02)***	-.05 (.01)***
A02 Touchy / easily annoyed	.66 (.03)***	.73 (.03)***	.20 (.04)***	-.08 (.04)	**.73 (.04)*** **	.10 (.05)*	-.08 (.04)	.74 (.03)***	.15 (.05)**	-.07 (.06)	**.72 (.04)*** **	.09 (.06)	-.04 (.05)
A03 Angry / resentful	.57 (.03)***	.79 (.03)***	.10 (.05)*	-.01 (.05)	**.54 (.05)*** **	.15 (.05)**	.04 (.05)	.65 (.03)***	.07 (.06)	-.02 (.06)	**.55 (.05)*** **	.15 (.06)**	.06 (.06)
A04 Argues with adults	.64 (.03)***	.63 (.04)***	.01 (.05)	.07 (.04)	**.58 (.04)*** **	.05 (.05)	.14 (.05)**	.71 (.03)***	.00 (.05)	.05 (.05)	**.59 (.04)*** **	.07 (.06)	.14 (.06)
A05 Complies with requests	.67 (.03)***	.78 (.03)***	.04 (.04)	.05 (.04)	**.49 (.04)*** **	.19 (.05)	.24 (.05)***	.75(.03)***	.03 (.06)	.01 (.06)	**.51 (.06)*** **	.22 (.07)**	.24 (.07)**
A06 Annoys	.56 (.04)***	.64 (.03)***	-.15 (.06)**	.21 (.05)***	**.35 (.05)*** **	.27 (.06)***	.17 (.06)**	.62 (.04)***	-.15 (.07)*	.16 (.07)*	**.39 (.06)*** **	.28 (.06)***	.16 (.08)*
A07 Blames others	.57 (.04)***	.69 (.03)***	.16 (.05)**	.09 (.05)	**.21 (.05)*** **	.17 (.06)*	.32 (.06)***	.63 (.04)***	.16 (.07)*	.05 (.07)	**.23 (.07)** **	.18 (.07)**	.33 (.07)***
A08 Spiteful / vindictive	.52 (.05)***	.76 (.03)***	-.03 (.07)	.01 (.07)	**.44 (.07)*** **	.25 (.08)**	.13 (.07)	.60 (.05)***	-.07 (.08)	-.03 (.08)	**.46 (.07)*** **	.27 (.08)***	.13 (.08)
D01 Recurrent temper outbursts	.75 (.02)***	.61 (.04)***	-.06 (.04)	.00 (.04)	**.94 (.03)*** **	-.17 (.04)**	-.03 (.04)	.81 (.02)***	-.07 (.05)	.00 (.05)	**.92 (.04)*** **	-.13 (.06)*	-.04 (.05)
D02 Persistently irritable or angry mood	.64 (.03)***	.77 (.03)***	.07 (.05)	-.06 (.05)	**.74 (.04)*** **	.18 (.05)*	-.10 (.05)	.72 (.03)***	.02 (.06)	-.06 (.07)	**.74 (.05)*** **	.20 (.07)**	-.08 (.07)
B01 Physical fights	.55 (.04)***	.71 (.04)***	-.06 (.06)	.17 (.06)**	.41 (.05)***	**.36 (.07)*** **	-.03 (.06)	.71 (.04)***	-.08 (.06)	.14 (.07)*	.44 (.05)***	**.37 (.06)*** **	-.03 (.07)
B02 Bullies, threatens, or intimidates	.59 (.04)***	.77 (.04)***	-.05 (.06)	.11 (.06)	.46 (.06)***	**.49 (.07)*** **	-.06 (.06)	.77 (.04)***	-.05 (.00)***	.11 (.00)***	.46 (.01)***	**.48 (.04)*** **	-.06 (.01)***
B03 Cruel to animals	.43 (.07)***	.55 (.08)***	.02 (.10)	.17 (.10)	.05 (.09)	**.25 (.12)***	.28 (.10)**	.55 (.08)***	.04 (.11)	.12 (.10)	.08 (.10)	**.27 (.15)***	.27 (.11)*
B04 Lies	.47 (.04)***	.62 (.05)***	.08 (.05)	.05 (.05)	.13 (.05)*	**.27 (.06)*** **	.31 (.06)***	.62 (.05)***	.09 (.07)	-.00 (.07)	.15 (.07)*	**.29 (.07)*** **	.31 (.07)***
B05 Steals without confrontation	.39 (.06)***	.50 (.07)***	-.08 (.07)	.22 (.07)**	.06 (.08)	**.31 (.09)** **	.20 (.08)*	.50 (.07)***	-.06 (.08)	.16 (.08)	.09 (.09)	**.34 (.09)*** **	.18 (.10)
C01a Lack of remorse / guilt	.56 (.04)***	.77 (.04)***	0.02 (.05)	-.04 (.05)	.09 (.05)	.42 (.06)***	**.51 (.06)*** **	.77 (.04)***	.01 (.08)	-.12 (.09)	.14 (.09)	.42 (.08)***	**.53 (.07)*** **
C01b Lack of concern	.55 (.04)***	.74 (.04)***	0.01 (.05)	.02 (.05)	.14 (.05)**	.26 (.06)***	**.48 (.06)*** **	.75 (.04)***	.01 (.07)	-.04 (.07)	.18 (.08)*	.27 (.08)**	**.49 (.07)*** **
C02a Cold and uncaring	.52 (.05)***	.73 (.04)***	0.06 (.06)	-.14 (.06)*	.10 (.05)	.19 (.07)**	**.63 (.06)*** **	.72 (.04)***	.05 (.08)	-.20 (.09)*	.13 (.09)	.20 (.09)*	**.65 (.07)*** **
C02b Self-serving	.54 (.04)***	.74 (.04)***	-0.02 (.05)	.11 (.05)*	-.01 (.05)	.12 (.05)**	**.68 (.05)*** **	.74 (.04)***	.01 (.08)	.04 (.08)	.03 (.09)	.15 (.09)	**.67 (.06)*** **
C03a Indifferent of poor performance	.21 (.05)***	.37 (.06)***	0.16 (.06)**	-.15 (.06)*	-.17 (.05)**	.03 (.07)	**.56 (.06)*** **	.37 (.06)***	.17 (.08)*	-.19 (.08)*	-.16 (.09)	.04 (.09)	**.58 (.07)*** **
C03b Avoids effort	.27 (.04)***	.40 (.05)***	0.14 (.06)*	-.08 (.06)	-.05 (.05)	.03 (.06)	**.45 (.06)*** **	.40 (.05)***	.16 (.06)*	-.12 (.07)	-.04 (.07)	.04 (.07)	**.46 (.06)*** **
C03c Blames others for poor performance	.37 (.05)***	.51 (.05)***	0.07 (.06)	.03 (.06)	-.03 (.05)	.12 (.06)	**.47 (.06)*** **	.51 (.05)***	.08 (.08)	-.02 (.08)	-.01 (.08)	.13 (.08)	**.47 (.06)*** **

Symptoms	Unidimensional CFA	First-Order CFA	First-Order ESEM					Higher-Order CFA	Higher-Order ESEM				
			IN	HI	ODD-AD	CD	CU		IN	HI	ODD-AD	CD	CU
C04a Shallow / deficient affect	.31 (.06)***	.48 (.06)***	-0.08 (.07)	-.01 (.06)	.04 (.05)	-.30 (.06)***	**.71 (.06)***	.48 (.06)***	-.09 (.01)***	-.01 (.00)***	.04 (.00)***	-.30 (.03)***	**.72 (.07)***
C04b Turns emotions 'on' or 'off' quickly	.36 (.05)***	.50 (.06)***	-0.09 (.07)	.16 (.07)*	.04 (.05)	-.11 (.07)	**.51 (.07)***	.50 (.06)***	-.04 (.08)	.11 (.08)	.05 (.07)	-.07 (.08)	**.48 (.07)***
C04c Manipulates	.26 (.05)***	.35 (.06)***	-0.14 (.06)*	.18 (.07)*	.09 (.05)	-.30 (.07)***	**.43 (.07)***	.35 (.06)***	-.07 (.07)	.15 (.08)	.09 (.07)	-.24 (.08)**	**.38 (.07)***
C04d Inconsistent affect	.44 (.05)***	.60 (.06)***	0.14 (.06)*	.07 (.07)	.15 (.05)*	-.39 (.07)***	**.58 (.08)****	.61 (.06)***	.19 (.08)*	.06 (.08)	.13 (.09)	-.35 (.09)***	**.56 (.08)***

Note. CFA = Confirmatory factor analysis; ESEM = Exploratory structural equation modeling; IN = Inattention; HI = Hyperactivity-impulsivity; ODD-AD = Oppositionality with chronic irritability/anger; CD = Conduct disorder; CU = Callous-unemotional. Target factor loadings are in bold. *p < .05; **p < .01; ***p < .001.

Table S3

Standardized factor correlations for the first-order CFA model (above the diagonal) and first-order ESEM model (below the diagonal)

Factors	1	2	3	4	5
1 IN	-	.50***	.36***	.30***	.38***
2 HI	.32***	-	.47***	.47***	.33***
3 ODD-AD	.25***	.36***	-	.75***	.58***
4 CD	.07	.11**	.32***	-	.59***
5 CU	.31***	.18***	.39***	.26***	-

Note. CFA = Confirmatory factor analysis; ESEM = Exploratory structural equation modeling; IN = Inattention; HI = Hyperactivity-impulsivity; ODD-AD = Oppositionality with chronic irritability/anger; CD = Conduct disorder; CU = Callous-unemotional.
**$p < .01$; ***$p < .001$.

Table S4

Standardized factor loadings (standard error) of the first-order factors on the second-order factor and residual variances (standard error) for the higher-order CFA model and higher-order ESEM solution

Factors	Higher-Order CFA Model		Higher-Order ESEM Model	
	Factor Loadings on Second-Order Factor	Residual Variances	Factor Loadings on Second-Order Factor	Residual Variances
Externalizing Spectrum Disorders				
Inattention	.55 (.05)***	.70 (.05)***	.43 (.11)***	.82 (.10)***
Hyperactivity-Impulsivity	.60 (.04)***	.64 (.05)***	.51 (.12)***	.74 (.12)***
Oppositionality with Chronic Irritability/Anger	.82 (.03)***	.34 (.05)***	.68 (.10)***	.50 (.14)***
Conduct Disorder	.85 (.04)***	.28 (.07)***	.37 (.10)***	.86 (.08)***
Callous-Unemotional	.67 (.04)***	.55 (.05)***	.54 (.11)***	.71 (.12)***

Note. CFA = Confirmatory Factor analysis; ESEM = Exploratory structural equation modeling.
***$p < .001$.

Table S5

Standardized factor loadings (standard error), omega factor reliability coefficients, and explained common variance for the bifactor CFA model and bifactor ESEM solution for clinician ratings

Symptoms	Bifactor CFA						Bifactor ESEM					
	G-EXT	S-IN	S-HI	S-ODD-AD	S-CD	S-CU	G-EXT	S-IN	S-HI	S-ODD-AD	S-CD	S-CU
A01 Careless	.18 (.06)**	.50 (.05)***					.13 (.06)*	**.49 (.05)*** **	.08 (.05)	-.00 (.05)	-.13 (.06)*	.19 (.05)***
A02 Sustaining attention	.26 (.05)***	.48 (.05)***					.14 (.05)**	**.49 (.05)*** **	.22 (.05)***	.09 (.04)*	.01 (.06)	.11 (.06)
A03 Does not listen	.25 (.05)***	.35 (.06)***					.22 (.06)***	**.35 (.06)*** **	.15 (.05)**	.04 (.05)	-.30 (.06)***	-.03 (.06)
A04 Does not finish work	.33 (.05)***	.60 (.04)***					.29 (.05)***	**.62 (.04)*** **	.07 (.04)	.01 (.04)	.08 (.05)	-.01 (.05)
A05 Organizational skills	.32 (.05)***	.53 (.04)***					.36 (.05)***	**.54 (.04)*** **	-.03 (.04)	-.06 (.05)	-.04 (.06)	-.07 (.05)
A06 Concentration	.27 (.05)***	.32 (.06)***					.20 (.05)***	**.37 (.05)*** **	-.01 (.05)	.10 (.05)*	.03 (.06)	.19 (.05) ***
A07 Loses things	.29 (.05)***	.39 (.05)***					.29 (.05)***	**.41 (.05)*** **	.03 (.05)	-.11 (.05)*	.17 (.06)*	.02 (.06)
A08 Easily distracted	.34 (.05)***	.41 (.06)***					.19 (.06)**	**.45 (.06)*** **	.37 (.05)***	.05 (.05)	-.00 (.06)	-.10 (.06)
A09 Forgetful	.34 (.05)***	.48 (.05)***					.33 (.05)***	**.50 (.04)*** **	.01 (.04)	-.07 (.05)	-.02 (.05)	.08 (.05)
B01 Fidgets	.39 (.05)***		.71 (.04)***				.33 (.06)***	.16 (.04)***	**.73 (.04)*** **	.03 (.04)	0.05 (.05)	-.06 (.05)
B02 Leaves seat	.41 (.05)***		.61 (.04)***				.37 (.05)***	.04 (.04)	**.63 (.04)*** **	.03 (.04)	0.20 (.05)***	-.05 (.05)
B03 Runs / climbs	.45 (.05)***		.66 (.04)***				.41 (.05)***	.08 (.04)*	**.68 (.03)*** **	.02 (.03)	0.11 (.04)*	-.04 (.04)
B04 Playing quietly	.43 (.05)***		.42 (.04)***				.40 (.05)***	.08 (.05)	**.42 (.04)*** **	.00 (.05)	0.14 (.05)**	.01 (.05)
B05 Driven / on the go	.44 (.05)***		.66 (.04)***				.40 (.05)***	.09 (.04)*	**.66 (.04)*** **	.02 (.04)	0.13 (.05)**	-.02 (.04)
B06 Blurts out answers	.37 (.05)***		.54 (.04)***				.35 (.05)***	.03 (.04)	**.56 (.04)*** **	.01 (.05)	-0.17 (.05)***	-.00 (.05)
B07 Awaiting turn	.46 (.05)***		.48 (.04)***				.42 (.05)***	.05 (.04)	**.52 (.04)*** **	.03 (.05)	-0.06 (.05)	.04 (.05)
B08 Interrupts / intrudes	.50 (.04)***		.54 (.04)***				.47 (.05)***	.08 (.04)*	**.57 (.04)*** **	.04 (.04)	-0.10 (.05)*	-.03 (.04)
B09 Talks excessively	.33 (.05)***		.52 (.04)***				.31 (.06)***	.06 (.04)	**.55 (.04)*** **	-.00 (.05)	-0.25 (.05)***	.08 (.05)
A01 Loses temper	.52 (.05)***			.75 (.04)***			.46 (.05)***	.05 (.03)	.07 (.04)*	**.78 (.04)*** **	-0.01 (.04)	.05 (.04)
A02 Touchy / easily annoyed	.59 (.04)***			.48 (.05)***			.64 (.05)***	.07 (.04)	-.03 (.04)	**.46 (.05)*** **	-0.31 (.05)***	-.14 (.05)**
A03 Angry / resentful	.56 (.04)***			.33 (.05)***			.58 (.04)***	.00 (.05)	-.01 (.04)	**.30 (.05)*** **	-0.12 (.05)*	-.05 (.05)
A04 Argues with adults	.62 (.04)***			.33 (.05)***			.57 (.04)***	-.01 (.04)	.06 (.04)	**.39 (.04)*** **	0.08 (.05)	.08 (.04)
A05 Complies with requests	.71 (.03)***			.22 (.05)***			.67 (.04)***	.00 (.04)	-.01 (.04)	**.28 (.04)*** **	0.17 (.05)***	.11 (.04)*
A06 Annoys	.61 (.04)***			.11 (.06)			.59 (.04)***	-.20 (.05)***	.10 (.05)*	**.16 (.05)** **	0.08 (.05)	.01 (.06)
A07 Blames others	.66 (.04)***			-.02 (.05)			.58 (.04)***	.10 (.04)*	.03 (.04)	**.06 (.04)**	0.07 (.05)	.17 (.06)**

Symptoms	Bifactor CFA						Bifactor ESEM					
	G-EXT	S-IN	S-HI	S-ODD-AD	S-CD	S-CU	G-EXT	S-IN	S-HI	S-ODD-AD	S-CD	S-CU
A08 Spiteful / vindictive	.54 (.06)***			.23 (.07)**			.55 (.06)***	-.11 (.06)*	-.05 (.06)	**.23 (.07)****	0.12 (.07)	.01 (.07)
D01 Recurrent temper outbursts	.51 (.05)***			.72 (.04)***			.45 (.05)***	.01 (.04)	.07 (.04)	**.76 (.04)*****	0.17 (.04)***	.05 (.04)
D02 Persistently irritable or angry mood	.56 (.05)***			.50 (.05)***			.60 (.05)***	-.02 (.04)	-.04 (.05)	**.46 (.05)*****	0.02 (.05)	-.15 (.05)**
B01 Physical fights	.59 (.05)***				.37 (.29)		.58 (.05)***	-.13 (.05)**	.08 (.05)	.19 (.06)**	**.18 (.07)***	-.16 (.06)**
B02 Bullies, threatens, or intimidates	.64 (.05)***				.80 (.61) [a]		.65 (.05)***	-.14 (.06)*	-.01 (.05)	.20 (.06)**	**.29 (.08)*****	-.23 (.06)***
B03 Cruel to animals	.50 (.07)***				-.04 (.12)		.45 (.07)***	.01 (.09)	.07 (.08)	-.04 (.09)	**.26 (.09)****	.13 (.10)
B04 Lies	.55 (.04)***				.06 (.09)		.52 (.05)***	.06 (.05)	-.05 (.05)	.01 (.04)	**.34 (.06)*****	.15 (.06)*
B05 Steals without confrontation	.45 (.06)***				.09 (.12)		.41 (.07)***	-.06 (.07)	.08 (.06)	-.03 (.06)	**.52 (.08)*****	.05 (.08)
C01a Lack of remorse / guilt	.55 (.05)***					.47 (.05)***	.72 (.04)***	-.11 (.04)*	-.18 (.05)	-.13 (.04)**	-.02 (.06)	**.20 (.09)***
C01b Lack of concern	.55 (.04)***					.41 (.05)***	.64 (.04)***	-.08 (.05)	-.08 (.05)	-.03 (.05)	-.02 (.05)	**.25 (.07)*****
C02a Cold and uncaring	.46 (.06)***					.60 (.05)***	.63 (.06)***	-.07 (.05)	-.21 (.05)	-.10 (.05)	-.27 (.06)***	**.38 (.10)*****
C02b Self-serving	.50 (.05)***					.53 (.05)***	.57 (.05)***	-.04 (.05)	.00 (.05)	-.10 (.05)*	.03 (.05)	**.46 (.06)*****
C03a Indifferent of poor performance	.13 (.06)*					.62 (.06)***	.29 (.07)***	.12 (.06)	-.18 (.06)	-.20 (.05)***	-.09 (.06)	**.40 (.08)*****
C03b Avoids effort	.23 (.05)***					.43 (.05)***	.31 (.06)***	.13 (.05)**	-.12 (.05)	-.09 (.05)	.03 (.06)	**.32 (.06)*****
C03c Blames others for poor performance	.37 (.05)***					.34 (.06)***	.41 (.05)***	.05 (.06)**	-.04 (.05)	-.10 (.05)	.08 (.06)	**.31 (.07)*****
C04a Shallow / deficient affect	.25 (.07)***					.55 (.06)***	.23 (.07)**	.01 (.06)	-.02 (.06)	.07 (.06)	-.05 (.06)	**.64 (.06)*****
C04b Turns emotions 'on' or 'off' quickly	.35 (.06)***					.33 (.07)***	.28 (.06)***	-.02 (.06)	.11 (.06)	.06 (.06)	.13 (.06)*	**.44 (.07)*****
C04c Manipulates	.26 (.06)***					.18 (.07)*	.09 (.06)	.01 (.06)	.18 (.06)	.18 (.06)**	.21 (.07)**	**.46 (.07)*****
C04d Inconsistent affect	.43 (.06)***					.37 (.07)***	.27 (.07)***	.24 (.06)***	.12 (.07)	.18 (.06)**	-.02 (.07)	**.59 (.06)*****
ω	.95	.79	.90	.91	.77	.85	.95	.82	.92	.92	.78	.87
ω_H / ω_S	.77	.56	.59	.25	.14	.50	.73	.68	.70	.32	.23	.47
Explained common variance (%)	48						43					

Note. CFA = Confirmatory factor analysis; ESEM = Exploratory structural equation modeling; G-EXT = General factor of externalizing spectrum disorders; S = Specific factor; IN = Inattention; HI = Hyperactivity-impulsivity; ODD-AD = Oppositionality with chronic irritability/anger; CD = Conduct disorder; CU = Callous-unemotional; ω = Omega (amount of variance accounted for by the g-factor and the s-factors taken together), ωH = Omega hierarchical general (amount of variance accounted for by the g-factor), ωS = Omega hierarchical subscale (amount of variance accounted for by the s-factors). Target factor loadings are in bold.

