Observation of ADHD Symptoms:
Prospects for a Behavior-Based, Objective and Context-Dependent Assessment

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Leona Hellwig
aus Alzenau in Unterfranken

Tübingen
2017

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	06.07.2017
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Caterina Gawrilow
2. Berichterstatter:	Prof. Dr. Tobias Renner

Danksagung

Allen MitarbeiterInnen der Graduiertenschule LEAD danke ich für die Förderung meiner Promotion, für das gut ausgestattete Forschungsumfeld und für viele wertvolle Lerngelegenheiten.

Caterina Gawrilow danke ich für die Möglichkeit zur Promotion, viel Beratung und Unterstützung und die Freiheit beim Nachdenken über wissenschaftliche Fragestellungen.

Tobias Renner danke ich für sein großes Engagement für interdisziplinäre Kooperationen und die Übernahme des Zweitgutachtens dieser Dissertation.

Meinen LEAD KollegInnen Michele, Juliane, Christina, Daniela und Thomas danke ich, besonders für die guten Zeiten am Neckar.

Ich danke allen Mitgliedern der Arbeitsgruppe Schulpsychologie für den Austausch, die Diskussionen und die gute Teamatmosphäre während meiner Promotionszeit.

Petra Haas und Merle Reuter danke ich, dass wir in den Mittagspausen die wichtigen Fragen des Lebens ansatzweise gelöst haben.

Für ihren Beitrag zu dieser Arbeit im Rahmen von Abschlussarbeiten danke ich Anna Rodner, Alisa Fitzpatrick, Judith Amann, Janna Holst, Mareike Bilger, Anna Mensinger und Lisa Schanbacher.

Ich danke allen Menschen herzlich, die mich während meiner Promotionszeit begleitet haben. Sei es in der Stille, bei Jovano, auf dem Coast Path, nach Vama Veche, im Chiemgau oder nach Hamburg, auf den Wackerstein, nach Sarajevo, in den Schönbuch, beim Zocken, zum Tanzen, zum Georgenhof oder in die Mauerstraße.

Contents

List of Tables

List of Abbreviations

ADHD	Attention deficit hyperactivity disorder
APA	American Psychiatric Association
DGKJP	Deutsche Gesellschaft für Kinder- und Jugendpsychiatrie Psychosomatik und Psychotherapie
DGPPN	Deutsche Gesellschaft für Psychiatrie und Psychotherapie, Psychosomatik und Nervenheilkunde
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, 5th edition
EE	Expressed Emotion
FMSS	Five-Minute Speech Sample
ICC	Intraclass Correlation Coefficient
ICD-10	International Classification of Diseases, 10th revision
WHO	World Health Organisation

List of Figures

Abstract

ADHD is a common child and adolescent psychiatric disorder (Willcutt, 2012) characterized by inattention, hyperactivity and impulsivity (APA, 2013; WHO, 1993). The identification of ADHD symptoms relies on their clinical manifestation, which is reflected in the behavioral descriptions contained in diagnostic manuals (APA, 2013; WHO, 1993). The occurrence of ADHD symptoms depends on the situation (APA, 2013; Burns, Servera, del Mar Bernad, Carrillo, & Geiser, 2014; WHO, 1993). Aside from unstructured clinical exploration, reports on ADHD symptoms are the main source of information on ADHD symptoms (Bundesärztekammer, 2005). Reported ADHD symptoms depict the behavioral descriptions of ADHD symptoms well, and can be considered to be behavior-based. However, reported ADHD symptoms are highly influenced by the informant and thus do not provide objective information (Lienert & Raatz, 1994). Moreover, reported ADHD symptoms usually express a general context-independent behavioral disposition. Information about ADHD symptoms on the cognitive and neurobiological level provides objective information, but is not proximal to behavioral descriptions. Neither reported ADHD symptoms nor measures on the cognitive and neurobiological level represent behavior-based, objective and context-dependent information on ADHD symptoms. Therefore, the present dissertation investigates observation as an assessment method which can fill this gap. To that end, three empirical studies were conducted applying observation techniques to the family context of ADHD symptoms (*Study 1*), ADHD symptoms in a delay of gratification situation (*Study 2*), and ADHD symptoms in a simulated classroom situation (*Study 3*).

The aim of the *Study 1* was to investigate the utility of Expressed Emotion (EE) scales measured by the Five Minute Speech Sample (FMSS) for assessing the family context of children with ADHD symptoms. To this end, inter-rater reliability as well as the association with the severity of ADHD symptoms in German school-aged children were assessed. Thirty-three children (19 female, $M(SD)_{age}$ = 10.65(1.34) years, n = 6 diagnosed with ADHD according to parental reports) took part in the study along with one of their parents. Parents and children gave information on ADHD severity. FMSS was obtained from one parent as part of a telephone interview. The inter-rater reliability of the EE scales for relationship and critical and positive comments were sufficient (ICC_{single} = .63 - .67). The EE scales critical and positive comments showed consistent associations with all ADHD symptom measures in parental and self-reports. The results

show that EE as measured by FMSS is a promising tool to capture the family context of children with ADHD symptoms. However, the utility of EE is limited by insufficient delimitation of the concept and therefore requires further development.