[a] Due to a negative residual variance of item B02 Bullies, threatens, intimates within the CD factor, the Bifactor CFA model should only be interpreted with caution. *p < .05; **p < .01; ***p < .001.

Table S6

Standardized factor loadings (standard error), omega factor reliability coefficients, and explained common variance for the bifactor S-1 CFA model and bifactor S-1 ESEM solution for clinician ratings

Symptoms	Bifactor S-1 CFA					Bifactor S-1 ESEM				
	G-HI-ref	S-IN	S-ODD-AD	S-CD	S-CU	G-HI-ref	S-IN	S-ODD-AD	S-CD	S-CU
A01 Careless	.16 (.05)**	.48 (.05)***				.19 (.05)***	**.47 (.05)*** **	-.02 (.04)	-.18 (.06)**	.13 (.05)**
A02 Sustaining attention	.28 (.05)***	.41 (.05)***				.31 (.05)***	**.42 (.05)*** **	.03 (.05)	-.20 (.06)***	-.01(.05)
A03 Does not listen	.26 (.05)***	.33 (.06)***				.26 (.05)***	**.37 (.06)*** **	.07 (.05)	-.09 (.06)	-.07 (.06)
A04 Does not finish work	.28 (.05)***	.60 (.04)***				.30 (.05)***	**.60 (.04)*** **	.05 (.04)	.04 (.06)	-.03 (.05)
A05 Organizational skills	.24 (.05)***	.60 (.04)***				.25 (.05)***	**.59 (.04)*** **	.04 (.05)	.17 (.06)**	-.03 (.05)
A06 Concentration	.18 (.05)***	.43 (.05)***				.17 (.05)**	**.34 (.05)*** **	.11 (.05)*	-.11 (.06)	.16 (.05)**
A07 Loses things	.24 (.05)***	.45 (.05)***				.26 (.05)***	**.40 (.05)*** **	-.07 (.05)	.16 (.06)*	.06 (.06)
A08 Easily distracted	.42 (.05)***	.29 (.06)***				.43 (.05)***	**.41 (.06)*** **	-.01 (.05)	-.09 (.06)	-.21 (.06)***
A09 Forgetful	.26 (.05)***	.58 (.04)***				.27 (.05)***	**.52 (.04)*** **	-.01 (.05)	.07 (.05)	.11 (.05)*
B01 Fidgets	.78 (.03)***					.77 (.03)***	.06 (.04)	-.13 (.04)***	-.10 (.05)*	-.18 (.05)***
B02 Leaves seat	.73 (.03)***					.73 (.03)***	-.05 (.04)	-.09 (.04)*	.01 (.05)	-.11 (.04)*
B03 Runs / climbs	.79 (.03)***					.78 (.02)***	-.01 (.04)	-.11 (.04)**	-.01 (.05)	-.11 (.05)*
B04 Playing quietly	.63 (.04)***					.60 (.04)***	.03 (.05)	-.05 (.05)	.06 (.06)	.00 (.05)
B05 Driven / on the go	.78 (.03)***					.77 (.03)***	-.00 (.04)	-.11 (.04)**	-.01 (.05)	-.09 (.04)*
B06 Blurts out answers	.64 (.03)***					.63 (.03)***	-.01 (.04)	-.06 (.05)	-.11 (.06)	-.04 (.05)
B07 Awaiting turn	.69 (.03)***					.66 (.03)***	.00 (.04)	-.03 (.04)	-.06 (.06)	.02 (.05)
B08 Interrupts / intrudes	.76 (.03)***					.73 (.03)***	.05 (.04)	-.02 (.04)	-.04 (.05)	-.06 (.04)
B09 Talks excessively	.60 (.04)***					.59 (.04)***	.02 (.04)	-.09 (.05)	-.19 (.05)**	.02 (.05)
A01 Loses temper	.29 (.05)***		.79 (.03)***			.43 (.05)***	.01 (.03)	**.86 (.04)*** **	-.26 (.04)***	-.06 (.04)
A02 Touchy / easily annoyed	.34 (.05)***		.66 (.03)***			.42 (.05)***	.17 (.04)	**.62 (.04)*** **	.10 (.04)*	-.08 (.04)*
A03 Angry / resentful	.32 (.05)***		.56 (.04)***			.39 (.05)***	.08 (.05)	**.45 (.05)*** **	.14 (.05)**	.03 (.05)
A04 Argues with adults	.37 (.05)***		.59 (.04)***			.47 (.05)***	-.01 (.04)	**.47 (.04)*** **	.03 (.05)	.11 (.05)*
A05 Complies with requests	.39 (.04)***		.64 (.03)***			.48 (.05)***	.01 (.04)	**.40 (.04)*** **	.16 (.05)**	.21 (.05)***
A06 Annoys	.37 (.05)***		.48 (.04)***			.47 (.05)***	-.17 (.05)	**.26 (.05)*** **	.22 (.06)***	.13 (.06)*
A07 Blames others	.39 (.05)***		.46 (.04)***			.44 (.05)***	.12 (.05)	**.16 (.05)** **	.15 (.05)**	.28 (.05)***

Symptoms	Bifactor S-1 CFA					Bifactor S-1 ESEM				
	G-HI-ref	S-IN	S-ODD-AD	S-CD	S-CU	G-HI-ref	S-IN	S-ODD-AD	S-CD	S-CU
A08 Spiteful / vindictive	.26 (.07)***		.56 (.05)***			.35 (.07)***	-.07 (.06)	**.36 (.07)*** **	.22 (.07)**	.11 (.07)
D01 Recurrent temper outbursts	.29 (.05)***		.79 (.03)***			.44 (.05)***	-.06 (.03)	**.79 (.04)*** **	-.19 (.04)***	-.04 (.04)
D02 Persistently irritable or angry mood	.30 (.05)***		.67 (.03)***			.40 (.05)***	.04 (.04)	**.63 (.04)*** **	.17 (.05)**	-.10 (.05)*
B01 Physical fights	.36 (.05)***			.60 (.05)***		.46 (.05)***	-.10 (.05)	.31 (.05)***	**.32 (.07)*** **	-.06 (.06)
B02 Bullies, threatens, or intimidates	.34 (.06)***			.72 (.05)***		.45 (.06)***	-.09 (.06)	.36 (.05)***	**.44 (.07)*** **	-.08 (.06)
B03 Cruel to animals	.32 (.09)***			.41 (.08)***		.38 (.09)***	-.01 (.09)	.00 (.08)	**.21 (.12)**	.24 (.10)*
B04 Lies	.28 (.05)***			.56 (.06)***		.35 (.05)***	.05 (.05)	.10 (.05)*	**.24 (.06)*** **	.28 (.08)***
B05 Steals without confrontation	.30 (.07)***			.37 (.08)***		.37 (.07)***	-.11 (.07)	.01 (.07)	**.26 (.09)** **	.16 (.08)*
C01a Lack of remorse / guilt	.25 (.06)***				.73 (.04)***	.34 (.06)***	-0.00 (.05)	.06 (.04)	.40 (.06)***	**.47 (.06)*** **
C01b Lack of concern	.29 (.05)***				.66 (.04)***	.37 (.05)***	-0.01 (.05)	.10 (.05)*	.24 (.06)***	**.44 (.05)*** **
C02a Cold and uncaring	.15 (.06)*				.77 (.04)***	.24 (.06)***	0.05 (.05)	.09 (.05)	.20 (.06)**	**.60 (.06)*** **
C02b Self-serving	.29 (.05)***				.65 (.04)***	.38 (.05)***	-0.04 (.04)	-.04 (.05)	.09 (.05)	**.62 (.05)*** **
C03a Indifferent of poor performance	-.01 (.06)				.49 (.05)***	.05 (.06)	0.15 (.05)**	-.13 (.05)*	.05 (.06)	**.54 (.06)*** **
C03b Avoids effort	.09 (.05)				.44 (.05)***	.14 (.05)**	0.13 (.05)*	-.04 (.05)	.04 (.06)	**.43 (.05)*** **
C03c Blames others for poor performance	.20 (.06)***				.45 (.05)***	.25 (.06)***	0.05 (.06)	-.04 (.05)	.10 (.06)	**.43 (.06)*** **
C04a Shallow / deficient affect	.08 (.07)				.53 (.06)***	.16 (.07)*	-0.06 (.06)	.03 (.06)	-.30 (.06)***	**.67 (.06)*** **
C04b Turns emotions ‘on’ or ‘off’ quickly	.23 (.06)***				.41 (.06)***	.30 (.06)***	-0.09 (.06)	.01 (.06)	-.14 (.07)*	**.46 (.07)*** **
C04c Manipulates	.19 (.06)**				.25 (.06)***	.24 (.06)***	-0.13 (.06)*	.05 (.06)	-.32 (.07)***	**.38 (.07)*** **
C04d Inconsistent affect	.27 (.07)***				.50 (.06)***	.32 (.07)***	0.14 (.06)*	.11 (.07)	-.39 (.07)***	**.53 (.07)*** **
ω	.93	.78	.91	.76	.84	.94	.80	.91	.66	.88
ω_H / ω_S	.67	.60	.71	.56	.75	.73	.60	.57	.24	.73
Explained common variance (%)	40					45				

Note. CFA = Confirmatory factor analysis; ESEM = Exploratory structural equation modeling; G-HI-ref = General HI reference factor; S = Specific factor; IN = Inattention; HI = Hyperactivity-impulsivity; ODD-AD = Oppositionality with chronic irritability/anger; CD = Conduct disorder; CU = Callous-unemotional; ω = Omega (amount of variance accounted for by the g-factor and the s-factors taken together), ω_H = Omega hierarchical general (amount of variance accounted for by the g-factor), ω_S = Omega hierarchical subscale (amount of variance accounted for by the s-factors). Target factor loadings are in bold.

*p < .05; **p < .01; ***p < .001.

Table S7

Residual correlations (standard error) for the bifactor S-1 CFA model (above the diagonal) and bifactor S-1 ESEM model (below the diagonal)

Factors	1	2	3	4	5
1 IN	-	n/a	.19 (.06)**	.10 (.08)	.30 (.06)***
2 HI	n/a	-	n/a	n/a	n/a
3 ODD-AD	.01 (.06)	n/a	-	.67 (.05)***	.49 (.04)***
4 CD	-.04 (.05)	n/a	.24 (.05)***	-	.51 (.06)***
5 CU	.17 (.06)**	n/a	.27 (.05)***	.21 (.04)***	-

Note. CFA = Confirmatory factor analysis; ESEM = Exploratory structural equation modeling; IN = Inattention; HI = Hyperactivity-impulsivity; ODD-AD = Oppositionality with chronic irritability/anger; CD = Conduct disorder; CU = Callous-unemotional.
n/a = Not applicable, as correlations between the reference factor and the specific factors were restrained to zero.
**$p < .01$; ***$p < .001$.

Table S8

Consistency and specificity values for all symptoms of the bifactor S-1 CFA model

Symptom	Consistency (%)	Specificity (%)
IN		
A01 Careless	10	90
A02 Sustaining attention	33	67
A03 Does not listen	38	62
A04 Does not finish work	18	82
A05 Organizational skills	14	86
A06 Concentration	15	85
A07 Loses things	23	77
A08 Easily distracted	68	32
A09 Forgetful	16	84
ODD-AD		
A01 Loses temper	12	88
A02 Touchy / easily annoyed	21	80
A03 Angry / resentful	24	76
A04 Argues with adults	28	73
A05 Complies with requests	28	73
A06 Annoys	38	62
A07 Blames others	42	58
A08 Spiteful / vindictive	18	82
D01 Recurrent temper outbursts	12	88
D02 Persistently irritable or angry mood	16	84
CD		
B01 Physical fights	26	74
B02 Bullies, threatens, or intimidates	18	82
B03 Cruel to animals	38	62
B04 Lies	20	79
B05 Steals without confrontation	40	60
CU		
C01a Lack of remorse / guilt	10	90
C01b Lack of concern	16	84
C02a Cold and uncaring	4	96
C02b Self-serving	17	84
C03a Indifferent of poor performance	0	100
C03b Avoids effort	4	96
C03c Blames others for poor performance	17	83
C04a Shallow / deficient affect	2	98
C04b Turns emotions ‘on’ or ‘off’ quickly	23	77
C04c Manipulates	37	63
C04d Inconsistent affect	23	77

Note. Consistency = Amount of true score variance explained by the g-factor; Specificity = Amount of true score variance attributable to a specific domain; IN = Inattention; ODD-AD = Oppositionality with chronic irritability/anger; CD = Conduct disorder; CU = Callous-unemotional.

Table S9

Goodness-of-fit statistics and information criteria of alternative factor models of the externalizing spectrum for parent and teacher ratings

Model	χ^2 (*df*)	CFI	TLI	RMSEA (90% CI)	SRMR	AIC [a]	BIC [a]
				Parent Ratings			
Unidimensional	5108.822* (902)	.695	.680	.104 (.102, .107)	.128	42432.776	42968.580
First-Order CFA	1877.912* (892)	.929	.924	.051 (.048, .054)	.070	39914.301	40490.696
First-Order ESEM	1366.000* (736)	.954	.941	.045 (.041, .048)	.044	39573.697	40783.315
Higher-Order CFA	2024.815* (897)	.918	.914	.054 (.051, .057)	.078	39947.395	40503.495
Higher-Order ESEM	1353.748* (741)	.956	.943	.044 (.040, .048)	.045	39580.913	40770.236
Bifactor CFA	1813.214* (858)	.931	.924	.051 (.048, .054)	.071	39669.635	40384.040
Bifactor ESEM	1181.019* (697)	.965	.952	.040 (.036, .044)	.039	39408.960	40776.884
Bifactor S-1 CFA	1795.684* (861)	.932	.926	.050 (.047, .054)	.065	39853.487	40555.715
Bifactor S-1 ESEM	1366.000* (736)	.954	.941	.045 (.041, .048)	.044	39573.697	40783.315
				Teacher Ratings [b]			
Unidimensional	4105.506* (819)	.773	.761	.123 (.119, .126)	.144	26711.712	27163.706
First-Order CFA	1895.955* (809)	.925	.920	.071 (.067, .075)	.087	24976.090	25463.956
First-Order ESEM	1238.998* (661)	.960	.948	.057 (.052, .062)	.047	24509.472	25528.251
Higher-Order CFA	1920.344* (814)	.923	.919	.071 (.067, .075)	.092	24995.278	25465.207
Higher-Order ESEM	1217.509* (666)	.962	.951	.056 (.051, .061)	.048	24530.517	25531.360
Bifactor CFA [c]	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Bifactor ESEM	1057.467* (624)	.970	.959	.051 (.046, .056)	.041	24356.842	25508.349
Bifactor S-1 CFA	1836.212* (780)	.927	.919	.071 (.067, .075)	.079	24938.404	25530.300
Bifactor S-1 ESEM	1238.998* (661)	.960	.948	.057 (.052, .062)	.047	24509.472	25528.251

Note. df = Degrees of freedom; CFI = Comparative fit index; TLI = Tucker-Lewis index; RMSEA = Root mean square error of approximation; CI = Confidence interval; SRMR = Standardized root mean square residual; AIC = Akaike information criterion; BIC = Bayesian information criterion; CFA = Confirmatory factor analysis; ESEM = Exploratory structural equation modeling.

[a] Information criteria were calculated using *maximum likelihood estimation with robust standard errors* (MLR) for continuous indicators to ensure that AIC and BIC were comparable across CFA and ESEM models.

[b] For teacher ratings, two items (B03: *Cruel to animals*, B05: *Steals without confrontation*) were omitted due to zero variance.

[c] Model did not converge because the standard errors of the model parameter estimates could not be computed. Problem involving item B01: *Physical fights*. n = 434 (Parents); n = 267 (Teachers). *p < .001.