The aim of *Study 2* was to investigate which variables produce performance differences between children with ADHD and their healthy peers in delay of gratification tasks. Concretely, *Study 2* investigated whether observed and reported ADHD symptoms explain performance differences in a delay of gratification task over and above attention orientation. Sixty-one children (14 female, $M(SD)_{age}$=10.40(1.58) years, 26 diagnosed with ADHD) participated in a video-recorded delay of gratification task. Videos were rated with regard to attention orientation, activity, and impulsivity during the delay. Ten children did not wait for the delayed reward. Attention orientation and observed activity during the waiting situation predicted performance in the delay of gratification task over and above an ADHD diagnosis and observed impulsivity.

The aim of *Study 3* was to investigate the utility and feasibility of an observation protocol for ADHD symptoms in a simulated classroom situation. Thirty-five children (20 female, $M(SD)_{age}$ = 10.67(1.36) years) took part in a video-recorded simulated classroom situation, which comprised a math test and a competitive card game. Both the children and one parent gave reports on ADHD symptom severity. Videos were analyzed with regard to inattention, hyperactivity and impulsivity. Inter-rater reliability of observed inattention during the math test and observed impulsivity during the competitive card game was satisfactory. Inter-rater reliability of hyperactivity items was partly sufficient. Associations between observed and reported ADHD symptoms were mostly medium. The video rating items for inattention were strongly negatively associated with performance in the math test. Further development of the implemented observation protocol promises to provide valuable information on ADHD symptoms at school that exceeds information gained through reported symptoms.

The results of the empirical studies show that behavioral observation provides behavior-based, objective and context-dependent information on ADHD symptoms, in accordance with expectations. Observed ADHD symptoms are associated with meaningful outcomes, and provide information beyond reported ADHD symptoms. Observation is a costly and complex assessment method. Implementation in practice is only justifiable if standardized observation protocols are developed that reduce the effort involved. Moreover, future research should develop diagnostic routines that allow

for the integration and use of discrepant information from different diagnostic sources, for instance observation and reports.

Introduction

ADHD comprises symptoms of inattention, hyperactivity and impulsivity behavior. Behavioral descriptions of inattentive, hyperactive and impulsive behavior in diagnostic manuals are the anchor point for the identification of ADHD symptoms (APA, 2015; WHO, 2011). Thus, the identification of symptoms relies on the clinical manifestation rather than on etiological psychological or physiological features of the disorder (Barkley, 1997a). In recent decades, the search for psychological and physiological features that would allow ADHD symptoms to be identified independently of their clinical manifestation has resulted in fruitful developments with regard to the psychological or physiological theoretical background of ADHD (Barkley, 1997a; Sonuga-Barke, 2002; Van der Meere, 2005). However, this theoretical background addresses psychological and physiological features of limited diagnostic utility, meaning that clinical manifestation remains the anchor point for the identification of ADHD symptoms (Tannock, 2013). The reliance on clinical manifestation for the identification of ADHD symptoms has important implications for assessment methods. First, the importance of assessment methods that provide information proximal to the behavioral descriptions of ADHD symptoms is stressed. Second, the context in which the symptoms develop is stressed because the behavioral descriptions usually mention specific contexts. To account for the proximity to concrete behavior and the context dependence, this thesis investigates observation as an important assessment instrument for the behavior-based, objective and context-dependent assessment of ADHD symptoms.

The present dissertation is structured in four chapters as follows. After this short introduction, chapter 1 provides theoretical background to support the notion that a behavior-based, objective and context-dependent assessment of ADHD symptoms is needed. Chapter 2 identifies the research aims of the present dissertation. Chapter 3 then describes three empirical studies conducted in accordance with these research aims. The subject of investigation of these studies is the assessment of the family context as a meaningful context for children with ADHD symptoms (*Study 1*, Chapter 3.1) as well as the observation of ADHD symptoms in a lab-based delay of gratification situation (*Study 2*, Chapter 3.2) and in an ecological valid classroom context (*Study 3*, Chapter 4.3). Chapter 4 comprises a general discussion of the results of the empirical studies with respect to the research aims.

1 THEORETICAL BACKGROUND

The overarching goal of the present dissertation is to investigate the potential of observation of ADHD symptoms as a behavior-based, objective and context-dependent assessment instrument. The theoretical background is divided into two sections. The first section (see 1.1) describes the nature of ADHD as the objective of the present dissertation. First, the description of ADHD symptoms in the diagnostic manuals as well as the prevalence and persistence of the disorder are summarized, and ADHD is introduced as a dimensional construct (see 1.1.1). Thereafter, etiological factors for ADHD on the neurobiological and psychological level are explained (see 1.1.2). The first section ends with a summary of German professional associations' (DGKJP & DGPPPN, 2016) diagnostic guidelines (see 1.1.3). In the second section, assessment methods of ADHD symptoms are reviewed in terms of their objectivity and proximity to behavior as well as the context-dependence of obtained information. Reviewing the information about ADHD symptoms obtained by reports and on the cognitive or neurobiological level revealed that available information is either objective or behavior-based, but not both (see 1.2.1). Therefore, observation as an assessment method which is both objective and behavior-based is introduced (see 1.2.2). Moreover, reasons for the context dependence of ADHD symptoms are summarized, with the conclusion that current assessment methods do not provide context-dependent information on ADHD symptoms (see 1.2.3). Thus, observation is introduced at the end of Chapter 2 as an assessment method which has the potential to provide context-dependent information on ADHD symptoms (see 1.2.4).