Table S10

Standardized factor loadings (standard error), omega factor reliability coefficients, and explained common variance for the bifactor CFA model and bifactor ESEM solution for parent ratings

Symptoms	Bifactor CFA						Bifactor ESEM					
	G-EXT	S-IN	S-HI	S-ODD-AD	S-CD	S-CU	G-EXT	S-IN	S-HI	S-ODD-AD	S-CD	S-CU
A01 Careless	.36 (.05)***	.59 (.04)***					.32 (.05)***	**.61 (.04)***	.09 (.05)*	-.02 (.04)	-.04 (.05)	.05 (.05)
A02 Sustaining attention	.41 (.05)***	.58 (.04)***					.30 (.05) ***	**.61 (.04) *****	.22 (.05)***	.04 (.05)	.06 (.06)	.05 (.05)
A03 Does not listen	.26 (.05)***	.52 (.04)***					.26 (.05)***	**.50 (.04)***	.01 (.05)	.04 (.05)	-.10 (.05)	-.08 (.06)
A04 Does not finish work	.40 (.05)***	.66 (.04)***					.33 (.05)***	**.69 (.04)***	.11 (.04)*	.02 (.04)	.02 (.05)	.06 (.054
A05 Organizational skills	.33 (.05)***	.74 (.03)***					.35 (.05)***	**.73 (.03)***	-.02 (.04)	-.07 (.04)	-.13 (.05)*	-.01 (.05)
A06 Concentration	.38 (.05)***	.59 (.04)***					.26 (.06)***	**.67 (.04)***	.09 (.05)	.16 (.06)**	.15 (.07)*	.09 (.06)
A07 Loses things	.32 (.05)***	.52 (.04)***					.36 (.05)***	**.51 (.04)***	-.01 (.05)	-.11 (.05)*	-.09 (.07)	-.09 (.05)
A08 Easily distracted	.40 (.05)***	.60 (.04)***					.31 (.06)***	**.62 (.04)***	.23 (.05)***	.02 (.05)	.00 (.05)	-.12 (.05)**
A09 Forgetful	.37 (.05)***	.68 (.03)***					.38 (.05)***	**.66 (.03)***	-.02 (.05)	-.06 (.04)	-.12 (.05)*	-.02 (.05)
B01 Fidgets	.43 (.05)***		.70 (.04)***				.38 (.05)***	.12 (.04)**	**.73 (.04)***	.07 (.04)	.07 (.06)	-.03 (.04)
B02 Leaves seat	.49 (.04)***		.61 (.04)***				.41 (.05)***	.16 (.04)***	**.65 (.04)***	.11 (.05)*	.26 (.05)***	-.06 (.06)
B03 Runs / climbs	.48 (.04)***		.65 (.03)***				.41 (.05)***	.15 (.04)***	**.68 (.04)***	.08 (.04)	.20 (.05)***	-.04 (.05)
B04 Playing quietly	.51 (.04)***		.51 (.04)***				.44 (.05)***	.17 (.04)***	**.55 (.04)***	.04 (.04)	-.02 (.06)	.09 (.05)
B05 Driven / on the go	.48 (.04)***		.65 (.03)***				.43 (.05)***	.11 (.04)**	**.69 (.04)***	.06 (.04)	-.05 (.06)	.05 (.04)
B06 Blurts out answers	.44 (.05)***		.62 (.04)***				.49 (.06)***	.02 (.04)	**.57 (.05)***	-.09 (.05)	-.21 (.05)***	-.07 (.06)
B07 Awaiting turn	.53 (.04)***		.51 (.04)***				.58 (.05)***	.02 (.04)	**.46 (.05)***	-.03 (.05)	-.16 (.07)*	-.14 (.05)*
B08 Interrupts / intrudes	.61 (.04)***		.52 (.04)***				.68 (.04)***	.03 (.04)	**.46 (.05)***	-.10 (.04)*	-.08 (.08)	-.14 (.04)**

Symptoms	Bifactor CFA						Bifactor ESEM					
	G-EXT	S-IN	S-HI	S-ODD-AD	S-CD	S-CU	G-EXT	S-IN	S-HI	S-ODD-AD	S-CD	S-CU
B09 Talks excessively	.41 (.05)***		.56 (.04)***				.48 (.06)***	.00 (.04)	**.48 (.05)*** **	-.08 (.05)	-.15 (.07)*	-.17 (.06)**
A01 Loses temper	.60 (.04)***			.71 (.04)***			.58 (.04)***	-.01 (.03)	.01 (.03)	**.72 (.04)*** **	.04 (.05)	.06 (.03)
A02 Touchy / easily annoyed	.69 (.04)***			.53 (.05)***			.73 (.04)***	.00 (.03)	.01 (.04)	**.49 (.04)*** **	.00 (.07)	-.16 (.04)***
A03 Angry / resentful	.65 (.04)***			.46 (.06)***			.68 (.04)***	.06 (.04)	.04 (.05)	**.41 (.06)*** **	-.12 (.07)	-.17 (.05)***
A04 Argues with adults	.72 (.03)***			.36 (.05)***			.68 (.03)***	.09 (.04)*	.04 (.03)	**.41 (.04)*** **	.04 (.04)	.03 (.04)
A05 Complies with requests	.74 (.03)***			.26 (.05)***			.66 (.04)***	.12 (.04)**	.08 (.04)*	**.36 (.05)*** **	.13 (.04)**	.08 (.04)
A06 Annoys	.75 (.03)***			.20 (.05)***			.76 (.03)***	-.19 (.04)***	.04 (.04)	**.21 (.05)*** **	.13 (.08)	-.01 (.04)
A07 Blames others	.75 (.03)***			-.03 (.06)			.73 (.04)***	.07 (.04)	-.06 (.04)	**-.02 (.05)**	.14 (.09)	.09 (.05)
A08 Spiteful / vindictive	.72 (.04)***			.23 (.06)***			.73 (.04)***	-.19 (.05)***	-.08 (.05)	**.23 (.06)*** **	.10 (.08)	.08 (.05)
D01 Recurrent temper outbursts	.54 (.04)***			.60 (.04)***			.48 (.05)***	-.03 (.04)	.07 (.04)	**.65 (.04)** **	.02 (.08)	.17 (.05)***
D02 Persistently irritable or angry mood	.57 (.05)***			.42 (.05)***			.56 (.04)***	.06 (.05)	-.05 (.05)	**.43 (.05)*** **	-.05 (.05)	.09 (.04)*
B01 Physical fights	.68 (.04)***				.48 (.09)***		.66 (.06)***	-.16 (.05)**	.04 (.05)	.13 (.06)*	**.36 (.09)*** **	-.04 (.07)
B02 Bullies, threatens, or intimidates	.78 (.04)***				.40 (.09)***		.76 (.05)***	-.20 (.05)***	.04 (.05)	.08 (.06)	**.31 (.11)** **	.03 (.06)
B03 Cruel to animals	.34 (.08)***				.11 (.13)		.39 (.08)***	-.16 (.07)*	-.07 (.09)	.08 (.07)	**.14 (.16)**	-.12 (.10)
B04 Lies	.74 (.03)***				-.25 (.08)**		.72 (.03)***	.11 (.04)**	-.08 (.04)	-.15 (.04)**	**.02 (.09)**	.22 (.05)***
B05 Steals without confrontation	.50 (.07)***				-.44 (.11)***		.48 (.07)***	.03 (.08)	-.05 (.08)	-.12 (.08)	**-.02 (.10)**	.20 (.07)**
C01a Lack of remorse / guilt	.52 (.04)***					.40 (.05)***	.55 (.05)***	-.04 (.05)	-.06 (.05)	.01 (.05)	.08 (.05)	**.35 (.05)*** **
C01b Lack of concern	.66 (.03)***					.39 (.05)***	.61 (.04)***	.04 (.04)	.10 (.04)**	.07 (.05)	.16 (.05)**	**.45 (.05)*** **
C02a Cold and uncaring	.56 (.05)***					.64 (.04)***	.62 (.05)***	-.22 (.04)***	.00 (.05)	-.03 (.05)	-.07 (.04)	**.57 (.05)*** **
C02b Self-serving	.58 (.04)***					.41 (.05)***	.65 (.04)***	-.09 (.04)**	-.09 (.05)	-.09 (.06)	-.02 (.06)	**.35 (.06)*** **

Symptoms	Bifactor CFA						Bifactor ESEM					
	G-EXT	S-IN	S-HI	S-ODD-AD	S-CD	S-CU	G-EXT	S-IN	S-HI	S-ODD-AD	S-CD	S-CU
C03a Indifferent of poor performance	.30 (.05)***					.63 (.04)***	.32 (.09)***	.07 (.05)	-.14 (.07)*	-.02 (.08)	.39 (.06)***	**.57 (.10)***
C03b Avoids effort	.31 (.05)***					.59 (.04)***	.31 (.09)***	.21 (.04)	-.18 (.07)*	.03 (.08)	.41 (.07)***	**.53 (.10)***
C03c Blames others for poor performance	.50 (.04)***					.36 (.05)***	.54 (.06)***	.13 (.04)	-.13 (.05)**	-.13 (.06)*	.22 (.06)***	**.30 (.07)***
C04a Shallow / deficient affect	.41 (.06)***					.73 (.045***	.43 (.07)***	-.06 (.06)	-.03 (.05)	.02 (.06)	-.20 (.07)**	**.71 (.07)***
C04b Turns emotions 'on' or 'off' quickly	.47 (.05)***					.56 (.05)***	.45 (.08)***	-.07 (.05)	.01 (.06)	.14 (.06)	-.33 (.08)***	**.57 (.09)***
C04c Manipulates	.52 (.05)***					.41 (.06)***	.52 (.08)***	.07 (.06)	-.05 (.07)	.02 (.07)	-.32 (.06)***	**.42 (.10)***
C04d Inconsistent affect	.53 (.05)***					.48 (.05)***	.53 (.07)***	.02 (.06)	-.03 (.06)	.08 (.06)	-.25 (.05)***	**.47 (.08)***
ω	.96	.90	.87	.94	.80	.92	.97	.91	.94	.95	.80	.93
ω_H / ω_S	.80	.67	.52	.22	.00	.48	.80	.74	.58	.26	.07	.44
Explained common variance (%)	51						47					

Note. CFA = Confirmatory factor analysis; ESEM = Exploratory structural equation modeling; G-EXT = General factor of externalizing spectrum disorders; S = Specific factor; IN = Inattention; HI = Hyperactivity-impulsivity; ODD-AD = Oppositionality with chronic irritability/anger; CD = Conduct disorder; CU = Callous-unemotional; ω = Omega (amount of variance accounted for by the g-factor and the s-factors taken together), ωH = Omega hierarchical general (amount of variance accounted for by the g-factor), ωS = Omega hierarchical subscale (amount of variance accounted for by the s-factors). Target factor loadings are in bold.
*p < .05; **p < .01; ***p < .001.

Table S11

Standardized factor loadings (standard error), omega factor reliability coefficients, and explained common variance for the bifactor S-1 CFA model and bifactor S-1 ESEM solution for parent ratings

Symptoms	Bifactor S-1 CFA					Bifactor S-1 ESEM				
	G-HI-ref	S-IN	S-ODD-AD	S-CD	S-CU	G-HI-ref	S-IN	S-ODD-AD	S-CD	S-CU
A01 Careless	.34 (.05)***	.60 (.04)***				.35 (.05)***	**.59 (.04)*** **	-.01 (.05)	-.02 (.05)	.06 (.05)
A02 Sustaining attention	.41 (.05)***	.57 (.04)***				.43 (.05)***	**.57 (.04)*** **	-.04 (.05)	.09 (.04)	-.03 (.05)
A03 Does not listen	.24 (.05)***	.53 (.04)***				.25 (.05)***	**.51 (.04)*** **	.11 (.05)*	-.11 (.05)*	-.06 (.05)
A04 Does not finish work	.36 (.05)***	.69 (.04)***				.38 (.05)***	**.67 (.04)*** **	.00 (.04)	.03 (.04)	.04 (.04)
A05 Organizational skills	.29 (.05)***	.76 (.03)***				.31 (.05)***	**.74 (.03)*** **	.02 (.04)	-.14 (.04)**	.07 (.05)
A06 Concentration	.33 (.05)***	.63 (.04)***				.32 (.05)***	**.64 (.04)*** **	.08 (.05)	.17 (.05)***	-.03 (.05)
A07 Loses things	.30 (.05)***	.54 (.04)***				.30 (.05)***	**.53 (.04)*** **	.04 (.05)	-.12 (.05)*	.03 (.05)
A08 Easily distracted	.45 (.05)***	.55 (.04)***				.48 (.05)***	**.57 (.04)*** **	-.03 (.05)	-.01 (.05)	-.18 (.05)***
A09 Forgetful	.32 (.05)***	.71 (.03)***				.33 (.05)***	**.69 (.04)*** **	.05 (.04)	-.14 (.04)**	.08 (.05)
B01 Fidgets	.77 (.03)***					.77 (.03)***	-.01 (.04)	-.17 (.04)***	.10 (.04)*	-.19 (.05)***
B02 Leaves seat	.78 (.03)***					.74 (.03)***	.05 (.04)	-.08 (.04)*	.26 (.04)***	-.22 (.05)***
B03 Runs / climbs	.80 (.02)***					.77 (.02)***	.04 (.04)	-.13 (.04)**	.22 (.04)***	-.19 (.04)***
B04 Playing quietly	.74 (.03)***					.71 (.03)***	.09 (.04)	-.13 (.04)**	.00 (.04)	.01 (.05)
B05 Driven / on the go	.79 (.03)***					.78 (.02)***	.00 (.03)	-.16 (.04)***	-.01 (.04)	-.08 (.04)
B06 Blurts out answers	.73 (.03)***					.74 (.03)***	-.05 (.04)	-.14 (.04)**	-.24 (.04)***	-.03 (.06)
B07 Awaiting turn	.76 (.03)***					.73 (.03)***	-.03 (.04)	.02 (.04)	-.20 (.04)***	-.06 (.06)
B08 Interrupts / intrudes	.84 (.02)***					.80 (.02)***	-.01 (.03)	-.01 (.04)	-.14 (.04)***	.00 (.05)
B09 Talks excessively	.67 (.03)***					.67 (.03)***	-.05 (.04)	-.05 (.04)	-.20 (.05)***	-.10 (.05)
A01 Loses temper	.36 (.05)***		.78 (.03)***			.49 (.04)***	.02 (.03)	**0.78 (.03)*** **	.05 (.03)	-.11 (.03)**
A02 Touchy / easily annoyed	.49 (.04)***		.69 (.03)***			.59 (.04)***	.04 (.03)	**0.71 (.04)*** **	-.01 (.04)	-.15 (.03)***
A03 Angry / resentful	.50 (.04)***		.57 (.04)***			.57 (.04)***	.10 (.03)**	**0.62 (.05)*** **	-.13 (.04)**	-.15 (.04)***
A04 Argues with adults	.48 (.04)***		.64 (.03)***			.57 (.04)***	.11 (.04)**	**0.54 (.04)*** **	.03 (.04)	.01 (.04)
A05 Complies with requests	.51 (.04)***		.59 (.03)***			.59 (.04)***	.13 (.04)**	**0.45 (.04)*** **	.13 (.04)**	.05 (.04)
A06 Annoys	.47 (.04)***		.64 (.04)***			.61 (.04)***	-.17 (.04)***	**0.41 (.05)*** **	.13 (.04)**	.12 (.04)**
A07 Blames others	.45 (.04)***		.57 (.04)***			.53 (.04)***	.10 (.04)**	**0.21 (.05)*** **	.13 (.04)**	.30 (.05)***

Symptoms	Bifactor S-1 CFA					Bifactor S-1 ESEM				
	G-HI-ref	S-IN	S-ODD-AD	S-CD	S-CU	G-HI-ref	S-IN	S-ODD-AD	S-CD	S-CU
A08 Spiteful / vindictive	.36 (.05)***		.73 (.04)***			.52 (.05)***	-.15 (.04)**	**0.45 (.06)*** **	.12 (.05)*	.21 (.05)***
D01 Recurrent temper outbursts	.31 (.05)***		.69 (.04)***			.44 (.05)***	-.03 (.04)	**0.65 (.04)*** **	.03 (.05)	-.04 (.05)
D02 Persistently irritable or angry mood	.33 (.05)***		.62 (.04)***			.42 (.05)***	.09 (.05)*	**0.54 (.05)*** **	-.02 (.05)	.06 (.05)
B01 Physical fights	.42 (.05)***			.61 (.04)***		.53 (.05)***	-.16 (.05)**	0.31 (.04)***	**.34 (.06)*** **	.08 (.05)
B02 Bullies, threatens, or intimidates	.45 (.05)***			.73 (.04)***		.61 (.05)***	-.19 (.05)***	0.29 (.06)***	**.32 (.06)*** **	.19 (.05)***
B03 Cruel to animals	.17 (.08)*			.36 (.10)***		.25 (.08)**	-.13 (.07)	0.26 (.09)**	**.12 (.13)**	.01 (.09)
B04 Lies	.42 (.05)***			.67 (.04)***		.51 (.05)***	.15 (.04)***	0.07 (.05)	**.04 (.06)**	.47 (.04)***
B05 Steals without confrontation	.26 (.07)***			.47 (.07)***		.33 (.07)***	.05 (.08)	0.02 (.08)	**.00 (.09)**	.37 (.08)***
C01a Lack of remorse / guilt	.28 (.05)***				.61 (.04)***	.38 (.05)***	-.02 (.05)	0.09 (.05)*	.08 (.05)	**.48 (.05)*** **
C01b Lack of concern	.44 (.04)***				.65 (.03)***	.54 (.04)***	.03 (.04)	0.04 (.04)	.17 (.04)***	**.50 (.04)*** **
C02a Cold and uncaring	.27 (.06)***				.79 (.03)***	.45 (.05)***	-.21 (.04)***	-0.01 (.05)	-.06 (.04)	**.74 (.04)*** **
C02b Self-serving	.32 (.05)***				.66 (.04)***	.43 (.05)***	-.05 (.04)	0.06 (.05)	.01 (.05)	**.56 (.05)*** **
C03a Indifferent of poor performance	.04 (.06)				.67 (.03)***	.16 (.06)**	.07 (.04)	-0.08 (.05)	.41 (.05)***	**.61 (.05)*** **
C03b Avoids effort	.06 (.05)				.64 (.04)***	.14 (.06)*	.21 (.05)***	-0.03 (.05)	.43 (.05)***	**.54 (.06)*** **
C03c Blames others for poor performance	.28 (.05)***				.57 (.04)***	.34 (.05)***	.16 (.05)**	0.01 (.05)	.22 (.05)***	**.48 (.05)*** **
C04a Shallow / deficient affect	.14 (.06)*				.77 (.03)***	.31 (.06)***	-.06 (.05)	-0.07 (.05)	-.17 (.05)***	**.79 (.05)*** **
C04b Turns emotions ‘on’ or ‘off’ quickly	.22 (.06)***				.68 (.04)***	.36 (.05)***	-.06 (.04)	0.07 (.05)	-.28 (.05)***	**.63 (.05)*** **
C04c Manipulates	.28 (.06)***				.61 (.04)***	.37 (.06)***	.10 (.05)*	0.07 (.06)	-.28 (.06)***	**.55 (.06)*** **
C04d Inconsistent affect	.29 (.06)***				.66 (.04)***	.39 (.06)***	.04 (.05)	0.09 (.05)	-.22 (.05)***	**.57 (.05)*** **
ω	.96	.90	.94	.80	.92	.97	.91	.95	.68	.93
ω_H / ω_S	.69	.69	.66	.59	.81	.77	.70	.52	.13	.71
Explained common variance (%)	40					46				

Note. CFA = Confirmatory factor analysis; ESEM = Exploratory structural equation modeling; G-HI-ref = General HI reference factor; S = Specific factor; IN = Inattention; HI = Hyperactivity-impulsivity; ODD-AD = Oppositionality with chronic irritability/anger; CD = Conduct disorder; CU = Callous-unemotional; ω = Omega (amount of variance accounted for by the g-factor and the s-factors taken together), ω_H = Omega hierarchical general (amount of variance accounted for by the g-factor), ω_S = Omega hierarchical subscale (amount of variance accounted for by the s-factors). Target factor loadings are in bold.
*p < .05; **p < .01; ***p < .001.

Table S12

Residual correlations (standard error) for the bifactor S-1 CFA model (above the diagonal) and bifactor S-1 ESEM model (below the diagonal) for parent ratings

Factors	1	2	3	4	5
1 IN	-	n/a	.16 (.05)**	.12 (.16)*	.19 (.05)***
2 HI	n/a	-	n/a	n/a	n/a
3 ODD-AD	-.01 (.05)	n/a	-	.71 (.04)***	.57 (.04)***
4 CD	.05 (.03)	n/a	.10 (.04)*	-	.69 (.04)***
5 CU	.06 (04)	n/a	.39 (.05)***	.09 (.05)*	-

Note. CFA = Confirmatory factor analysis; ESEM = Exploratory structural equation modeling; IN = Inattention; HI = Hyperactivity-impulsivity; ODD-AD = Oppositionality with chronic irritability/anger; CD = Conduct disorder; CU = Callous-unemotional.
n/a = Not applicable, as correlations between the reference factor and the specific factors were restrained to zero.
*p < .05; **p < .01; ***p < .001.