1.1 Attention deficit hyperactivity disorder

Assessing ADHD symptoms is the objective of the present dissertation. Therefore, the following section provides an overview of the most important features of ADHD symptoms. First, descriptions of ADHD in diagnostic manuals are summarized. After that, estimates of ADHD's prevalence and its persistence into adulthood are discussed. Finally, ADHD symptoms are introduced as a dimensional construct.

1.1.1 Phenomenology of ADHD symptoms

Two official diagnostic manuals that contain behavioral descriptions currently exist. The ICD-10 is the 10th revision of the classification of diseases developed by the World Health Organization (WHO, 2011). It is used worldwide, particularly in clinical practice and in health insurance providers' accounting systems (WHO, 1993, 2011). The DSM-5 is the fifth revision of the classification system for mental disorders developed by the American Psychiatric Association. It is predominantly used in research (APA, 2013, 2015). Diagnostic criteria for ADHD symptoms are explained in the following paragraphs with reference to the English versions of the ICD-10 diagnostic criteria for research (WHO, 1993) and the DSM-5 (WHO, 1993).

Characteristics of ADHD in ICD-10

In the ICD-10 (WHO, 1993), diagnostic codes for ADHD are subsumed under "behavioral and emotional disorders with onset usually occurring in childhood and adolescence" (F90-98). The diagnostic code "F90 hyperkinetic disorders" lists behavioral descriptions of inattentive, hyperactive and impulsive symptoms. The ICD-10 differentiates between a *disturbance of activity and attention* (F90.0) that excludes conduct disorder and *hyperkinetic conduct disorder* (F90.1), for which the criteria for both hyperkinetic disorders (F90) and conduct disorder have to be met. Furthermore, the ICD-10 includes the residual categories *other hyperkinetic disorder* (F90.8), and *hyperkinetic disorders, unspecified* (F90.9). Apart from hyperkinetic disorders (F.90), a residual category *other specified behavioral and emotional disorder with onset usually occurring in childhood and adolescence* (F98.8) captures attention disorder without hyperactivity. A diagnosis requires (a) a pervasiveness of symptoms across settings and time, (b) the inappropriateness of inattention, hyperactivity and impulsivity for the age of the child, (c) the direct observation of symptoms (apart from teacher and parental reports) or significant impairment on psychometric tests on attention, (d) onset before the age of seven years, (e) a duration of at least six months, and (f) an IQ above 50. Moreover, pervasive developmental disorders (F84), mania (F30), and depression (F32) cannot be diagnosed together with ADHD. The descriptions of symptoms differentiate between occurrence at *home* and in a *school or nursery* setting. For a diagnosis, the following number of symptoms have to be present: For inattention criteria, three at home and two in a school or nursery setting; for hyperactivity criteria, three at home and three in a school or nursery setting; and one impulsivity criterion at home. Tables 1 and 2 list all criteria in their original wording (F90).

Characteristics of ADHD in DSM-5

In the DSM-5, ADHD is listed in the chapter on neurodevelopmental disorders. The DSM-5 has been released in 2013 and is the most current diagnostic manual. Important changes from the fourth revision to the current fifth revision are: the elimination of subtypes and the introduction of presentations, the introduction of reduced symptom thresholds for persons older than 17, the change of the age of onset from seven to twelve, the possibility to diagnose autism spectrum disorders together with ADHD, and the requirement for multiple informants (Tannock, 2013). According to DSM-5, a diagnosis can be specified as a *combined presentation* (314.01), a *predominantly inattentive presentation* (314.00) or a *hyperactive/impulsive presentation* (314.01). These presentations can be further specified with the attributes mild, moderate or severe. A diagnosis requires symptoms (a) to have persisted for at least six months, (b) to be inappropriate for the age of the child, (c) to not be a manifestation of oppositional behavior, defiance, hostility, or failure to understand tasks or instructions, (d) to be present prior the age of 12 years, (e) to be present in two or more settings, (f) to reduce social, academic or occupational functioning. Moreover, (g) symptoms cannot be better explained by another mental disorder. For persons younger than 17, six or more of the inattention criteria have to apply, as do six or more of the hyperactivity-impulsivity criteria. For persons older than 17, five of the inattention criteria and five of the hyperactivity-impulsivity criteria have to apply. Tables 1 and 2 list all criteria in their original wording.