Table S13

Standardized factor loadings (standard error), omega factor reliability coefficients, and explained common variance for the bifactor ESEM model for teacher ratings

Symptoms	Bifactor ESEM					
	G-EXT	S-IN	S-HI	S-ODD-AD	S-CD	S-CU
A01 Careless	.33 (.06)***	**.57 (.05)*** **	.16 (.05)**	.03 (.05)	-.07 (.06)	.04 (.06)
A02 Sustaining attention	.31 (.07)***	**.67 (.04)*** **	.27 (.05)***	.10 (.06)	-.29 (.06)***	.05 (.06)
A03 Does not listen	.27 (.06)***	**.50 (.05)*** **	-.06 (.06)	-.05 (.07)	-.07 (.05)	.10 (.06)
A04 Does not finish work	.29 (.07)***	**.80 (.04)*** **	.15 (.05)***	.07 (.05)	-.22 (.05)***	.15 (.06)
A05 Organizational skills	.27 (.06)***	**.73 (.04)*** **	.01 (.05)	-.07 (.06)	.13 (.04)**	.12 (.06)
A06 Concentration	.51 (.05)***	**.52 (.05)*** **	-.07 (.05)	.15 (.05)**	-.18 (.05)***	.11 (.05)
A07 Loses things	.30 (.07)***	**.66 (.05)*** **	-.08 (.06)	-.17 (.07)*	.44 (.06)***	-.06 (.06)
A08 Easily distracted	.36 (.07)***	**.55 (.06)*** **	.33 (.06)***	.01 (.07)	-.31 (.05)***	-.03 (.07)
A09 Forgetful	.31 (.07)***	**.69 (.05)*** **	-.10 (.06)	-.16 (.06)	.40 (.05)***	-.07 (.06)
B01 Fidgets	.42 (.06)***	.11 (.05)*	**.66 (.04)*** **	.05 (.05)	.14 (.05)**	.02 (.05)
B02 Leaves seat	.41 (.06)***	.19 (.05)***	**.81 (.03)*** **	.14 (.04)**	.08 (.05)	.11 (.04)
B03 Runs / climbs	.40 (.06)***	.18 (.04)***	**.77 (.04)*** **	.15 (.04)***	.15 (.06)**	.15 (.04)
B04 Playing quietly	.54 (.06)***	.11 (.05)*	**.52 (.05)*** **	.00 (.05)	.08 (.05)	-.12 (.06)
B05 Driven / on the go	.54 (.05)***	.08 (.05)	**.62 (.04)*** **	.08 (.04)**	.09 (.04)*	-.01 (.04)
B06 Blurts out answers	.65 (.05)***	-.07 (.04)	**.59 (.06)*** **	-.16 (.04)***	-.16 (.06)**	-.18 (.04)
B07 Awaiting turn	.72 (.05)***	-.07 (.05)	**.51 (.06)*** **	-.14 (.04)**	-.16 (.06)**	-.19 (.04)
B08 Interrupts / intrudes	.70 (.04)***	-.02 (.04)	**.50 (.05)*** **	-.10 (.03)**	-.04 (.05)	-.20 (.04)
B09 Talks excessively	.62 (.05)***	.01 (.05)	**.50 (.05)*** **	.00 (.05)	-.14 (.05)**	-.18 (.05)
A01 Loses temper	.67 (.05)***	-.04 (.05)	.07 (.04)	**.60 (.05)*** **	.08 (.04)*	-.03 (.04)
A02 Touchy / easily annoyed	.78 (.04)***	-.10 (.04)**	.03 (.04)	**.52 (.05)*** **	.01 (.03)	-.16 (.03)
A03 Angry / resentful	.68 (.04)***	-.04 (.04)	-.09 (.04)*	**.59 (.05)** **	-.06 (.04)	-.04 (.04)
A04 Argues with adults	.83 (.03)***	.08 (.05)	-.06 (.04)	**.20 (.06)** **	-.04 (.04)	.07 (.05)
A05 Complies with requests	.84 (.03)***	.04 (.04)	.06 (.03)	**.16 (.06)**	.02 (.04)	.24 (.04)
A06 Annoys	.84 (.02)***	-.15 (.04)***	.09 (.04)*	**-.01 (.06)**	.14 (.05)**	.01 (.04)
A07 Blames others	.85 (.02)***	-.07 (.05)	-.07 (.04)	**.04 (.06)**	.00 (.05)	-.02 (.04)
A08 Spiteful / vindictive	.83 (.03)***	-.23 (.05)***	-.06 (.05)	**.11 (.07)**	.14 (.06)*	.02 (.05)
D01 Recurrent temper outbursts	.68 (.05)***	-.03 (.06)	.21 (.05)***	**.47 (.06)*** **	.15 (.05)**	.10 (.06)
D02 Persistently irritable or angry mood	.59 (.06)***	.14 (.06)*	.00 (.07)	**.30 (.06)*** **	-.21 (.07)**	.11 (.06)
B01 Physical fights	.75 (.04)***	-.08 (.04)	.21 (.05)***	.10 (.05)*	**.38 (.06)*** **	.01 (.05)
B02 Bullies, threatens, or intimidates	.81 (.03)***	-.06 (.05)	.12 (.05)*	-.01 (.05)	**.28 (.06)*** **	.02 (.05)
B04 Lies	.73 (.04)***	.01 (.05)	-.18 (.05)**	-.10 (.06)	**.04 (.07)**	.12 (.06)
C01a Lack of remorse / guilt	.69 (.04)***	-.10 (.04)	-.04 (.05)	-.04 (.06)	.06 (.05)	**.39 (.05)**
C01b Lack of concern	.73 (.04)***	.01 (.04)	.06 (.04)	-.09 (.04)	.10 (.04)*	**.49 (.05)**
C02a Cold and uncaring	.64 (.05)***	.01 (.04)	.02 (.05)	-.02 (.04)	.05 (.04)	**.63 (.05)**
C02b Self-serving	.78 (.03)***	-.13 (.05)**	-.17 (.05)**	-.03 (.06)	-.07 (.05)	**.24 (.06)**
C03a Indifferent of poor performance	.46 (.06)***	.21 (.05)***	-.19 (.05)***	-.10 (.06)	-.16 (.05)**	**.55 (.06)**
C03b Avoids effort	.47 (.06)***	.43 (.05)***	-.21 (.05)***	-.01 (.06)	-.31 (.05)***	**.46 (.06)**
C03c Blames others for poor performance	.77 (.04)***	.11 (.06)	-.16 (.06)**	.05 (.06)	-.14 (.06)*	**.08 (.07)**
C04a Shallow / deficient affect	.48 (.06)***	.00 (.05)	-.06 (.05)	-.11 (.05)*	.05 (.05)	**.73 (.05)**
C04b Turns emotions 'on' or 'off' quickly	.44 (.06)***	-.08 (.06)	.07 (.06)	.16 (.07)*	.01 (.07)	**.48 (.06)**
C04c Manipulates	.59 (.06)***	-.01 (.05)	.00 (.07)	.27 (.08)**	.14 (.07)*	**.38 (.07)**
C04d Inconsistent affect	.50 (.06)***	.15 (.06)*	-.05 (.06)	.13 (.08)	.21 (.07)**	**.32 (.07)**
ω	.98	.94	.96	.97	.87	.95

Symptoms	Bifactor ESEM					
	G-EXT	S-IN	S-HI	S-ODD-AD	S-CD	S-CU
ω_H / ω_S	.83	.77	.54	.14	.08	.35
Explained common variance (%)	53					

Note. CFA = Confirmatory factor analysis; G-EXT = General factor of externalizing spectrum disorders; S = Specific factor; IN = Inattention; HI = Hyperactivity-impulsivity; ODD-AD = Oppositionality with chronic irritability/anger; CD = Conduct disorder; CU = Callous-unemotional; ω = Omega (amount of variance accounted for by the g-factor and the s-factors taken together), ω_H = Omega hierarchical general (amount of variance accounted for by the g-factor), ω_S = Omega hierarchical subscale (amount of variance accounted for by the s-factors). **p* < .05; ***p* < .01; ****p* < .001.

Table S14

Standardized factor loadings (standard error), omega factor reliability coefficients, and explained common variance for the bifactor S-1 CFA model and bifactor S-1 ESEM solution for teacher ratings

Symptoms	Bifactor S-1 CFA					Bifactor S-1 ESEM				
	G-HI-ref	S-IN	SODD-AD	S-CD	S-CU	G-HI-ref	S-IN	S-ODD-AD	S-CD	S-CU
A01 Careless	.34 (.06)***	.57 (.05)***				.38 (.06)***	**.56 (.05)*** **	.01 (.05)	-.08 (.05)	-.01 (.06)
A02 Sustaining attention	.37 (.06)***	.66 (.04)***				.41 (.06)***	**.63 (.05)*** **	.05 (.05)	-.33 (.05)***	-.05 (.05)
A03 Does not listen	.17 (.06)**	.57 (.05)***				.19 (.06)**	**.51 (.05)*** **	.01 (.07)	.03 (.05)	.12 (.06)
A04 Does not finish work	.30 (.06)***	.81 (.03)***				.34 (.06)***	**.76 (.04)*** **	.01 (.05)	-.28 (.04)***	.07 (.06)
A05 Organizational skills	.19 (.07)**	.75 (.03)***				.26 (.06)***	**.73 (.04)*** **	-.11 (.06)	.09 (.04)	.09 (.06)
A06 Concentration	.44 (.06)***	.64 (.04)***				.39 (.06)***	**.54 (.05)*** **	.26 (.05)***	-.08 (.04)	.12 (.05)*
A07 Loses things	.15 (.07)*	.75 (.03)***				.24 (.07)***	**.73 (.05)*** **	-.13 (.06)*	.45 (.05)***	-.05 (.05)
A08 Easily distracted	.45 (.06)***	.52 (.05)***				.47 (.06)***	**.52 (.06)*** **	.02 (.07)	-.26 (.05)***	-.09 (.06)
A09 Forgetful	.15 (.07)*	.77 (.03)***				.23 (.07)***	**.77 (.05)*** **	-.11 (.05)	.44 (.04)***	-.05 (.05)
B01 Fidgets	.74 (.04)***					.73 (.04)***	.02 (.05)	-.18 (.05)**	-.17 (.04)***	-.07 (.06)
B02 Leaves seat	.88 (.02)***					.81 (.04)***	.04 (.05)	-.24 (.04)***	-.38 (.04)***	-.02 (.04)
B03 Runs / climbs	.86 (.02)***					.79 (.04)***	.04 (.04)	-.23 (.04)***	-.34 (.04)***	.01 (.05)
B04 Playing quietly	.76 (.04)***					.74 (.03)***	.08 (.05)	-.04 (.06)	-.02 (.06)	-.15 (.06)*
B05 Driven / on the go	.83 (.03)***					.81 (.03)***	.02 (.05)	-.07 (.04)	-.15 (.04)***	-.08 (.05)
B06 Blurts out answers	.88 (.02)***					.86 (.02)***	-.04 (.04)	-.13 (.04)**	.18 (.04)***	-.20 (.05)***
B07 Awaiting turn	.92 (.02)***					.88 (.02)***	-.02 (.04)	-.06 (.04)	.21 (.04)***	-.18 (.05)***
B08 Interrupts / intrudes	.89 (.02)***					.86 (.02)***	.00 (.04)	-.02 (.04)	.14 (.04)*	-.18 (.04)***
B09 Talks excessively	.82 (.03)***					.79 (.03)***	.03 (.05)	.04 (.05)	.05 (.04)	-.21 (.05)***
A01 Loses temper	.50 (.05)***		.67 (.04)***			.62 (.05)***	-.04 (.04)	**.69 (.05)*** **	-.16 (.04)***	-.10 (.04)*
A02 Touchy / easily annoyed	.55 (.05)***		.71 (.04)***			.67 (.04)***	-.04 (.03)	**.73 (.04)*** **	.00 (.03)	-.19 (.04)***
A03 Angry / resentful	.41 (.06)***		.76 (.03)***			.53 (.05)***	.01 (.03)	**.77 (.04)*** **	-.08 (.04)*	-.09 (.04)
A04 Argues with adults	.54 (.05)***		.66 (.04)***			.65 (.05)***	.13 (.05)**	**.42 (.05)*** **	.12 (.05)*	.15 (.05)**
A05 Complies with requests	.59 (.05)***		.69 (.04)***			.73 (.04)***	.05 (.04)	**.26 (.05)*** **	.06 (.05)	.33 (.04)***
A06 Annoys	.63 (.04)***		.56 (.05)***			.74 (.04)***	-.12 (.04)**	**.21 (.06)*** **	.25 (.05)***	.16 (.04)***
A07 Blames others	.55 (.05)***		.63 (.04)***			.65 (.04)***	.01 (.05)	**.34 (.06)*** **	.28 (.04)***	.14 (.04)**

Symptoms	Bifactor S-1 CFA					Bifactor S-1 ESEM				
	G-HI-ref	S-IN	SODD-AD	S-CD	S-CU	G-HI-ref	S-IN	S-ODD-AD	S-CD	S-CU
A08 Spiteful / vindictive	.50 (.06)***		.70 (.05)***			.65 (.05)***	-.18 (.05)***	**.37 (.06)*** **	.23 (.06)***	.19 (.05)**
D01 Recurrent temper outbursts	.58 (.06)***		.57 (.05)***			.70 (.05)***	-.07 (.06)	**.47 (.07)*** **	-.17 (.06)**	.04 (.06)
D02 Persistently irritable or angry mood	.42 (.07)***		.50 (.06)***			.49 (.06)***	.16 (.06)**	**.40 (.06)*** **	-.12 (.06)	.11 (.07)
B01 Physical fights	.65 (.04)***			.54 (.05)***		.76 (.04)***	-.09 (.04)*	.18 (.65)**	**.17 (.05)** **	.09 (.05)
B02 Bullies, threatens, or intimidates	.63 (.05)***			.60 (.06)***		.74 (.04)***	-.04 (.05)	.16 (.05)**	**.26 (.06)*** **	.15 (.06)**
B04 Lies	.36 (.06)***			.74 (.05)***		.49 (.06)***	.07 (.05)	.19 (.06)**	**.29 (.06)*** **	.32 (.06)***
C01a Lack of remorse / guilt	.41 (.06)***				.67 (.04)***	.54 (.05)***	-.11 (.04)	.05 (.05)	.14 (.06)*	**.55 (.05)*** **
C01b Lack of concern	.48 (.05)***				.73 (.04)***	.63 (.05)***	-.03 (.04)	-.08 (.05)	.09 (.04)*	**.64 (.05)*** **
C02a Cold and uncaring	.35 (.06)***				.80 (.03)***	.54 (.05)***	-.06 (.04)	-.09 (.05)	-.02 (.05)	**.76 (.05)*** **
C02b Self-serving	.44 (.06)***				.68 (.04)***	.53 (.06)***	-.07 (.05)	.23 (.05)***	.22 (.05)***	**.43 (.05)*** **
C03a Indifferent of poor performance	.11 (.07)				.74 (.03)***	.25 (.07)***	.18 (.05)***	-.03 (.06)	-.07 (.06)	**.70 (.05)*** **
C03b Avoids effort	.16 (.07)*				.74 (.04)***	.26 (.07)***	.42 (.05)***	.09 (.06)	-.15 (.05)**	**.57 (.05)*** **
C03c Blames others for poor performance	.48 (.06)***				.62 (.05)***	.53 (.06)***	.17 (.06)**	.33 (.06)***	0.18 (.06)**	**.23 (.06)*** **
C04a Shallow / deficient affect	.16 (.07)*				.79 (.04)***	.36 (.07)***	-.08 (.04)	-.22 (.05)***	-.02 (.05)	**.88 (.05)*** **
C04b Turns emotions ‘on’ or ‘off’ quickly	.29 (.07)***				.52 (.05)***	.41 (.07)***	-.15 (.06)*	.03 (.06)	-.14 (.07)*	**.51 (.07)*** **
C04c Manipulates	.40 (.07)***				.61 (.05)***	.51 (.06)***	-.05 (.06)	.23 (.09)**	-.07 (.07)	**.43 (.07)*** **
C04d Inconsistent affect	.31 (.07)***				.55 (.05)***	.41 (.06)***	.13 (.06)*	.11 (.08)	.05 (.07)	**.37 (.08)*** **
ω	.97	.92	.96	.88	.94	.98	.93	.96	.79	.94
ω_H / ω_S	.74	.78	.58	.50	.76	.83	.76	.36	.12	.59
Explained common variance (%)	46					53				

Note. CFA = Confirmatory factor analysis; ESEM = Exploratory structural equation modeling; G-HI-ref = General HI reference factor; S = Specific factor; IN = Inattention; HI = Hyperactivity-impulsivity; ODD-AD = Oppositionality with chronic irritability/anger; CD = Conduct disorder; CU = Callous-unemotional; ω = Omega (amount of variance accounted for by the g-factor and the s-factors taken together), ω_H = Omega hierarchical general (amount of variance accounted for by the g-factor), ω_S = Omega hierarchical subscale (amount of variance accounted for by the s-factors). Target factor loadings are in bold.

*p < .05; **p < .01; ***p < .001.

Table S15

Residual correlations (standard error) for the bifactor S-1 CFA model (above the diagonal) and bifactor S-1 ESEM model (below the diagonal) for teacher ratings

Factors	1	2	3	4	5
1 IN	-	n/a	.17 (.06)**	.21 (.07)**	.38 (.06)***
2 HI	n/a	-	n/a	n/a	n/a
3 ODD-AD	.03 (.05)	n/a	-	.77 (.04)***	.69 (.04)***
4 CD	-.01 (.04)**	n/a	.15 (.04)***	-	.70 (.05)***
5 CU	.20 (.05)***	n/a	.40 (.05)***	.14 (.04)***	-

Note. CFA = Confirmatory factor analysis; ESEM = Exploratory structural equation modeling; IN = Inattention; HI = Hyperactivity-impulsivity; ODD-AD = Oppositionality with chronic irritability/anger; CD = Conduct disorder; CU = Callous-unemotional.
n/a = Not applicable, as correlations between the reference factor and the specific factors were restrained to zero.
**$p < .01$; ***$p < .001$.

7.3 Supplementary material for:

Identifying symptoms of ADHD and disruptive behavior disorders most strongly associated with functional impairment in children: A symptom-level approach.

Table S1

Scale characteristics: Cronbach's alpha, item-total correlations, and scale correlations

Scale	Number of Items	Cronbach's α	Range of Item-Total Correlations	Mean (*SD*)	*n*	Scale Correlations						
ADHD Scales						1	2	3	4	5	6	7
1 Inattention	9	.71	.29 - .49	1.95 (0.48)	474		.35	.42	.30	.21	.30	.24
2 Hyperactivity-Impulsivity	9	.87	50 - .68	1.64 (0.73)	474		-	.45	.42	.32	.25	.31
3 ADHD-Related Functional Impairment	5	.62	.24 - .48	1.61 (0.59)	472			-	.46	.35	.26	.53
ODD/CD/CU Scales												
4 Oppositionality with Chronic Irritability / Anger	10	.86	.43 - .66	1.05 (0.62)	452				-	.57	.47	.61
5 Conduct Disorder Symptoms	5	.60	.22 - .47	0.40 (0.43)	450					-	.39	.56
6 Callous-Unemotional Symptoms	11	.77	.23 - .58	0.50 (0.42)	446						-	.43
7 ODD/CD/CU-Related Functional Impairment	5	.86	.61 - .71	0.93 (0.78)	442							-

Note. Symptoms of ADHD, ODD, CD, and CU were assessed using the semi-structured *Clinical Parent Interview for Externalizing Disorders in Children and Adolescents* (ILF-EXTERNAL) conducted with the parents. ADHD = attention-deficit/hyperactivity disorder; ODD = oppositional defiant disorder; CD = conduct disorder; CU = callous-unemotional. All Pearson correlation coefficients significant ($p < .001$, not adjusted).