Comparison of characteristics of ADHD symptoms in ICD-10 and DSM-5

The definitions of ADHD symptoms in ICD-10 and DSM-5 show differences and similarities. The age of onset is an important difference; in ICD-10, symptoms have to be present before the age of seven, whereas in DSM-5 before the age of 12. Furthermore, the exclusion of diagnoses that can be diagnosed together with ADHD is stricter in ICD-10 than in DSM-5. Moreover, hyperactivity and impulsivity are combined into one category in DSM-5, whereas in ICD-10 they are separated. The number of symptoms that have to be present for a diagnosis varies between the DSM-5 and ICD-10. Similarities between ICD-10 and DSM-5 are the behavioral descriptions of symptoms (see Table 1 and 2). In both ICD-10 and DSM-5, symptoms are required to be present in different settings and last at least six months.

Prevalence of ADHD

The general prevalence of ADHD in children and adolescents ranges from around 3.4 % to 7.2 % (Faraone et al., 2015; Polanczyk, 2007; Polanczyk, Salum, Sugaya, Caye, & Rohde, 2015; Thomas, Sanders, Doust, Beller, & Glasziou, 2015; Willcutt, 2012). However, a number of factors significantly influence these prevalence estimates. These factors are (1) diagnostic procedure, (2) the assessment of functional impairment, (3) socio-economic factors, (4) geographical location, (5) sex and (6) race and ethnicity. The (1) diagnostic procedure is an important moderator of the prevalence estimate. Single informant diagnoses are more frequent than diagnoses from several informants (Willcutt, 2012). In the same vein, the type of diagnostic interview influences the prevalence estimate (Polanczyk et al., 2015). Moreover, revisions of the DSM have entailed changes in the age of onset over the years, which have caused changes in prevalence rates (Faraone et al., 2015; Thomas et al., 2015). The DSM requires ADHD to cause (2) functional impairment (reduction of social, academic or occupational functioning, D-criterion, APA, 2013). Although, it is difficult to measure ADHD-specific functional impairment, studies that have included measures of functional impairment show slightly reduced prevalence estimates (Polanczyk et al., 2015; Willcutt, 2012). Lower (3) socio-economic status seems to increase prevalence estimates (Willcutt, 2012). A child's (4) geographical location, meaning the country in which he or she lives does not have a significant impact on prevalence estimates (Polanczyk, 2007; Willcutt, 2012). Across all studies, (5) sex has a large influence on prevalence estimates. The probability of a boy receiving an ADHD diagnosis is two to four times higher than that of a girl (Faraone et al., 2015; Willcutt, 2012). Studies on the influence of (6) race and ethnicity on prevalence estimates are rare. Current evidence does not reveal a significant influence of race and ethnicity on prevalence estimates (Faraone et al., 2015; Willcutt, 2012).

Persistence and development of ADHD symptoms

ADHD is a disorder with onset in childhood. Therefore, whether symptoms persist into adulthood is an important question. The literature reveals that ADHD is a disorder which frequently persists into adulthood (Faraone et al., 2015). However, prevalence estimates for adults are lower than for children (Willcutt, 2012); thus, some people who meet the criteria for a diagnosis in childhood do not meet the criteria in adulthood. Predictors for a persistent ADHD diagnosis are family members who suffer from ADHD, adverse psychosocial conditions, impairment caused by ADHD and comorbid disorders (Biederman et al., 1996; Biederman, Petty, Clarke, Lomedico, &

Faraone, 2011). Importantly, symptom expression changes from childhood to adulthood, with hyperactivity and impulsivity are more likely to decline than inattention (Biederman, 2000). Symptoms of hyperactivity and impulsivity are internalized and can be described as inner restlessness (see examples for adolescents in DSM-5 criteria Tables 1 and 2, Faraone et al., 2015; Millstein, Wilens, Biederman, & Spencer, 1998).

ADHD as a dimensional construct

ADHD is defined in the diagnostic manuals as a category, with a distinction made between persons with and without ADHD. This categorization implies that ADHD as a psychological construct is either present in a person or not. Assessment methods should therefore be required to classify people as either persons who suffer from ADHD or healthy persons. Naturally, this is not the case. Empirical studies suggest that ADHD symptoms are spread dimensionally rather than categorically (Balázs & Keresztény, 2014; Coghill & Sonuga-Barke, 2012). Everyone can be arranged on a spectrum from fewer to more ADHD symptoms. In line with this, assessment instruments usually capture ADHD symptoms dimensionally (Lidzba, Christiansen, & Drechsler, 2013) and define cut-off values. Thus, categorization into persons with and without ADHD is important for practical reasons, as it allows a subgroup of people who need treatment to be defined. Moreover, the importance of the dimensional perspective on ADHD symptoms becomes clear when taking a look at the impairment that comes along with ADHD symptoms. The severity of ADHD symptoms is associated with the severity of psychosocial problems (Norén Selinus et al., 2016) in the subclinical range, too. For instance, the association between impairment and subclinical ADHD symptoms is evident for academic achievement (Merrell & Tymms, 2001) and unfavorable family climate (Schloß et al., 2015). Whether a given study should take a categorical or dimensional approach to ADHD symptoms depends on its design and research question. In the present dissertation, ADHD symptoms are approached both dimensionally (*Studies 1* and *3*) and categorically (*Study 2*). In the present dissertation, the term ADHD symptoms always refers to a dimensional approach, whereas ADHD always refers to the diagnosis, meaning a categorical approach.

In sum, ADHD symptoms are defined by behavioral descriptions in the current diagnostic manuals. ADHD is a frequently occurring disorder that often persists into adulthood. ADHD is best modeled as a dimensional construct. The official diagnostic

manuals do not refer to etiologic factors at all. Thus, etiology has to be looked at separately in the next section.

Table 1
ICD-10 and DSM-5 Diagnostic Criteria for ADHD (Part1)

ICD-10		DSM-5	
Inattention		**Inattention**	
1	Short duration of spontaneous activities (*home*)	1	Often fails to give close attention to details or makes careless mistakes in schoolwork, at work, or during other activities (e.g., overlooks or misses details, work is inaccurate).
2	Often leaving play activities unfinished (*home*)	2	Often has difficulty sustaining attention in tasks or play activities (e.g., has difficulty remaining focused during lectures, conversations, or lengthy reading).
3	Over-frequent changes between activities (*home*)	3	Often does not seem to listen when spoken to directly (e.g., mind seems elsewhere, even in the absence of any obvious distraction).
4	Undue lack of persistence at tasks set by adults (*home*)	4	Often does not follow through on instructions and fails to finish schoolwork, chores, or duties in the workplace (e.g., starts tasks but quickly loses focus and is easily sidetracked).
5	Unduly high distractibility during study e.g. homework or reading assignment (*home*)	5	Often has difficulty organizing tasks and activities (e.g., difficulty managing sequential tasks; difficulty keeping materials and belongings in order; messy, disorganized work; has poor time management; fails to meet deadlines).
6	Undue lack of persistence at tasks (*school/nursery*)	6	Often avoids, dislikes, or is reluctant to engage in tasks that require sustained mental effort (e.g., schoolwork or homework; for older adolescents and adults, preparing reports, completing forms, reviewing lengthy papers).
7	Unduly high distractibility, i.e. often orienting towards extrinsic stimuli (*school/nursery*)	7	Often loses things necessary for tasks or activities (e.g., school materials, pencils, books, tools, wallets, keys, paperwork, eyeglasses, mobile telephones).
8	Over-frequent changes between activities when choice is allowed (*school/nursery*)	8	Is often easily distracted by extraneous stimuli (for older adolescents and adults, may include unrelated thoughts).
9	Excessively short duration of play activities (*school/nursery*)	9	Is often forgetful in daily activities (e.g., doing chores, running errands; for older adolescents and adults, returning calls, paying bills, keeping appointments).

Note. ICD-10 (WHO, 1993), DSM-5 (APA, 2013)

Table 2
ICD-10 and DSM-5 Diagnostic Criteria for ADHD (Part 2)

ICD-10	DSM-5
Hyperactivity	**Hyperactivity-Impulsivity**
1 Very often runs about or climbs excessively in situations where it is inappropriate; seems unable to remain still (*home*)	1 Often fidgets with or taps hands or feet or squirms in seat.
2 Markedly excessive fidgeting & wriggling during spontaneous activities (*home*)	2 Often leaves seat in situations when remaining seated is expected (e.g., leaves his or her place in the classroom, in the office or other workplace, or in other situations that require remaining in place).
3 Markedly excessive activity in situations expecting relative stillness (e.g. mealtimes, travel, visiting church) (*home*)	3 Often runs about or climbs in situations where it is inappropriate. (Note: In adolescents or adults, may be limited to feeling restless.)
4 Often leaves seat in classroom or other situations when remaining seated is expected (*home*)	4 Often unable to play or engage in leisure activities quietly.
5 Often has difficulty playing quietly (*home*)	5 Is often "on the go," acting as if "driven by a motor" (e.g., is unable to be or uncomfortable being still for extended time, as in restaurants, meetings; may be experienced by others as being restless or difficult to keep up with).
6 Continuous (or almost continuous) and excessive motor restlessness (running, jumping, etc.) in situations allowing free activity (*school or nursery*)	6 Often talks excessively.
7 Markedly excessive fidgeting and wriggling in structured situations (*school or nursery*)	7 Often blurts out an answer before a question has been completed (e.g., completes people's sentences; cannot wait for turn in conversation).
8 Excessive levels of off-task activity during tasks (*school or nursery*)	8 Often has difficulty waiting his or her turn (e.g., while waiting in line).
9 Unduly often out of seat when required to be sitting (*school or nursery*)	9 Often interrupts or intrudes on others (e.g., butts into conversations, games, or activities; may start using other people's things without asking or receiving permission; for adolescents and adults, may intrude into or take over what)
10 Often has difficulty playing quietly (*school or nursery*)	
Impulsivity	
1 Often has difficulty awaiting turns in games or group situations (*home*)	
2 Often interrupts or intrudes on others (e.g. butts in to others' conversations or games) (*home*)	
3 Often blurts out answers to questions before questions have been completed (*home*)	

1.1.2 Etiology

Questions on the etiology of ADHD can be addressed from different perspectives. The following section summarizes etiological approaches on the neurobiological and psychological level.