Table S2

Item characteristics: Skewness, kurtosis, and variance inflation factor

Item	Mean (*SD*)	Skewness	Kurtosis	VIF	*n*
	ADHD Symptoms				
A01 Careless	2.03 (0.77)	-0.52	-0.00	1.25	474
A02 Difficulties Sustaining Attention	2.23 (0.64)	-0.45	0.19	1.31	474
A03 Does not Listen	1.96 (0.93)	-0.56	-0.56	1.25	474
A04 Does not Finish Work	2.09 (0.80)	-0.73	0.25	1.38	474
A05 Organizational Skills	2.07 (0.86)	-0.75	0.02	1.43	474
A06 Concentration	1.85 (0.87)	-0.47	-0.39	1.20	474
A07 Loses Things	1.39 (1.11)	0.10	-1.34	1.34	474
A08 Easily Distracted	2.48 (0.63)	-1.03	0.91	1.40	474
A09 Forgetful	1.49 (1.03)	-0.10	-1.15	1.35	474
B01 Fidgets	2.08 (0.99)	-0.78	-0.51	2.05	474
B02 Leaves Seat	1.72 (1.01)	-0.34	-0.97	1.87	474
B03 Runs, Climbs	1.51 (1.08)	-0.08	-1.25	2.12	474
B04 Playing Quietly	1.14 (1.01)	0.32	-1.12	1.51	474
B05 Driven, on the go	1.51 (1.14)	-0.07	-1.41	2.00	474
B06 Blurts out Answers	1.72 (1.08)	-0.35	-1.14	1.54	474
B07 Waits for Turn	1.46 (1.06)	-0.05	-1.23	1.56	474
B08 Interrupts, Intrudes	1.93 (0.92)	-0.59	-0.43	1.84	474
B09 Talks Excessively	1.69 (1.14)	-0.28	-1.34	1.50	474
F01 Psychological Strain	1.63 (0.90)	-0.34	-0.62	1.23	473
F02 Functional Impairment Related to Home Life and Family Members	1.84 (0.86)	-0.47	-0.31	1.29	473
F03 Functional Impairment Related to Relationships with Adults	1.26 (0.97)	0.12	-1.07	1.35	473
F04 Functional Impairment Related to Relationships with Children/Adolescents; Recreational Activities	1.31 (1.05)	0.16	-1.20	1.55	472
F05 Functional Impairment Related to Academic Performance	2.02 (0.85)	-0.69	0.00	1.22	473
	ODD/CD/CU Symptoms				
A01 Loses Temper	1.47 (1.01)	-0.11	-1.09	2.65	452
A02 Touchy, Easily Annoyed	1.31 (1.01)	0.11	-1.14	2.06	452
A03 Angry, Resentful	1.15 (0.93)	0.23	-0.99	1.53	452
A04 Argues with Adults	1.22 (0.98)	0.19	-1.07	1.93	452
A05 Refuses to Comply with Requests	1.47 (1.02)	0.03	-1.10	1.97	452
A06 Annoys	0.87 (0.93)	0.77	-0.43	1.58	452
A07 Blames Others	1.21 (1.01)	0.22	-1.15	1.53	452
A08 Spiteful, Vindictive	0.30 (0.68)	2.45	5.53	1.46	452
D01 Recurrent Temper Outbursts	0.87 (0.95)	0.77	-0.49	2.75	452
D02 Persistently Irritable or Angry Mood	0.59 (0.85)	1.22	0.41	1.76	452
B01 Physical Fights	0.47 (0.75)	1.55	1.78	1.65	451
B02 Bullies, Threatens, or Intimidates	0.34 (0.71)	2.13	3.71	1.76	451
B03 Cruel to Animals	0.09 (0.37)	4.43	21.33	1.23	451
B04 Lies	0.89 (0.93)	0.70	-0.54	1.53	451
B05 Steals Without Confrontation	0.22 (0.55)	2.59	6.11	1.25	450
C01a Lack of Remorse	0.55 (0.80)	1.27	0.58	1.86	447
C01b Lack of Concern	0.68 (0.84)	1.02	0.14	1.74	447

Item	Mean (*SD*)	Skewness	Kurtosis	VIF	*n*
C02a Cold and Uncaring	0.35 (0.70)	2.03	3.49	1.79	446
C02b Self-Serving	0.57 (0.81)	1.18	0.37	1.70	447
C03a Indifferent to Poor Performance	0.46 (0.77)	1.62	1.84	1.40	447
C03b Avoids Effort	0.88 (0.89)	0.65	-0.55	1.38	447
C03c Blames Others for Poor Performance	0.63 (0.86)	1.14	0.24	1.34	446
C04a Shallow, Deficient Affect	0.32 (0.65)	2.04	3.60	1.39	446
C04b Turns Emotions 'On' or 'Off' Quickly	0.41 (0.72)	1.81	2.74	1.31	446
C04c Manipulates	0.41 (0.67)	1.61	2.05	1.34	446
C04d Inconsistent Affect	0.24 (0.60)	2.75	7.51	1.31	447
F01 Psychological Strain	0.96 (0.94)	0.36	-1.21	1.93	446
F02 Functional Impairment Related to Home Life and Family Members	1.33 (1.04)	-0.01	-1.27	2.25	443
F03 Functional Impairment Related to Relationships with Adults	0.75 (0.91)	0.86	-0.43	2.49	444
F04 Functional Impairment Related to Relationships with Children/Adolescents; Recreational Activities	0.85 (0.98)	0.73	-0.76	2.44	445
F05 Functional Impairment Related to Academic Performance	0.76 (1.10)	0.94	-0.47	2.19	445

Note. Each symptom was rated by clinicians using the DSM-5 based, semi-structured *Clinical Parent Interview for Externalizing Disorders in Children and Adolescents* (ILF-EXTERNAL) conducted with the parents. VIF = Variance inflation factor; ADHD = attention-deficit/hyperactivity disorder; ODD = oppositional defiant disorder; CD = conduct disorder; CU = callous-unemotional.

Table S3

Multivariate linear regression analysis predicting global functional impairment from individual ADHD symptoms

Predictors	Estimate	*SE*	95% CI	*t*	*p*
A01 Careless	0.02	0.03	-.04, .09	0.68	.498
A02 Difficulties Sustaining Attention	0.11	0.04	.04, .19	2.92**	.004
A03 Does not Listen	0.05	0.03	.00, .10	1.96	.051
A04 Does not Finish Work	0.02	0.03	-.04, .09	0.67	.051
A05 Organizational Skills	0.02	0.03	-.04, .08	0.71	.480
A06 Concentration	0.05	0.03	.00, .11	1.97*	.049
A07 Loses Things	-0.02	0.02	-.07, .02	-0.90	.367
A08 Easily Distracted	0.18	0.04	.10, .26	4.41***	< .001
A09 Forgetful	0.04	0.03	-.01, .09	1.47	.141
B01 Fidgets	-0.07	0.03	-.14, -.01	-2.31*	.021
B02 Leaves Seat	0.06	0.03	.00, .12	2.06*	.040
B03 Runs, Climbs	0.07	0.03	.01, .13	2.44*	.015
B04 Playing Quietly	0.01	0.03	-.04, .06	0.35	.727
B05 Driven, on the go	0.04	0.03	-.01, .09	1.49	.138
B06 Blurts out Answers	0.05	0.03	.00, .10	1.81	.071
B07 Waits for turn	-0.02	0.03	-.07, .04	-0.60	.550
B08 Interrupts, Intrudes	0.10	0.03	.03, .16	2.98**	.003
B09 Talks excessively	0.01	0.02	-.04, .05	0.36	.719
Age	0.02	0.02	-.01, .05	1.12	.260
Gender	0.13	0.06	.02, .25	2.25*	.025

Note. Significant predictors are shaded in gray. $F = 12.69$, $df = 451$, $p < .001$, $R^2 = .36$; ADHD = attention-deficit/hyperactivity disorder.
* $p < .05$; ** $p < .01$; *** $p < .001$.

Table S4

Multivariate linear regression analysis predicting global functional impairment from individual ODD/CD/CU symptoms

Predictors	Estimate	*SE*	95% CI	*t*	*p*
A01 Loses Temper	0.12	0.04	.03, .20	2.63**	.009
A02 Touchy, Easily Annoyed	0.02	0.04	-.06, .20	0.41	.679
A03 Angry, Resentful	0.08	0.04	.01, .16	2.31*	.022
A04 Argues with Adults	0.13	0.04	.06, .20	3.51***	< .001
A05 Refuses to Comply	0.06	0.04	-.02, .13	1.56	.119
A06 Annoys	0.08	0.04	.02, .15	2.39*	.018
A07 Blames Others	0.06	0.03	-.01, .12	1.74	.082
A08 Spiteful, Vindictive	-0.15	0.05	-.24, -.06	-3.18**	.002
D01 Recurrent Temper Outbursts	-0.10	0.05	-.19, -.01	-2.16*	.032
D02 Persistently Irritable or Angry Mood	0.07	0.04	-.02, .15	1.49	.134
B01 Physical Fights	0.17	0.05	.08, .26	3.68***	< .001
B02 Bullies, Threatens, or Intimidates	0.08	0.05	-.13, .18	1.70	.091
B03 Cruel to Animals	0.14	0.08	-.01, .30	1.79	.074
B04 Lies	0.06	0.04	-.01, .13	1.74	.084
B05 Steals Without Confrontation	0.11	0.05	.00, .21	1.99*	.047
C01a Lack of Remorse	0.04	0.05	-.05, .13	0.89	.377
C01b Lack of Concern	0.05	0.04	-.03, .13	1.17	.244
C02a Cold and Uncaring	-0.05	0.05	-.14, .05	-0.92	.356
C02b Self-Serving	0.06	0.04	-.01, .15	1.53	.127
C03a Indifferent to Poor Performance	-0.07	0.04	-.15, .01	-1.67	.096
C03b Avoids Effort	0.08	0.04	.01, .14	2.20*	.028
C03c Blames Others for Poor Performance	0.04	0.04	-.03, .11	1.07	.283
C04a Shallow, Deficient Affect	-0.06	0.05	-.16, .03	-1.31	.189
C04b Turns Emotions ‘On’ or ‘Off’ Quickly	0.05	0.04	-.03, .14	1.38	.169
C04c Manipulates	-0.02	0.05	-.10, .08	-0.19	.853
C04d. Inconsistent Affect	0.12	0.05	.02, .22	2.26*	.024
Age	0.02	0.02	-.02, .05	0.86	.389
Gender	-0.01	0.07	-.15, .14	-0.08	.938

Note. Significant predictors are shaded in gray. $F = 16.06$, $df = 411$, $p < .001$, $R^2 = .52$; ODD = oppositional defiant disorder; CD = conduct disorder; CU = callous-unemotional. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table S5

Multivariate linear regression analysis predicting global functional impairment from all externalizing symptoms

Predictors	Estimate	*SE*	95% CI	*t*	*p*
A01 Careless	-0.01	0.03	-.06, .05	-0.24	.810
A02 Difficulties Sustaining Attention	0.06	0.04	-.01, .12	1.51	.131
A03 Does not Listen	0.03	0.03	-.02, .08	1.08	.280
A04 Does not Finish Work	0.03	0.03	-.03, .09	1.11	.268
A05 Organizational Skills	0.01	0.023	-.04, .07	0.45	.655
A06 Concentration	-0.01	0.03	-.06, .04	-0.28	.776
A07 Loses Things	-0.03	0.02	-.01, .01	-1.30	.195
A08 Easily Distracted	0.13	0.04	.06, .20	3.34***	< .001
A09 Forgetful	0.01	0.02	-.04, .05	0.24	.808
B01 Fidgets	-0.02	0.03	-.08, .04	-0.66	.513
B02 Leaves Seat	0.00	0.03	-.05, .06	0.15	.878
B03 Runs, Climbs	0.06	0.03	.01, .11	2.17*	.030
B04 Playing Quietly	0.01	0.03	-.04, .06	0.36	.716
B05 Driven, on the go	0.00	0.03	-.05, .05	0.11	.914
B06 Blurts out Answers	-0.03	0.02	-.08, .02	-1.18	.241
B07 Waits for turn	-0.03	0.02	-.08, .02	-1.15	.249
B08 Interrupts, Intrudes	0.05	0.03	-.01, .11	1.58	.116
B09 Talks excessively	-0.01	0.02	-.05, .04	-0.23	.822
A01 Loses Temper	0.07	0.03	.01, .14	2.14*	.033
A02 Touchy, Easily Annoyed	0.02	0.03	-.04, .08	0.53	.595
A03 Angry, Resentful	0.08	0.03	.03, .14	3.00**	.003
A04 Argues with Adults	0.09	0.03	.04, .15	3.23**	.001
A05 Refuses to Comply	0.04	0.03	-.02, .09	1.42	.157
A06 Annoys	0.10	0.03	.05, .15	3.69***	< .001
A07 Blames Others	0.03	0.03	-.02, .07	0.99	.322
A08 Spiteful, Vindictive	-0.10	0.04	-.17, -.03	-2.92**	.004
D01 Recurrent Temper Outbursts	-0.07	0.04	-.14, .00	-1.86	.064
D02 Persistently Irritable or Angry Mood	0.05	0.03	-.02, .11	1.37	.172
B01 Physical Fights	0.14	0.03	.07, .20	3.99***	< .001
B02 Bullies, Threatens, or Intimidates	0.03	0.04	-.04, .11	0.87	.385
B03 Cruel to Animals	0.08	0.06	-.04, .20	1.36	.174
B04 Lies	0.03	0.03	-.02, .09	1.29	.198
B05 Steals Without Confrontation	0.06	0.04	-.02, .14	1.57	.118
C01a Lack of Remorse	-0.01	0.04	-.08, .06	-0.36	.721
C01b Lack of Concern	0.02	0.03	-.05, .08	0.47	.639
C02a Cold and Uncaring	0.03	0.04	-.05, .10	0.70	.485
C02b Self-Serving	0.02	0.03	-.05, .08	0.52	.606
C03a Indifferent to Poor Performance	-0.03	0.03	-.10, .03	-1.08	.283
C03b Avoids Effort	0.08	0.03	.03, .13	2.89**	.004
C03c Blames Others for Poor Performance	0.02	0.03	-.04, .07	0.61	.543
C04a Shallow, Deficient Affect	-0.04	0.04	-.12, .03	-1.17	.242
C04b Turns Emotions ‘On’ or ‘Off’ Quickly	0.02	0.03	-.04, .09	0.74	.461
C04c Manipulates	-0.03	0.04	-.10, .04	-0.85	.396
C04d Inconsistent Affect	0.08	0.04	.00, .16	1.94	.053
Age	0.02	0.02	-.01, .05	1.39	.165
Gender	0.03	0.06	-.08, .14	0.49	.627

Note. Significant predictors are shaded in gray. $F = 11.40$, $df = 397$, $p < .001$, $R^2 = .57$.
* $p < .05$; ** $p < .01$; *** $p < .001$.

Table S6

Ordinal logistic regression analysis predicting functional impairment domains from individual ADHD symptoms

Predictors	Estimate	*SE*	*z*	*p*	*OR*	*95% CI*
F01 Psychological Strain						
A01 Careless	0.25	0.13	1.95	.052	1.28	0.99, 1.65
A02 Difficulties Sustaining Attention	0.39	0.16	2.49*	.013	1.48	1.09, 2.02
A03 Does not Listen	0.04	0.11	0.38	.705	1.04	0.85, 1.28
A04 Does not Finish Work	0.12	0.13	0.93	.350	1.13	0.88, 1.45
A05 Organizational Skills	0.07	0.12	0.58	.562	1.07	0.85, 1.36
A06 Concentration	-0.09	0.11	-0.82	.411	0.91	0.74, 1.13
A07 Loses Things	0.06	0.09	0.62	.537	1.06	0.90, 1.26
A08 Easily Distracted	0.45	0.16	2.80**	.005	1.56	1.14, 2.14
A09 Forgetful	0.09	0.10	0.83	.405	1.09	0.89, 1.37
B01 Fidgets	-0.20	0.3	-1.53	.126	0.82	0.64, 1.06
B02 Leaves Seat	0.06	0.12	0.47	.642	1.06	0.84, 1.32
B03 Runs, Climbs	0.35	0.12	2.96**	.003	1.42	1.12, 1.79
B04 Playing Quietly	-0.07	0.10	-0.64	.521	0.94	0.76, 1.15
B05 Driven, on the go	0.03	0.11	0.27	.790	1.03	0.84, 1.27
B06 Blurts out Answers	0.04	0.10	0.41	.683	1.05	0.85, 1.27
B07 Waits for turn	-0.09	0.10	-0.88	.381	0.91	0.75, 1.12
B08 Interrupts, Intrudes	0.112	0.13	0.91	.363	1.12	0.87, 1.44
B09 Talks excessively	-0.03	0.09	-0.32	.752	0.97	0.81, 1.17
Age	0.06	0.07	0.85	.396	1.06	0.93, 1.20
Gender	-0.05	0.23	-0.23	.815	0.95	0.60, 1.49
F02 Functional Impairment Related to Home Life and Family Members						
A01 Careless	0.09	0.13	0.68	.498	1.09	0.84, 1.42
A02 Difficulties Sustaining Attention	0.39	0.16	2.49*	.013	1.48	1.08, 2.02
A03 Does not Listen	0.33	0.11	3.02**	.004	1.38	1.12, 1.71
A04 Does not Finish Work	-0.16	0.13	-1.22	.222	0.85	0.65, 1.10
A05 Organizational Skills	0.07	0.13	0.54	.588	1.07	0.84, 1.36
A06 Concentration	0.28	0.11	2.44*	.015	1.32	1.06, 1.65
A07 Loses Things	-0.07	0.09	-0.78	.433	0.93	0.78, 1.11
A08 Easily Distracted	0.13	0.17	0.79	.428	1.14	0.82, 1.59
A09 Forgetful	0.01	0.10	0.15	.883	1.01	0.83, 1.23
B01 Fidgets	-0.25	0.13	-1.99*	.047	0.78	0.60, 1.00
B02 Leaves Seat	0.27	0.12	2.24*	.025	1.30	1.03, 1.66
B03 Runs, Climbs	0.23	0.12	1.91	.057	1.26	0.99, 1.59
B04 Playing Quietly	0.00	0.11	-0.02	.981	0.99	0.81, 1.23
B05 Driven, on the go	0.26	0.11	2.35*	.019	1.29	1.04, 1.60
B06 Blurts out Answers	0.16	0.10	1.53	.127	1.17	0.96, 1.43
B07 Waits for turn	-0.06	0.10	-0.55	.580	0.94	0.77, 1.16
B08 Interrupts, Intrudes	0.04	0.13	0.31	.760	1.04	0.81, 1.34
B09 Talks excessively	0.03	0.09	0.31	.759	1.03	0.86, 1.24
Age	-0.06	0.07	-0.98	.326	0.94	0.83, 1.07
Gender	0.20	0.23	0.87	.383	1.22	0.78, 1.91
F03 Functional Impairment Related to Relationships with Adults						

Predictors	Estimate	*SE*	*z*	*p*	*OR*	*95% CI*
A01 Careless	-0.11	0.13	-0.88	.380	0.89	0.69, 1.14
A02 Difficulties Sustaining Attention	0.16	0.16	1.05	.293	1.17	0.87, 1.60
A03 Does not Listen	0.12	0.10	1.16	.244	1.13	0.92, 1.39
A04 Does not Finish Work	0.03	0.13	0.24	.808	1.03	0.81, 1.32
A05 Organizational Skills	-0.05	0.12	-0.43	.670	0.95	0.75, 1.21
A06 Concentration	0.09	0.11	0.80	.421	1.09	0.88, 1.35
A07 Loses Things	0.10	0.09	1.08	.279	1.10	0.93, 1.31
A08 Easily Distracted	0.10	0.16	0.60	.551	1.10	0.80, 1.50
A09 Forgetful	0.11	0.10	1.15	.246	1.12	0.92, 1.36
B01 Fidgets	-0.04	0.13	-0.32	.752	0.96	0.75, 1.22
B02 Leaves Seat	0.29	0.12	2.50*	.013	1.34	1.07, 1.69
B03 Runs, Climbs	0.07	0.12	0.57	.571	1.07	0.85, 1.34
B04 Playing Quietly	0.04	0.11	0.35	.724	1.04	0.84, 1.28
B05 Driven, on the go	0.18	0.11	1.73	.084	1.20	0.98, 1.48
B06 Blurts out Answers	0.06	0.10	0.63	.527	1.07	0.87, 1.30
B07 Waits for turn	-0.01	0.10	-0.10	.923	0.99	0.81, 1.21
B08 Interrupts, Intrudes	0.21	0.13	1.60	.110	1.22	0.95, 1.58
B09 Talks excessively	0.014	0.09	0.15	.882	1.01	0.85, 1.22
Age	0.16	0.06	2.49*	.013	1.17	1.03, 1.33
Gender	0.53	0.23	2.30*	.021	1.69	1.08, 2.65
F04 Functional Impairment Related to Relationships with Children/Adolescents; Recreational Activities						
A01 Careless	-0.12	0.13	-0.91	.363	0.89	0.69, 1.14
A02 Difficulties Sustaining Attention	0.26	0.16	1.63	.102	1.29	0.95, 1.76
A03 Does not Listen	0.10	0.10	0.93	.354	1.10	0.90, 1.35
A04 Does not Finish Work	-0.08	0.13	-0.59	.555	0.93	0.72, 1.19
A05 Organizational Skills	0.17	0.12	1.40	.161	1.19	0.93, 1.51
A06 Concentration	0.01	0.11	0.11	.910	1.01	0.82, 1.25
A07 Loses Things	-0.21	0.09	-2.27*	.023	0.81	0.68, 0.97
A08 Easily Distracted	0.69	0.16	4.22***	< .001	1.99	1.45, 2.75
A09 Forgetful	0.14	0.10	1.40	.162	1.14	0.95, 1.39
B01 Fidgets	-0.19	0.13	-1.47	.142	0.83	0.65, 1.06
B02 Leaves Seat	0.11	0.12	0.90	.366	1.11	0.88, 1.40
B03 Runs, Climbs	0.05	0.11	0.43	.660	1.05	0.84, 1.32
B04 Playing Quietly	-0.07	0.11	-0.70	.482	0.93	0.75, 1.14
B05 Driven, on the go	0.12	0.10	1.18	.238	1.13	0.92, 1.39
B06 Blurts out Answers	0.23	0.10	2.26*	.024	1.25	1.03, 1.53
B07 Waits for turn	0.10	0.10	1.01	.315	1.10	0.91, 1.36
B08 Interrupts, Intrudes	0.48	0.13	3.63***	< .001	1.65	1.25, 2.10
B09 Talks excessively	0.17	0.09	1.80	.072	1.19	0.99, 1.43
Age	-0.07	0.07	-1.07	.286	0.93	0.82, 1.06
Gender	0.82	0.23	3.37***	< .001	2.26	1.44, 3.60
F05 Functional Impairment Related to Academic Performance						
A01 Careless	0.22	0.13	1.63	.103	1.24	0.96, 1.61
A02 Difficulties Sustaining Attention	0.29	0.16	1.82	.069	1.34	0.98, 1.85

Predictors	Estimate	*SE*	*z*	*p*	*OR*	*95% CI*
A03 Does not Listen	0.07	0.11	0.61	.540	1.07	0.87, 1.32
A04 Does not Finish Work	0.32	0.13	2.49*	.013	1.38	1.07, 1.78
A05 Organizational Skills	-0.05	0.15	-0.44	.663	0.95	0.74, 1.21
A06 Concentration	0.28	0.11	2.50*	.012	1.32	1.06, 1.64
A07 Loses Things	-0.20	0.09	-2.10*	.036	0.82	0.68, 0.99
A08 Easily Distracted	0.50	0.17	2.96**	.003	1.65	1.19, 2.23
A09 Forgetful	0.09	0.10	0.88	.378	1.09	0.90, 1.33
B01 Fidgets	-0.17	0.13	-1.35	.177	0.84	0.66, 1.08
B02 Leaves Seat	0.08	0.12	0.65	.513	1.08	0.86, 1.36
B03 Runs, Climbs	0.18	0.12	1.47	.143	1.19	0.94, 1.51
B04 Playing Quietly	0.17	0.11	1.57	.118	1.19	0.96, 1.47
B05 Driven, on the go	-0.15	0.11	-1.36	.174	0.86	0.69, 1.07
B06 Blurts out Answers	0.11	0.10	1.10	.271	1.12	0.91, 1.37
B07 Waits for turn	-0.17	0.11	-1.65	.099	0.84	0.68, 1.03
B08 Interrupts, Intrudes	0.20	0.13	1.54	.123	1.22	0.95, 1.58
B09 Talks excessively	-0.10	0.09	-1.05	.293	0.91	0.75, 1.09
Age	0.15	0.07	2.28*	.023	1.16	1.02, 1.33
Gender	0.00	0.23	-0.02	.986	1.00	0.63, 1.67

Note. Significant predictors are shaded in light gray and the response variables are shaded in dark gray. McFadden's Pseudo R^2 = .06 (F01); .09 (F02); .07 (F03); .12 (F04); .08 (F05). ADHD = attention-deficit/hyperactivity disorder, OR = odds ratio.
* $p < .05$; ** $p < .01$; *** $p < .001$.

Table S7

Ordinal logistic regression analysis predicting functional impairment domains from individual ODD/CD/CU symptoms

Predictors	Estimate	*SE*	*z*	*p*	*OR*	*95% CI*
F01 Psychological Strain						
A01 Loses Temper	0.43	0.16	2.71**	.007	1.54	1.13, 2.11
A02 Touchy, Easily Annoyed	0.07	0.14	0.47	.640	1.07	0.81, 1.40
A03 Angry, Resentful	0.38	0.13	2.86**	.004	1.46	1.12, 1.84
A04 Argues with Adults	0.25	0.13	1.92	.055	1.28	0.99, 1.65
A05 Refuses to Comply with Requests	-0.03	0.13	-0.21	.835	0.97	0.75, 1.26
A06 Annoys	0.06	0.12	0.48	.634	1.06	0.83, 1.35
A07 Blames Others	0.15	0.11	1.30	.195	1.16	0.93, 1.44
A08 Spiteful, Vindictive	-0.12	0.16	-0.72	.470	0.89	0.64, 1.26
D01 Recurrent Temper Outbursts	-0.09	0.16	-0.58	.563	0.91	0.66, 1.25
D02 Persistently Irritable or Angry Mood	0.07	0.16	0.47	.641	1.07	0.79, 1.56
B01 Physical Fights	0.31	0.16	2.02*	.043	1.37	1.01, 1.86
B02 Bullies, Threatens, or Intimidates	0.24	0.17	1.40	.163	1.27	0.91, 1.78
B03 Cruel to Animals	-0.02	0.28	-0.07	.945	0.98	0.56, 1.72
B04. Lies	0.35	0.12	2.79**	.005	1.41	1.10, 1.81
B05. Steals Without Confrontation	0.05	0.19	0.28	.777	1.05	0.73, 1.52
C01a Lack of Remorse	-0.08	0.16	-0.53	.596	0.92	0.67, 1.25
C01b Lack of Concern	0.09	0.15	0.64	.521	1.10	0.82, 1.46
C02a Cold and Uncaring	-0.45	0.18	-2.52*	.012	0.64	0.45, 0.90
C02b Self-Serving	0.27	0.15	1.81	.071	1.31	0.98, 1.75
C03a Indifferent to Poor Performance	-0.32	0.15	-2.08*	.038	0.73	0.54, 0.98
C03b Avoids Effort	0.04	0.12	0.33	.740	1.04	0.82, 1.33
C03c Blames Others for Poor Performance	-0.03	0.13	-0.26	.792	0.97	0.76, 1.24
C04a Shallow, Deficient Affect	0.06	0.17	0.34	.732	1.06	0.76, 1.48
C04b Turns Emotions ‘On’ or ‘Off’ Quickly	0.06	0.15	0.42	.676	1.06	0.80, 1.42
C04c Manipulates	-0.02	0.16	-0.10	.919	0.98	0.71, 1.35
C04d Inconsistent Affect	0.19	0.18	1.04	.297	1.21	0.85, 1.75
Age	0.02	0.07	0.31	.760	1.02	0.90, 1.16
Gender	-0.15	0.25	-0.61	.544	0.86	0.52, 1.42
F02 Functional Impairment Related to Home Life and Family Members						
A01 Loses Temper	0.39	0.15	2.52*	.012	1.48	1.09, 2.00
A02 Touchy, Easily Annoyed	-0.09	0.14	-0.68	.496	0.91	0.69, 1.19
A03 Angry, Resentful	0.15	0.13	1.14	.253	1.16	0.90, 1.49
A04 Argues with Adults	0.56	0.13	4.31***	< .001	1.75	1.36, 2.27
A05 Refuses to Comply with Requests	0.25	0.13	1.96*	.050	1.28	1.00, 1.65
A06 Annoys	0.26	0.12	2.10*	.035	1.30	1.02, 1.66
A07 Blames Others	0.12	0.11	1.03	.305	1.12	0.90, 1.41
A08 Spiteful, Vindictive	-0.01	0.17	-0.03	.975	0.99	0.71, 1.40
D01 Recurrent Temper Outbursts	-0.30	0.16	-1.80	.072	0.74	0.54, 1.03
D02 Persistently Irritable or Angry Mood	0.37	0.16	2.34*	.019	1.44	1.06, 1.97
B01 Physical Fights	0.12	0.16	0.71	.476	1.12	0.82, 1.55
B02 Bullies, Threatens, or Intimidates	0.03	0.18	0.15	.881	1.03	0.73, 1.45
B03 Cruel to Animals	0.17	0.29	0.58	.559	1.19	0.66, 2.10

Predictors	Estimate	*SE*	*z*	*p*	*OR*	*95% CI*
B04. Lies	0.23	0.12	1.94	.052	1.26	1.00, 1.60
B05. Steals Without Confrontation	0.22	0.18	1.24	.215	1.25	0.88, 1.79
C01a Lack of Remorse	0.12	0.16	0.79	.430	1.13	0.83, 1.54
C01b Lack of Concern	0.03	0.14	0.21	.832	1.03	0.78, 1.36
C02a Cold and Uncaring	-0.03	0.17	-0.18	.858	0.97	0.69, 1.36
C02b Self-Serving	0.23	0.15	1.56	.119	1.26	0.94, 1.70
C03a Indifferent to Poor Performance	-0.18	0.15	-1.22	.224	0.83	0.62, 1.11
C03b Avoids Effort	0.05	0.12	0.38	.702	1.05	0.83, 1.33
C03c Blames Others for Poor Performance	0.07	0.13	0.56	.574	1.07	0.84, 1.37
C04a Shallow, Deficient Affect	-0.09	0.17	-0.50	.615	0.92	0.66, 1.28
C04b Turns Emotions 'On' or 'Off' Quickly	-0.07	0.15	-0.47	.640	0.93	0.70, 1.25
C04c Manipulates	0.10	0.16	0.61	.5425	1.10	0.80, 1.53
C04d Inconsistent Affect	0.53	0.19	2.77**	.006	1.70	1.17, 2.48
Age	-0.01	0.07	-0.14	.888	0.99	0.87, 1.13
Gender	0.04	0.25	0.15	.884	1.03	0.64, 1.70
F03 Functional Impairment Related to Relationships with Adults						
A01 Loses Temper	0.26	0.17	1.56	.118	1.30	0.94, 1.80
A02 Touchy, Easily Annoyed	-0.08	0.15	-0.57	.569	0.92	0.69, 1.23
A03 Angry, Resentful	0.23	0.14	1.67	.095	1.25	0.96, 1.64
A04 Argues with Adults	0.37	0.13	2.77**	.006	1.45	1.12, 1.89
A05 Refuses to Comply with Requests	0.15	0.14	1.06	.289	1.16	0.88, 1.52
A06 Annoys	0.36	0.13	2.86**	.004	1.44	1.12, 1.85
A07 Blames Others	0.26	0.12	2.16*	.030	1.29	1.02, 1.63
A08 Spiteful, Vindictive	-0.36	0.16	-2.26*	.024	0.70	0.51, 0.95
D01 Recurrent Temper Outbursts	-0.40	0.17	-2.34*	.019	0.67	0.48, 0.94
D02 Persistently Irritable or Angry Mood	0.18	0.16	1.14	.254	1.20	0.88, 1.63
B01 Physical Fights	0.45	0.16	2.81**	.005	1.57	1.15, 2.15
B02 Bullies, Threatens, or Intimidates	0.12	0.17	0.69	.493	1.12	0.81, 1.56
B03 Cruel to Animals	0.30	0.27	1.12	.262	1.36	0.80, 2.32
B04. Lies	0.08	0.13	0.65	.516	1.09	0.85, 1.39
B05. Steals Without Confrontation	0.45	0.19	2.35*	.019	1.56	1.08, 2.27
C01a Lack of Remorse	0.10	0.16	0.64	.523	1.11	0.81, 1.52
C01b Lack of Concern	0.31	0.15	2.13*	.033	1.37	1.03, 1.83
C02a Cold and Uncaring	-0.36	0.18	-1.95	.051	0.70	0.49, 1.00
C02b Self-Serving	0.29	0.15	1.97*	.049	1.34	1.00, 1.79
C03a Indifferent to Poor Performance	-0.13	0.16	-0.84	.401	0.88	0.64, 1.19
C03b Avoids Effort	0.31	0.13	2.38*	.017	1.36	1.06, 1.75
C03c Blames Others for Poor Performance	0.29	0.13	2.25*	.024	1.34	1.04, 1.72
C04a Shallow, Deficient Affect	-0.41	0.18	-2.25*	.024	0.66	0.46, 0.94
C04b Turns Emotions 'On' or 'Off' Quickly	0.22	0.16	1.40	.160	1.25	0.92, 1.70
C04c Manipulates	-0.05	0.17	-0.31	.757	0.95	0.67, 1.32
C04d Inconsistent Affect	0.21	0.19	1.10	.274	1.23	0.85, 1.78
Age	0.05	0.07	0.65	.516	1.05	0.91, 1.21
Gender	0.34	0.29	1.15	.250	1.40	0.80, 2.52
F04 Functional Impairment Related to Relationships with Children/Adolescents; Recreational Activities						
A01 Loses Temper	0.35	0.17	2.13*	.034	1.42	1.03, 1.97

Predictors	Estimate	*SE*	*z*	*p*	*OR*	*95% CI*
A02 Touchy, Easily Annoyed	0.33	0.15	2.26*	.024	1.39	1.05, 1.85
A03 Angry, Resentful	0.12	0.14	0.90	.367	1.13	0.86, 1.48
A04 Argues with Adults	0.21	0.13	1.61	.107	1.23	0.96, 1.59
A05 Refuses to Comply with Requests	0.19	0.13	1.45	.148	1.21	0.93, 1.58
A06 Annoys	0.28	0.18	2.18*	.029	1.32	1.03, 1.69
A07 Blames Others	0.14	0.12	1.17	.242	1.15	0.91, 1.45
A08 Spiteful, Vindictive	-0.69	0.17	-4.05***	< .001	0.50	0.36, 0.70
D01 Recurrent Temper Outbursts	-0.30	0.17	-1.77	.077	0.74	0.53, 1.03
D02 Persistently Irritable or Angry Mood	0.04	0.16	0.26	.795	1.04	0.77, 1.41
B01 Physical Fights	0.61	0.16	3.79***	< .001	1.83	1.34, 2.52
B02 Bullies, Threatens, or Intimidates	0.23	0.17	1.33	.183	1.25	0.90, 1.76
B03 Cruel to Animals	0.63	0.27	2.32*	.021	1.88	1.10, 3.23
B04. Lies	0.21	0.13	1.69	.091	1.24	0.97, 1.58
B05. Steals Without Confrontation	0.21	0.18	1.17	.241	1.24	0.86, 1.77
C01a Lack of Remorse	0.21	0.16	1.35	.178	1.24	0.91, 1.69
C01b Lack of Concern	-0.03	0.15	-0.23	.815	0.97	0.72, 1.29
C02a Cold and Uncaring	0.15	0.17	0.88	.378	1.17	0.83, 1.64
C02b Self-Serving	0.00	0.15	0.02	.980	1.00	0.74, 1.35
C03a Indifferent to Poor Performance	-0.19	0.16	-1.18	.238	0.83	0.61, 1.12
C03b Avoids Effort	0.23	0.13	1.80	.072	1.25	0.98, 1.61
C03c Blames Others for Poor Performance	-0.03	0.13	-0.25	.800	0.97	0.75, 1.25
C04a Shallow, Deficient Affect	-0.07	0.18	-0.38	.705	0.94	0.66, 1.32
C04b Turns Emotions ‘On’ or ‘Off’ Quickly	0.08	0.15	0.56	.573	1.09	0.81, 1.46
C04c Manipulates	0.04	0.17	0.21	.833	1.04	0.74, 1.44
C04d Inconsistent Affect	0.49	0.19	2.62	.009**	1.63	1.13, 2.36
Age	0.02	0.07	0.23	.815	1.02	0.89, 1.17
Gender	0.07	0.27	0.25	.804	1.07	0.63, 1.84
F05 Functional Impairment Related to Academic Performance						
A01 Loses Temper	0.21	0.17	1.23	.219	1.23	0.88, 1.71
A02 Touchy, Easily Annoyed	0.05	0.15	0.38	.708	1.06	0.79, 1.41
A03 Angry, Resentful	0.38	0.14	2.23*	.026	1.36	1.04, 1.79
A04 Argues with Adults	0.26	0.13	1.92	.055	1.29	0.99, 1.68
A05 Refuses to Comply with Requests	0.27	0.14	1.96	.050	1.31	1.00, 1.72
A06 Annoys	0.16	0.13	1.28	.201	1.18	0.92, 1.51
A07 Blames Others	0.09	0.12	0.74	.460	1.09	0.86, 1.39
A08 Spiteful, Vindictive	-0.54	0.16	-3.30***	< .001	0.58	0.42, 0.80
D01 Recurrent Temper Outbursts	-0.15	0.17	-0.86	.393	0.86	0.62, 1.21
D02 Persistently Irritable or Angry Mood	0.01	0.16	0.08	.935	1.01	0.74, 1.38
B01 Physical Fights	0.42	0.16	2.66**	.008	1.52	1.12, 2.08
B02 Bullies, Threatens, or Intimidates	0.17	0.17	1.01	.314	1.19	0.85, 1.67
B03 Cruel to Animals	0.52	0.29	1.80	.072	1.67	0.95, 2.93
B04. Lies	0.02	0.13	0.14	.892	1.02	0.79, 1.30
B05. Steals Without Confrontation	0.32	0.19	1.68	.093	1.37	0.95, 1.99
C01a Lack of Remorse	0.10	0.16	0.60	.552	1.10	0.80, 1.52
C01b Lack of Concern	0.25	0.15	1.69	.090	1.28	0.96, 1.70
C02a Cold and Uncaring	0.13	0.18	0.69	.491	1.13	0.79, 1.63
C02b Self-Serving	-0.08	0.15	-0.54	.591	0.92	0.68, 1.24
C03a Indifferent to Poor Performance	-0.16	0.16	-0.99	.321	0.85	0.62, 1.17