Etiological approaches on the neurobiological level

ADHD is a disorder with strong heritability. Twin and adoption studies suggest a heritability of .80 (scale 0-1, Banaschewski, Roessner, Uebel, & Rothenberger, 2004; Faraone et al., 2005; Faraone & Biederman, 1998; Zhou et al., 2008). Genes attributable to the expression of dopamine receptors seem to be relevant for ADHD (Faraone et al., 2005). On a neurological level, a recent meta-analysis depicts structural brain abnormalities in ADHD compared to controls. Patients with ADHD show reduced grey matter volume in the basal ganglia, the insula, prefrontal cortex, orbitofrontal cortex, anterior cingulate cortex and occipital lobe. In addition to these structural abnormalities, ADHD patients also show functional abnormalities. Research findings point to underactivation during inhibitory tasks in the basal ganglia, prefrontal cortex and insula, in particular (Norman et al., 2016). In summary, the neurobiology of persons with ADHD shows distinctive features in the genome as well as brain structure and function. These results give some indication of initial approaches to neurobiological etiologic factors of ADHD but do not provide a conclusive picture.

Etiological approaches on the psychological level

Key psychological theories of ADHD are the executive dysfunction theory, pathway theories and state regulation theory (Johnson, Wiersema, & Kuntsi, 2009). As diagnostic criteria merely provide a description of the clinical manifestation of ADHD, these theories try to specify the nature of ADHD and isolate psychological causes of the disorder (Barkley, 1997a).

The *executive dysfunction theory* supposes that ADHD symptoms develop due to impairments in behavioral inhibition as one part of executive functioning, specifically the inhibition of prepotent responses, which stop ongoing responses to feedback on errors and interference control. Deficits in behavioral inhibition lead to deficits in working memory, the self-regulation of affect, motivation, arousal and reconstitution of novel behavior (Barkley, 1997b). This theory is corroborated by empirical evidence. For instance, persons with ADHD perform worse in the stop-signal task, a widely used

measure of behavioral inhibition, than persons without ADHD (Alderson, Rapport, & Kofler, 2007). In the same vein, a meta-analytic review shows consistently worse performance in multiple measures of executive functioning in persons with ADHD compared to persons without ADHD, particularly in measures of response inhibition, vigilance, working memory, and planning. However, despite these group differences in measures of executive functioning, deficient executive functioning can be only one piece of the jigsaw explaining ADHD. Deficient executive functions are present in some but not all children with ADHD (Willcutt, Doyle, Nigg, Faraone, & Pennington, 2005). This explains the group differences between persons with and without ADHD, but highlights that executive dysfunction is not a necessary precondition for ADHD. Moreover, effect size measures for group differences (ADHD, non-ADHD) in ADHD symptoms are much bigger than group differences in executive functioning measures (Alderson et al., 2007). In sum, executive dysfunction theory has been confirmed empirically and explains important parts of ADHD. However, it is not exhaustive, as not all children with ADHD have an inhibition deficit. This fact resulted in the subsequent development of the pathway models.

The *pathway models* have been developed to account for the psychological heterogeneity of ADHD. The dual pathway model suggests that ADHD symptoms develop due to an inhibitory dysfunction resulting in a disorder of thought and action (inhibitory dysfunction pathway) *as well as* due to a motivational style characterized by a preference for immediacy and an aversion of delay (delay aversion pathway). These two pathways contribute independently to the occurrence of ADHD symptoms (Sonuga-Barke, 2002, 2003). Evidence for the inhibitory dysfunction pathway has already been mentioned with regard to executive dysfunction theory (Alderson et al., 2007; Willcutt et al., 2005). The delay aversion pathway is based on a suboptimal reward process that can be described as a preference for immediacy. Children with delay aversion try to escape delay, exhibiting impulsive behavior when choices are available and hyperactive and inattentive behavior when no choices are available. Evidence for a delay aversive response pattern in children with ADHD has been found across different age groups (Patros et al., 2016; Pauli-Pott & Becker, 2011). An important implication of the delay aversion pathway is the inclusion of environmental factors in a theory of ADHD. Parenting style and environmental responses to a child's preference for immediacy influence the degree to which a child develops delay aversion (Sonuga-Barke, 2003). Another important contribution of the dual pathway model is the notion that multiple

and different psychological foundations can lie at the root of ADHD with the same clinical picture, and the clinical picture might not be due to one core deficit. Recently, deficits in timing processing have been suggested as another independent etiologic factor leading to ADHD (Sonuga-Barke, Bitsakou, & Thompson, 2010). Nevertheless, despite the empirical underpinnings and the comprehensiveness of the pathway models, they mostly fail to explain the frequently observed behavioral variability of ADHD symptoms (Kofler, Rapport, & Matt Alderson, 2008; Rapport, Kofler, Alderson, Timko, & Dupaul, 2009; Tamm et al., 2012).