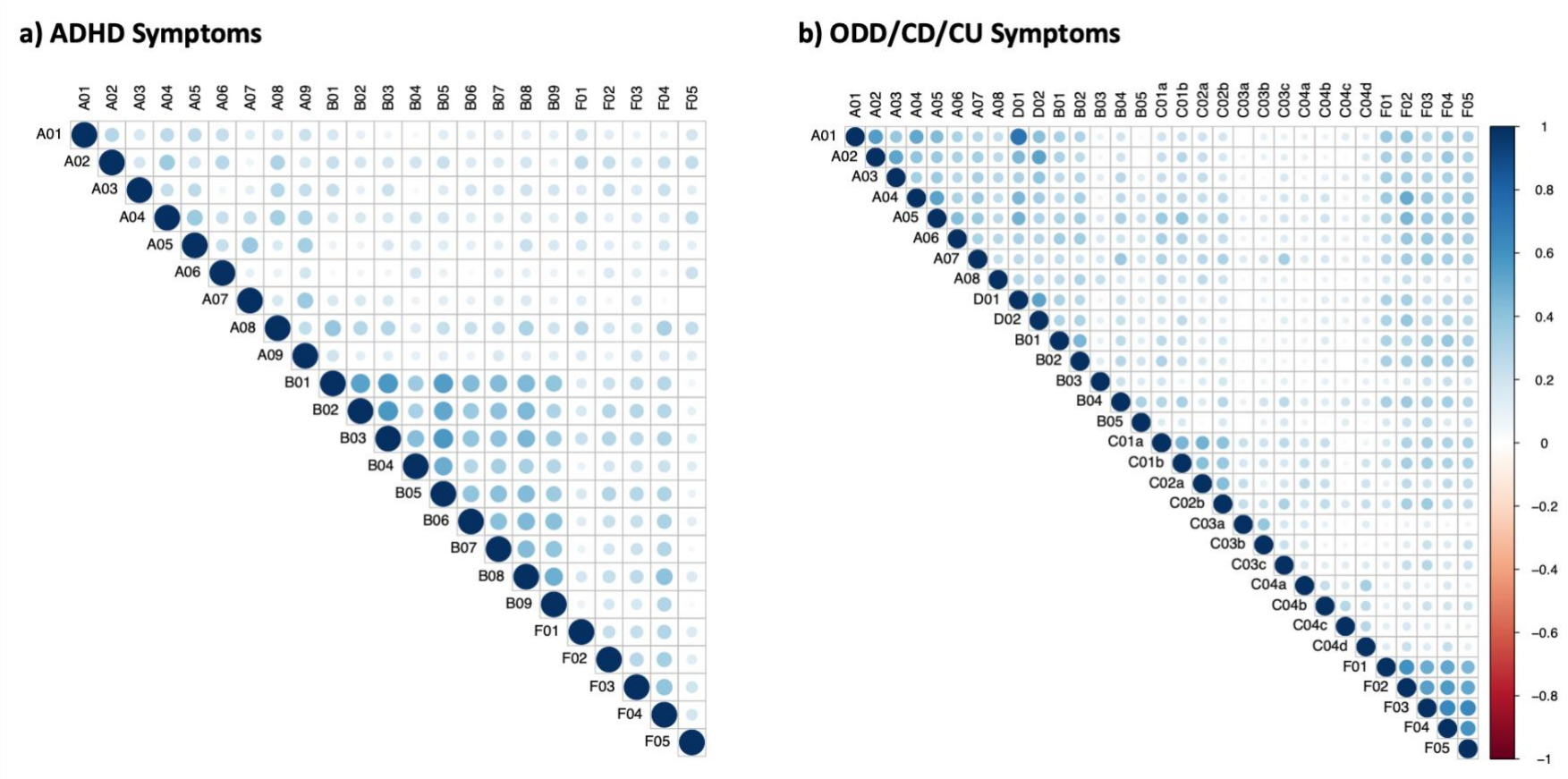
Predictors	Estimate	*SE*	*z*	*p*	*OR*	*95% CI*
C03b Avoids Effort	0.46	0.13	3.50***	< .001	1.58	1.22, 2.05
C03c Blames Others for Poor Performance	0.10	0.13	0.76	.449	1.10	0.85, 1.43
C04a Shallow, Deficient Affect	-0.38	0.19	-2.02*	.043	0.69	0.47, 0.98
C04b Turns Emotions 'On' or 'Off' Quickly	0.19	0.16	1.21	.225	1.21	0.89, 1.66
C04c Manipulates	-0.01	0.18	-0.06	.951	0.99	0.70, 1.40
C04d Inconsistent Affect	0.21	0.19	1.10	.272	1.24	0.85, 1.81
Age	0.07	0.07	0.94	.350	1.07	0.93, 1.23
Gender	0.39	0.30	1.31	.189	1.47	0.84, 2.67

Note. Significant predictors are shaded in light gray and the response variables are shaded in dark gray. McFadden's Pseudo R^2 = .16 (F01); .20 (F02); .22 (F03); .21 (F04); .18 (F05). ODD = oppositional defiant disorder; CD = conduct disorder; CU = callous-unemotional, OR = odds ratio.

* $p < .05$; ** $p < .01$; *** $p < .001$

Figure S1

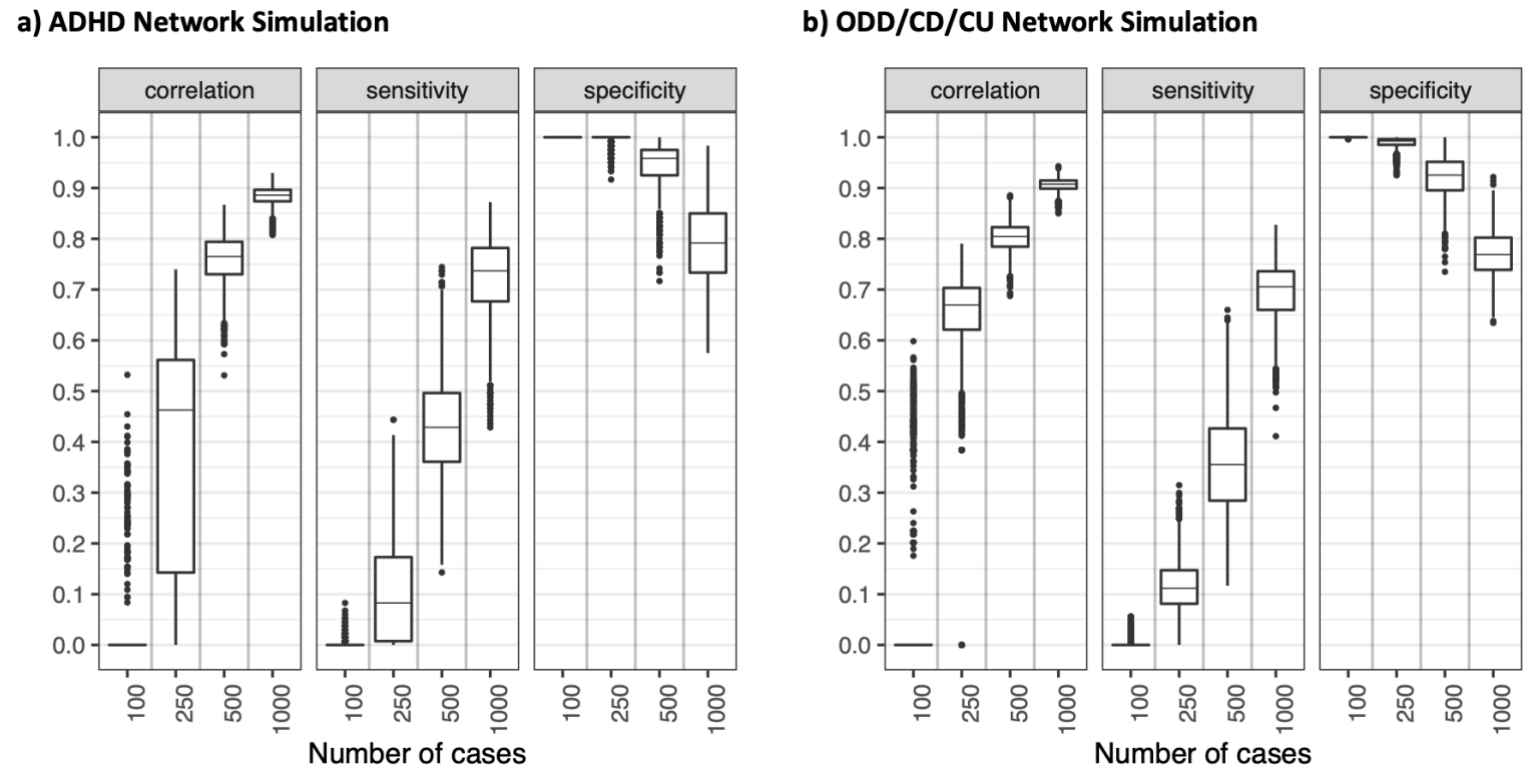
Zero-order correlation matrix for ADHD and ODD/CD/CU symptoms



Note. Depicted are the zero-order Spearman intercorrelations between the ADHD symptoms (Figure S1a) and between the ODD/CD/CU symptoms (Figure S1b). Dark blue indicates strong positive correlations, dark red indicates strong negative correlations, and white indicates absence of correlation. The complete wording for each item can be found in Figure 1. ADHD = attention-deficit/hyperactivity disorder; ODD = oppositional defiant disorder; CD = conduct disorder; CU = callous-unemotional.

Figure S2

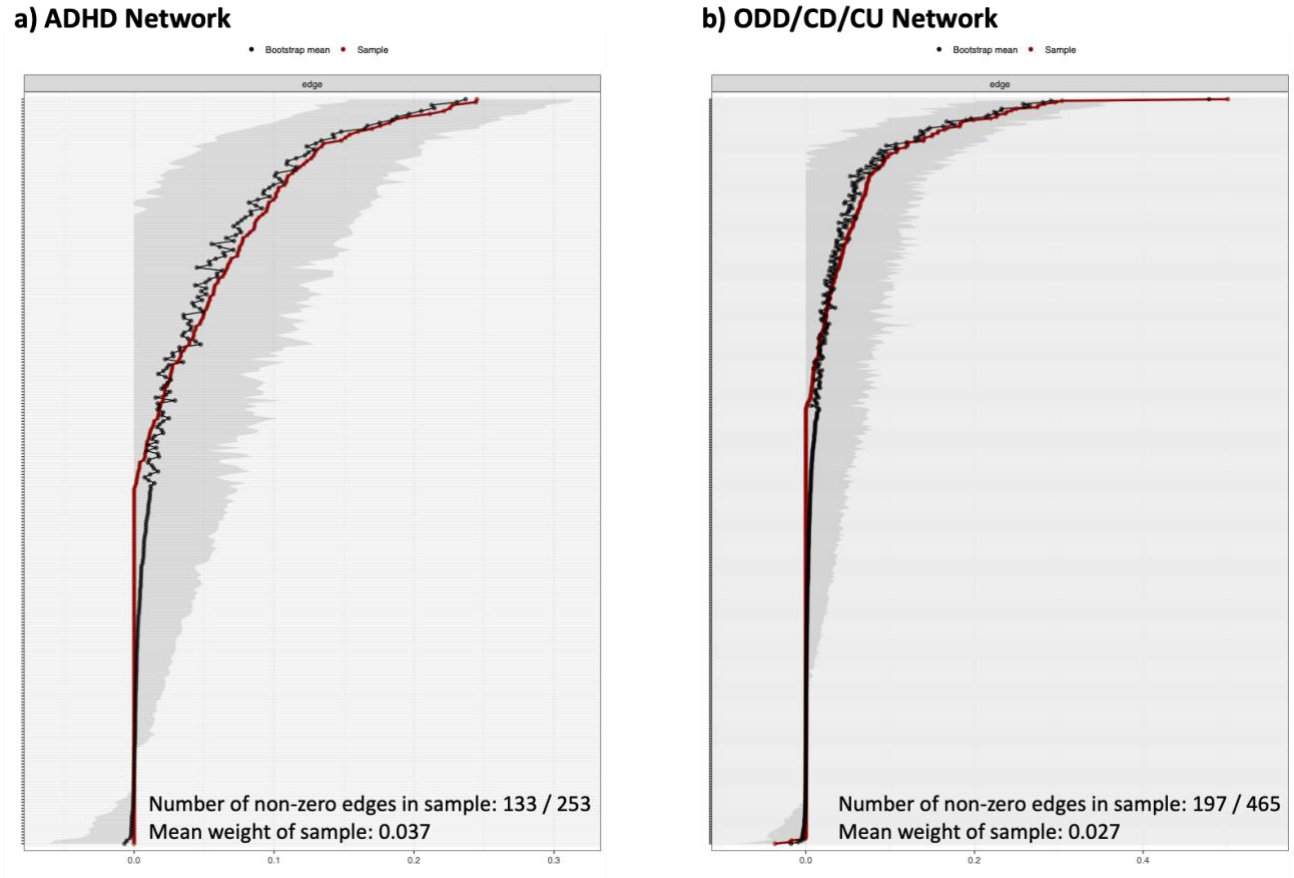
Simulation studies estimating the correlation, sensitivity, and specificity of the ADHD and ODD/CD/CU networks



Note. Simulation results of simulated data sets for varying sample sizes (n = 100, 250, 500, 1,000). Data sets were generated using the *netSimulator* function from the *bootnet* package and were based on the EBICglasso estimation, Spearman correlations, 2,500 repetitions, and a tuning parameter of $\lambda = 0.5$. For a moderate sample size (n = 500), as in our case, we can derive a moderate correlation accuracy between the "true" and the estimated edge weights, moderate to low sensitivities (i.e. the proportion of edges in the "true" networks that were also included in the estimated networks), and moderate to high specificities (i.e. the proportion of missing edges in the "true" networks that were correctly not included in the estimated networks). These results indicate that it is highly probable that the edges we observed "truly" exist. ADHD = attention-deficit/hyperactivity disorder; ODD = oppositional defiant disorder; CD = conduct disorder; CU = callous-unemotional.

Figure S3

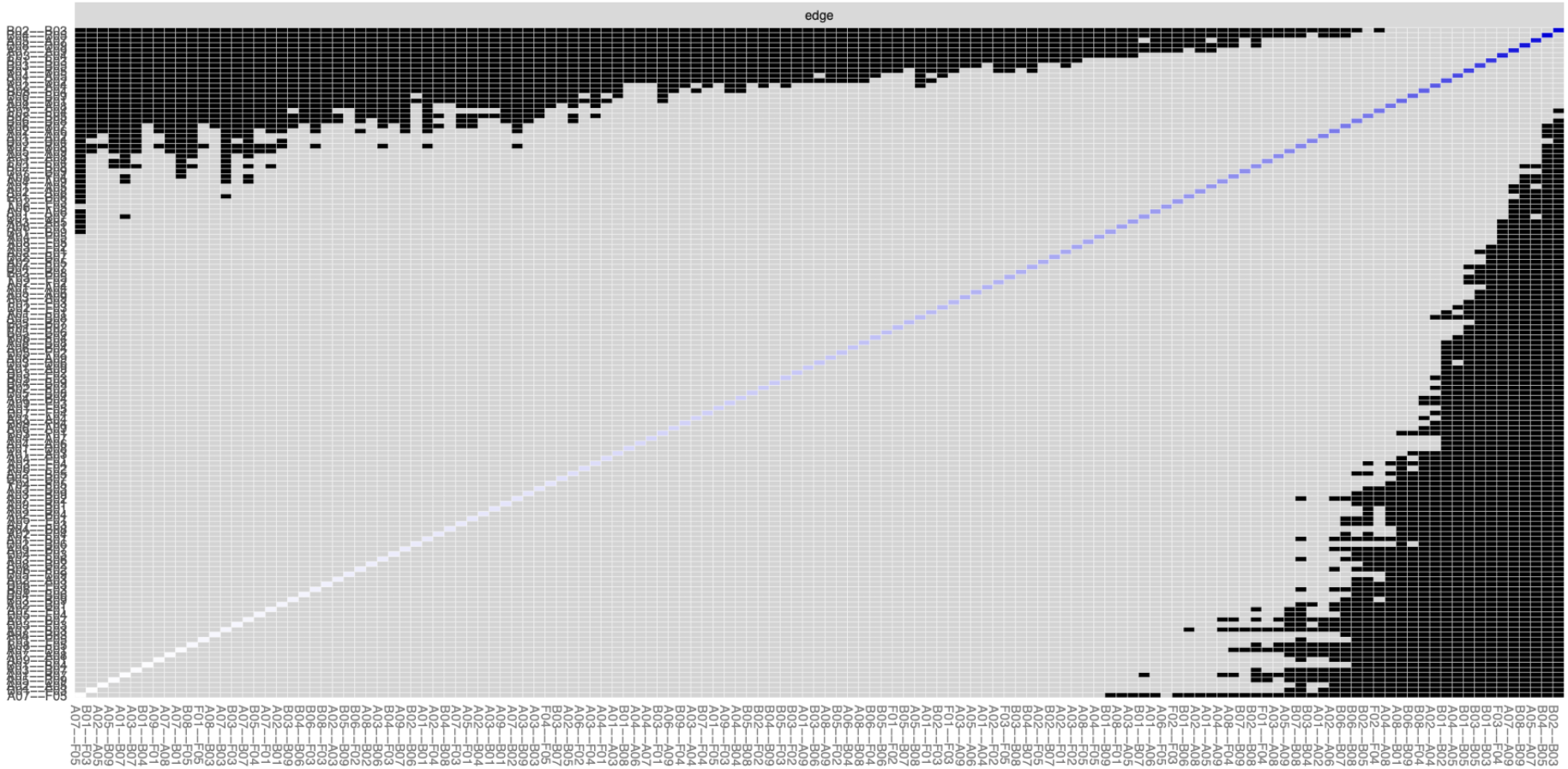
Edge weight accuracies for the ADHD and ODD/CD/CU networks



Note. The horizontal gray area represents the 95% quantile range of the parameter values based on 2,500 bootstrap samples. The red dots represent the original sample values and the black dots represent the bootstrap mean values. The y-axis shows all edges in the network (labels omitted for ease of legibility) and the x-axis shows the strength of the edge weights. ADHD = attention-deficit/hyperactivity disorder; ODD = oppositional defiant disorder; CD = conduct disorder; CU = callous-unemotional.

Figure S4

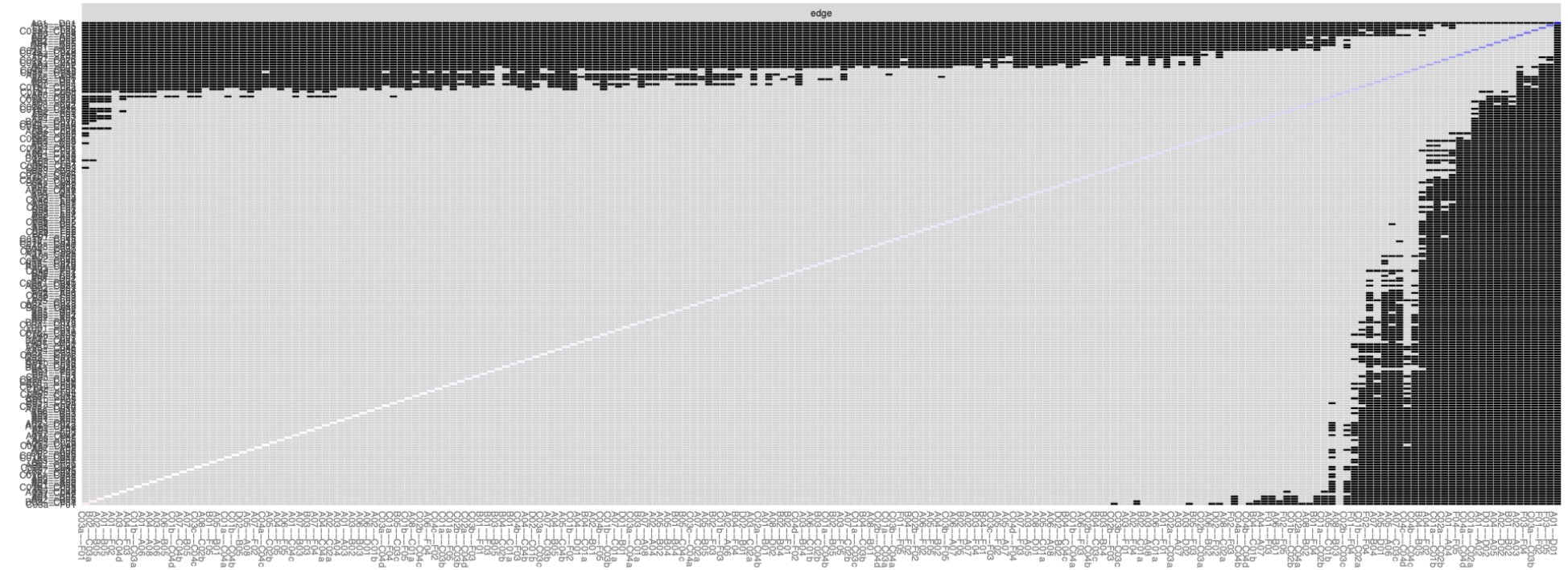
Bootstrapped difference test for the edge weights from the attention-deficit/hyperactivity disorder (ADHD) network



Note. The gray boxes indicate that two edge weights do not significantly differ from each other; the black boxes mark a significant difference between two edge weights (α = .05). Most of the strongest edges differ significantly from the other edges in the network.