The most recent theoretical model, *state regulation theory,* directly deals with the behavioral variability of ADHD (Van der Meere, 2005). Attempting to explain variability across situations requires context to be incorporated into the explanatory model. State regulation theory builds on the cognitive energetic model (Sander, 1983). This model combines a cognitive processing stage and an energetic stage to predict task efficiency. Thus, task efficiency is explained by the processing of information and the activation and arousal provided. If arousal and activation are provided in accordance with current needs, task efficiency will be high. If there is a mismatch between activation and arousal and current needs, task efficiency will be low. The state regulation theory of ADHD assumes that persons with ADHD suffer from deficits in the ability to create a fit between the activation and arousal needs resulting from the cognitive processing stage and the actually provided activation and arousal resulting from the energetic stage. This notion has been empirically confirmed by research on the difficulties adolescents with ADHD have balancing the speed-accuracy trade-off (Mulder et al., 2010), for example. Persons with ADHD are prone to underactivation. Thus, state regulation, i.e. adaptation to task requirements, often fails for them in situations in which stimulation is low, resulting in low activation when tasks require high activation. Evidence confirms this notion. Event-rate studies show that children with ADHD perform particularly worse under a low stimulation rate (Van der Meere, 2002). A classroom observation study revealed that children with ADHD showed off-task behavior during idle time (low stimulation) rather than during structured classroom activities (high stimulation, Imeraj et al., 2016). State regulation theory explains task efficiency via the current match or mismatch between demanded and provided activation and arousal. This element incorporates the context, because activation and arousal demands vary with context. Therefore, the theory accounts for behavioral variability in ADHD (Kofler et al., 2008; Rapport et al., 2009; Tamm et al., 2012) and

can be contrasted with general deficit theories (Van der Meere, 2002; Willcutt et al., 2005). In this vein, it has been proposed that attention deficit disorders might not be the correct diagnostic term; state regulation disorder would be more appropriate (Van der Meere, 2002). However, this theory is difficult to empirically test, because the optimal fit between demanded and provided activation and arousal varies by person and situation and is difficult to operationalize (Johnson et al., 2009).

In sum, the variety of etiologic approaches implies that the different etiological theories approach the same construct, ADHD, from different perspectives, as for instance the neurobiological, cognitive and behavioral perspective (Johnson et al., 2009). Therefore, the following sections go into detail regarding the quality of information on ADHD symptoms obtained on neurobiological, cognitive and behavioral level (see 1.1.3 and 1.2).

1.1.3 Diagnostic process

So far, ADHD criteria and the phenomenology of ADHD symptoms have been explained. This section identifies diagnostic methods for ADHD symptoms. Professional associations provide official guidelines on methods that should be used in the diagnostic process for ADHD. Here, the official guidelines of the German Medical Association (German: Bundesärztekammer) and the German Association for Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy (German: Deutsche Gesellschaft für Kinder- und Jugendpsychiatrie Psychosomatik und Psychotherapie, DGKJP) are reviewed.

The German Medical Association published guidelines on the diagnosis, treatment and course of the disorder in 2005 (Bundesärztekammer, 2005). To exclude other disorders with a similar clinical manifestation, an in-depth *differential diagnosis* is important. Important questions are: Can ADHD symptoms be traced back to medication? Can a neurological disorder, such as epilepsy or a traumatic brain injury, explain the symptoms? Does the patient suffer from a pervasive developmental disorder? Can inattentive and hyperactive-impulsive behavior be distinguished from oppositional behavior? Especially in adolescents, it is important to consider that inattentive and hyperactive-impulsive symptoms may be related to other disorders (depression, anxiety disorder or psychosis). In these cases, the onset of ADHD symptoms was usually not during childhood. Moreover, comorbid disorders, especially

tic disorder, emotional disorder, conduct disorder and developmental disorders, have to be taken into consideration. The German Medical Association mentions clinical exploration, questionnaire instruments, neuropsychological tests, organic examination and behavioral observation as diagnostic methods.

During *clinical exploration*, the diagnostician has a conversation with the caregiver about symptoms, intensity, frequency, and situational variability, the development of symptoms, comorbid conditions and psychosocial circumstances. Older children can provide information on symptoms themselves. Clinical explorations are usually unstructured. Structured clinical interviews can be used, but are not common in clinical practice (Kinder-DIPS, Schneider, Unnewehr, & Margraf, 2009).

Standardized questionnaire, for instance reports, complement the information obtained during clinical exploration. For ADHD symptoms, information from teachers and caregiver (usually parents) should be obtained. Self-report questionnaires on ADHD symptoms exist for children around age 11 and above, but are often less informative compared to teacher reports and parental reports. Standardized questionnaires like the Conners-3 rating scales (Lidzba et al., 2013) provide information on the fulfillment of diagnostic criteria according to DSM-5 and ICD-10 (APA, 2013; WHO, 2011). Aside from specific questionnaires, the Child Behavior Checklist (parental report), which captures general psychopathology in children and adolescents, is widely used, as are self-report and teacher report forms (Achenbach et al., 2008).