Figure S5

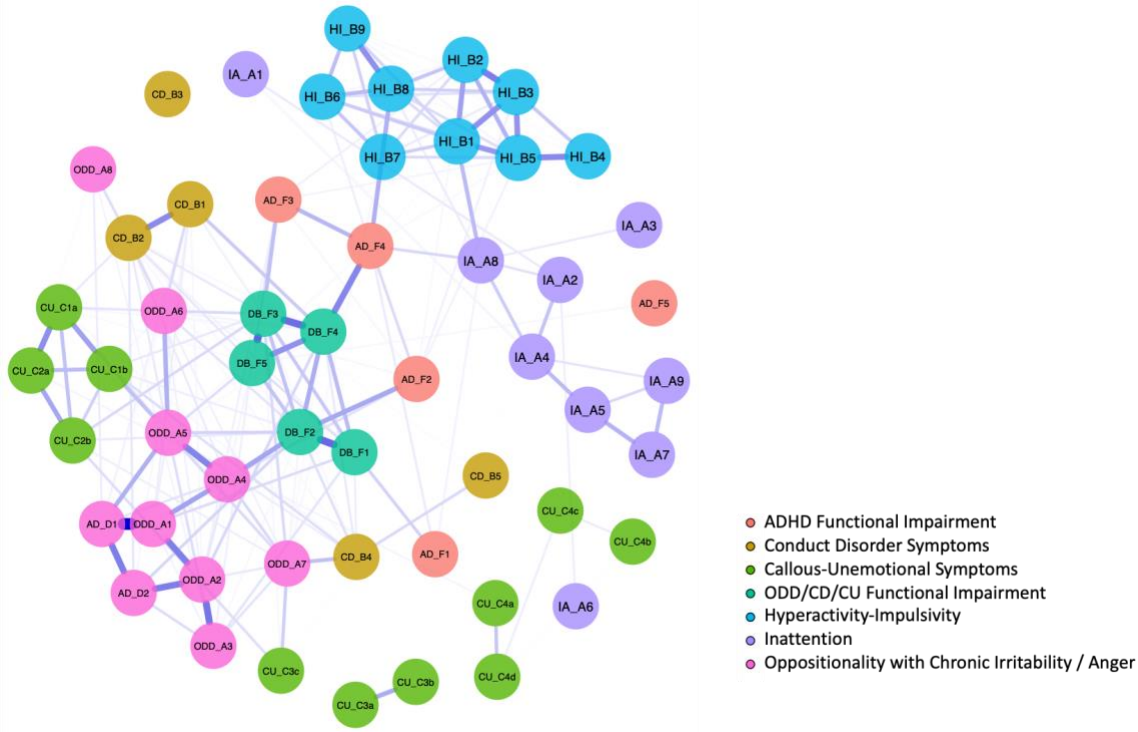
Bootstrapped difference test for the edge weights from the ODD/CD/CU network



Note. The gray boxes indicate that two edge weights do not significantly differ from each other; the black boxes mark a significant difference between two edge weights (α = .05). Most of the strongest edges differ significantly from the other edges in the network. ODD = oppositional defiant disorder; CD = conduct disorder; CU = callous-unemotional.

Figure S6

Network of all externalizing symptoms with associated functional impairment domains



Note. The figure depicts the network structure of all externalizing symptoms and functional impairment related to psychological strain (F1) and in the domains of home life and family members (F2), relationships with adults (F3), relationships with children/adolescents and recreational activities (F4), and academic performance (F5). Item dimensions are differentiated by color. Blue edges represent positive partial correlations, and the thickness of an edge represents the strength of the partial correlation. A short description of each item is provided in Figure 1. Regarding the strongest associations of functional impairment domains with symptoms, the strongest partial correlations (i.e., the largest edge weights) were as follows: ADHD F1 and A8 *Easily Distracted* ($\rho_{xy \bullet z}$ = .02), ADHD F2 and A4 *Argues with Adults* ($\rho_{xy \bullet z}$ = .05), ADHD F3 and B2 *Leaves Seat* ($\rho_{xy \bullet z}$ = .02), ADHD F4 and B8 *Interrupts, Intrudes* ($\rho_{xy \bullet z}$ =.12), ADHD F5 showed no associations with other symptoms, DB F1 and A1 *Loses Temper* ($\rho_{xy \bullet z}$ = .06), DB F2 and A4 *Argues with Adults* ($\rho_{xy \bullet z}$ = .14), DB F3 and A7 *Blames Others* ($\rho_{xy \bullet z}$ = .06), DB F4 and B1 *Physical Fights* ($\rho_{xy \bullet z}$ = .08), and DB F5 and A5 *Refuses to Comply* ($\rho_{xy \bullet z}$ = .05). ADHD = attention-deficit/hyperactivity disorder; IA = Inattention; HI = hyperactivity-impulsivity; DB = disruptive behavior symptoms; ODD = oppositional defiant disorder; AD = affective dysregulation; CD = conduct disorder; CU = callous-unemotional.

Figure S7

Edge weight accuracies for the network of externalizing symptoms

Note. The horizontal gray area represents the 95% quantile range of the parameter values based on 2,500 bootstrap samples. The red dots represent the original sample values and the black dots represent the bootstrap mean values. The y-axis shows all edges in the network (labels omitted for ease of legibility) and the x-axis shows the strength of the edge weights.

ANN-KATHRIN THÖNE

PSYCHOLOGIST

Pohligstr. 9, 50969 Cologne, Germany | Email: ann-kathrin.thoene@uk-koeln.de

EDUCATION

09/2018 – 06/2014 | PhD in Health Sciences

Interdisciplinary Program Health Sciences (IPHS), Medical Faculty, University of Cologne, Germany

- Dissertation: Categorical and dimensional concepts of externalizing disorders in children

10/2018 – Present | Child and Adolescent Psychotherapist in Training

School of Child and Adolescent Cognitive Behavior Therapy (AKiP), University Hospital Cologne, Germany

10/2013 – 01/2017 | Master of Science in Neurocognitive Psychology

University of Oldenburg, Germany

- Master thesis: Atypical brain connectivity in autism: An MEG study on resting state network profiles

10/2010 – 07/2013 | Bachelor of Science in Psychology

University of Bremen, Germany

- Bachelor thesis: Alzheimer's dementia and the amyloid hypothesis

07/2003 – 06/2010 | Abitur

Hermann-Böse-Gymnasium, Bremen, Germany

PUBLICATIONS

Dose, C.*, **Thöne, A. K.*,** Del Giudice, T., Görtz-Dorten, A., Faber, M., Benesch, C., Hautmann, C., Lindenschmidt, T., & Döpfner, M. (2022). Child-therapist and parent-therapist alliances and outcome in the treatment of children with oppositional defiant/conduct disorder. *Psychother. Res.*, 1–14. https://doi.org/10.1080/10503307.2022.2138791

Görtz-Dorten, A., **Thöne, A.-K.,** & Döpfner, M. (2022). *DISYPS-ILF Interview-Leitfäden zum Diagnostik-System für psychische Störungen nach DSM-5 für Kinder und Jugendliche [Clinical Parent Interviews for Diagnosing Mental Disorders Based on the DSM-5 in Children and Adolescents]*. Hogrefe.

Jendreizik, L. T., Hautmann, C., von Wirth, E., Dose, C., **Thöne, A.-K.,** Treier, A. K., Banaschewski, T., Becker, K., Brandeis, D., Geissler, J., Hebebrand, J., Hohmann, S., Holtmann, M., Huss, M., Jans, T., Kaiser, A., Millenet, S., Poustka, L., Schneider, P., & Döpfner, M. (2022). The importance of familial risk factors in children with ADHD: direct and indirect effects of family adversity, parental psychopathology and parenting practices on externalizing symptoms. *Child Adolesc. Psychiatry Ment. Health, 16*(1), 1–14. https://doi.org/10.1186/s13034-022-00529-z

Junghänel, M.*, **Thöne, A.-K.*,** Dose, C., Breuer, D., Görtz-Dorten, A., & Döpfner, M. (2023). Conceptualizing anxiety and depression in children and adolescents: A latent factor and network analysis. *Curr. Psychol.*, 1-16. https://doi.org/10.1007/s12144-023-04321-w

Junghänel, M., **Thöne, A.-K.,** Ginsberg, C., Görtz-Dorten, A., Frenk, F., Mücke, K., Treier, A. K., Labarga, S. Z., Banaschewski, T., Millenet, S., Fegert, J. M., Bernheim, D., Hanisch, C., Kölch, M., Schüller, A., Ravens-Sieberer, U., Kaman, A., Roessner, V., Hinz, J., & Döpfner, M. (2022). Irritability and emotional impulsivity as core feature of ADHD and ODD in children. *J. Psychopathol. Behav. Assess., 44,* 679–697. https://doi.org/10.1007/s10862-022-09974-8

Junghänel, M., Wand, H., Dose, C., **Thöne, A.-K.,** Treier, A. K., Hanisch, C., Ritschel, A., Kölch, M., Lincke, L., Roessner, V., Kohls, G., Ravens-Sieberer, U., Kaman, A., Banaschewski, T., Aggensteiner, P. M., Görtz-Dorten, A., & Döpfner, M. (2022). Validation of a new emotion regulation self-report questionnaire for children. *BMC Psychiatry, 22*(1), 820. https://doi.org/10.1186/s12888-022-04440-x

Möller, T. J., Braun, N**., Thöne, A.-K.,** Herrmann, C. S., & Philipsen, A. (2020). The senses of agency and ownership in patients with borderline personality disorder. *Front. Psychiatry, 11,* 474. https://doi.org/10.3389/fpsyt.2020.00474

Thöne, A. K., Dose, C., Junghänel, M., Hautmann, C., Jendreizik, L. T., Treier, A. K., Vetter, P., von Wirth, E., Banaschewski, T., Becker, K., Brandeis, D., Dürrwächter, U., Geissler, J., Hebebrand, J., Hohmann, S., Holtmann, M., Huss, M., Jans, T., Kaiser, A., … Döpfner, M. (2023). Identifying symptoms of ADHD and disruptive behavior disorders most strongly associated with functional impairment in children: A symptom-level

approach. *J. Psychopathol. Behav. Assess.*, *45*(2), 277–293. https://doi.org/10.1007/s10862-023-10025-z

Thöne, A.-K., Görtz-Dorten, A., Altenberger, P., Dose, C., Geldermann, N., Hautmann, C., Jendreizik, L. T., Treier, A.-K., von Wirth, E., Banaschewski, T., Brandeis, D., Millenet, S., Hohmann, S., Becker, K., Ketter, J., Hebebrand, J., Wenning, J., Holtmann, M., Legenbauer, T., … Döpfner, M. (2020). Toward a dimensional assessment of externalizing disorders in children: Reliability and validity of a semi-structured parent interview. *Front. Psychol., 11,* 1840. https://doi.org/10.3389/fpsyg.2020.01840

Thöne, A.-K.*, Junghänel, M.*, Görtz-Dorten, A., Breuer, D., del Giudice, T., Hanisch, C., Hennemann, T., & Döpfner, M. (2022). Empirically based dimensions of externalizing symptoms in children and adolescents: A multitrait-multisource approach. *J. Psychopathol. Behav. Assess., 44*(3), 844–861. https://doi.org/10.1007/s10862-022-09983-7

Thöne, A.-K., Junghänel, M., Görtz-Dorten, A., Dose, C., Hautmann, C., Jendreizik, L. T., Treier, A.-K., Vetter, P., von Wirth, E., Banaschewski, T., Becker, K., Brandeis, D., Dürrwächter, U., Geissler, J., Hebebrand, J., Hohmann, S., Holtmann, M., Huss, M., Jans, T., … Döpfner, M. (2021). Disentangling symptoms of externalizing disorders in children using multiple measures and informants. *Psychol. Assess.*, *33*(11), 1065–1079. https://doi.org/10.1037/pas0001053

Wiebe, A., Kannen, K., Selaskowski, B., Mehren, A., **Thöne, A.-K.,** Pramme, L., Blumenthal, N., Li, M., Asché, L., Jonas, S., Bey, K., Schulze, M., Steffens, M., Pensel, M. C., Guth, M., Rohlfsen, F., Ekhlas, M., Lügering, H., Fileccia, H., … Braun, N. (2022). Virtual reality in the diagnostic and therapy for mental disorders: A systematic review. *Clin. Psychol. Rev., 98,* 102213. https://doi.org/10.1016/j.cpr.2022.102213

Zich, C., Debener, S., **Thoene, A.-K.,** Chen, L. C., & Kranczioch, C. (2017). Simultaneous EEG-fNIRS reveals how age and feedback affect motor imagery signatures. *Neurobiol. Aging, 49,* 183–197. https://doi.org/10.1016/j.neurobiolaging.2016.10.011

* denotes shared first authorship

7.5 Erklärung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Stellen - einschließlich Tabellen, Karten und Abbildungen -, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, sind in jedem Einzelfall als Entlehnung kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertationsschrift noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss der Promotion nicht ohne Genehmigung der / des Vorsitzenden des IPHS-Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Univ.-Prof. a.D. Dr. Manfred Döpfner betreut worden.

Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen.

Übersicht der Publikationen:

Thöne, A.-K., Görtz-Dorten, A., Altenberger, P., Dose, C., Geldermann, N., Hautmann, C., Jendreizik, L. T., Treier, A.-K., von Wirth, E., Banaschewski, T., Brandeis, D., Millenet, S., Hohmann, S., Becker, K., Ketter, J., Hebebrand, J., Wenning, J., Holtmann, M., Legenbauer, T., Huss, M., Romanos, M., Jans, T., Geissler, J., Poustka, L., Uebel-von Sandersleben, H., Renner, T., Dürrwächter, U., & Döpfner, M. (2020). Toward a dimensional assessment of externalizing disorders in children: Reliability and validity of a semi-structured parent interview. *Front. Psychol., 11,* 1840. https://doi.org/10.3389/fpsyg.2020.01840

Thöne, A.-K., Junghänel, M., Görtz-Dorten, A., Dose, C., Hautmann, C., Jendreizik, L. T., Treier, A.-K., Vetter, P., von Wirth, E., Banaschewski, T., Becker, K., Brandeis, D., Dürrwächter, U., Geissler, J., Hebebrand, J., Hohmann, S., Holtmann, M., Huss, M., Jans, T., Ketter, J., Legenbauer, T., Millenet, S., Poustka, L., Renner, T., Romanos, M.,

Uebel-von Sandersleben, H., Wenning, J., Ziegler, M., & Döpfner, M. (2021). Disentangling symptoms of externalizing disorders in children using multiple measures and informants. *Psychol. Assess.*, *33*(11), 1065-1079. https://doi.org/10.1037/pas0001053

Thöne, A.-K., Dose, C., Junghänel, Hautmann, C., Jendreizik, L. T., Treier, A.-K., Vetter, P., von Wirth, E., Banaschewski, T., Becker, K., Brandeis, D., Dürrwächter, U., Geissler, J., Hebebrand, J., Hohmann, S., Holtmann, M., Huss, M., Jans, T., Kaiser, A., Ketter, J., Legenbauer, T., Millenet, S., Poustka, L., Renner, T., Romanos, M., Uebel-von Sandersleben, H., Schneider, P., Wenning, J., Ziegler, M., Görtz-Dorten, A., & Döpfner, M. (2023). Identifying symptoms of ADHD and disruptive behavior disorder most strongly associated with functional impairment in children: A symptom-level approach. *J. Psychopathol. Behav. Assess.*, *45*(2), 277–293. https://doi.org/10.1007/s10862-023-10025-z

Ich versichere, dass ich alle Angaben wahrheitsgemäß nach bestem Wissen und Gewissen gemacht habe und verpflichte mich, jedmögliche, die obigen Angaben betreffenden Veränderungen, dem IPHS-Promotionsausschuss unverzüglich mitzuteilen.

Die dieser Dissertation zugrundeliegenden Daten wurden im Rahmen eines von dem Bundesministerium für Bildung und Forschung (BMBF) geförderten Forschungsprojekts an der Klinik und Poliklinik für Psychiatrie, Psychosomatik und Psychotherapie des Kindes- und Jugendalters der Universität zu Köln erhoben (Förderkennzeichen 01EE1408B). Die Daten der multizentrischen, klinischen Studie zur evidenz-basierten, stufenweisen Versorgung von ADHS bei Schulkindern (ESCAschool) wurde an neun Studienstandorten erhoben. ESCAschool ist Teil des Forschungsverbunds ESCAlife.

Beiträge der Erstautorin: Ann-Kathrin Thöne konzipierte die Studiendesign der drei Publikationen, war an der Rekrutierung und Datenerhebung des ESCAschool Studienzentrums in Köln beteiligt, führte die formale Datenanalyse durch, interpretierte und visualisierte die Ergebnisse, integrierte die wissenschaftliche Literatur und schrieb den ersten Entwurf des jeweiligen Manuskriptes. Sie ist Ko-Autorin des Interview-Leitfadens für Externale Störungen (ILF-EXTERNAL) aus den Interview-Leitfäden zum Diagnostik-System für psychische

Störungen nach DSM-5 für Kinder und Jugendliche (DISYPS-ILF; Görtz-Dorten, Thöne, & Döpfner, 2022), auf dem die Ergebnisse der vorliegenden Studien basieren.

Beiträge der Ko-Autor*innen:

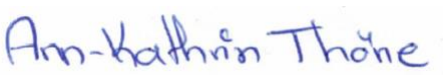
Studienkonzeption (der ESCAschool Studie): Manfred Döpfner, Christopher Hautmann, Elena von Wirth, Tobias Banaschewski, Katja Becker, Daniel Brandeis, Johannes Hebebrand, Sarah Hohmann, Martin Holtmann, Michael Huss, Thomas Jans, Marcel Romanos.

Koordination und Leitung der ESCAschool Studie: Manfred Döpfner, Christina Dose, Lea Teresa Jendreizik, Anna Kaiser, Anne-Katrin Treier, Elena von Wirth, Ute Dürrwächter, Julia Geissler, Johanna Ketter, Tanja Legenbauer, Sabina Millenet, Tobias Renner, Henrik Uebel-von Sandersleben, Priska S. Schneider, Jasmin Wenning.

Rekrutierung der Patient*innen und Datenerhebung: Lea Teresa Jendreizik, Nina Geldermann, Anne-Katrin Treier, Paula Vetter (geb. Altenberger), Ute Dürrwächter, Julia Geissler, Anna Kaiser, Johanna Ketter, Henrik Uebel-von Sandersleben, Priska S. Schneider, Mirjam Ziegler.

Dateninterpretation: Christina Dose, Michaela Junghänel, Manfred Döpfner, Anja Görtz-Dorten.

Alle Autor*innen haben das jeweilige Manuskript kritisch überarbeitet, die letzte Version des Manuskripts abschließend genehmigt und sich bereit erklärt, für alle Aspekte der Arbeit verantwortlich zu sein, um sicherzustellen, dass Fragen im Zusammenhang mit der Genauigkeit oder Integrität eines Teil der Arbeit angemessen untersucht und gelöst werden.

28.05.2024 

Datum Unterschrift