The German Medical Association advises against conducting *neuropsychological tests* that assess impulsivity, attention and executive functions during the diagnostic process. These tests provide information on specific cognitive functions, but their classification accuracy is too low for the identification of ADHD cases. An intelligence test, however, might be in order to check the appropriateness of a child's schooling arrangements, with the test situation providing an opportunity to observe ADHD symptoms.

For differential diagnostic purposes, *organic examinations*, especially neurological examinations, are recommended to screen for organic causes of symptoms similar to ADHD symptoms, such as epilepsy, for instance. It is not recommended that physiological signs be used in the identification of ADHD itself.

The guidelines mention, that *behavioral observation* of ADHD symptoms during clinical exploration, neuropsychological tests and physical examinations can also be included in assessments of ADHD symptoms. However, symptoms are often diminished

in novel situations like examination situations in hospitals or in the offices of health practitioners. Therefore, the absence of observed symptoms during examinations does not indicate the absence of relevant ADHD symptoms. Furthermore, the guidelines mention very briefly that observations in natural environments are desirable, but often not feasible. Video-recorded behavioral observations in meaningful situations (dinner, homework) in natural environments are recommended.

Aside from the German Medical Association, the German Association for Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy has also developed guidelines for the diagnosis of ADHD symptoms. The guidelines are currently (spring 2017) under review for further development. The former guidelines (DGKJP, 2007) are outdated, but the current guidelines have not been published yet (DGKJP, 2016). In line with the German Medical Association guidelines, the previous DGKJP guidelines cite clinical exploration with caregivers and teachers and standardized questionnaires as the most crucial means of acquiring information on ADHD symptoms (DGKJP, 2007). In the same vein, the unpublished guidelines, which are currently in the developmental process, stress the importance of clinical exploration. As a complement, diagnoses can also be based on standardized questionnaires and behavioral observation during test sessions or in natural environments (DGKJP, 2016).

In sum, the methods recommended by professional associations in Germany for obtaining information on ADHD symptoms are first of all clinical exploration with primary caregivers and teachers and standardized questionnaires. The official guidelines mention behavioral observation, but do not attach great importance to it. Importantly, the quality and nature of the information obtained via different assessment methods is likely to differ quite a bit. For instance, it is intuitively plausible that observed ADHD symptoms do not provide the same kind of information on ADHD symptoms as an anecdotal experience told by a parent during the clinical exploration. Therefore, the following section looks beyond officially recommended assessment methods. It reviews information on ADHD symptoms obtained through reports and on the cognitive and neurobiological level with regard to the following criteria: proximity to behavioral descriptions, objectivity and context dependence.

1.2 Review of assessment methods

The first section of the theoretical portion of this dissertation depicted the nature of ADHD symptoms, etiological factors and professional associations' recommendations regarding the diagnostic process. Questions regarding the quality of information on ADHD symptoms obtained via different assessment methods remained open. How does information on ADHD symptoms obtained via different methods differ? This section reviews several assessment methods, namely reported ADHD symptoms as well as information on the cognitive and neurobiological level, in order to answer this question.

1.2.1 Objective and behavior-based information on ADHD symptoms

In evaluating behavioral descriptions of ADHD, two questions seem to be particularly relevant. First, the *proximity* between the information obtained and the *behavioral descriptions* of ADHD symptoms in the diagnostic manuals (APA, 2013; WHO, 1993) is relevant. To what extent does the information on ADHD symptoms encapsulate the behavioral descriptions? Behavioral descriptions focus on concrete behavior. In order to evaluate the proximity of the information to behavioral descriptions, we can ask the question: How *behavior-based* is the obtained information on ADHD symptoms? Second, the *objectivity* of the information on ADHD symptoms is pivotal. For psychological instruments, objectivity is defined as the independence of the informant and the instrument. That means, that every informant who uses the instrument gives the same or similar answers (Lienert & Raatz, 1994). Objectivity is a desirable quality for assessment methods. Information that is not objective is biased through the informant, and because of this confounding it is impossible to determine which information refers to ADHD symptoms and which to the informant. Therefore, the second evaluation criterion is the question: How *objective* is the obtained information on ADHD symptoms?

Assessment of ADHD symptoms via report

Reports on ADHD symptoms are based on answers by the concerned person (self-report), caregivers (usually parental report) or a third party (usually teacher report) to the question of to what extent a description of ADHD symptoms applies to the concerned person. The items of rating scales refer to ICD-10 or DSM-5 criteria. In

German speaking countries, two comprehensive rating scale systems for children and adolescents exist. The DYSIPS III (Döpfner & Görtz-Dorten, 2016) provides rating scales including standardized values for self-reports (SBB-ADHS), parental or teacher reports (FBB-ADHS), and health practitioners (IFL-EXTERNAL) for different age groups. The second rating scale system is the Conners-3 rating scales (Lidzba et al., 2013), which similarly provide comprehensive parental, teacher and self-report scales with standardized values.

Do these reports provide behavior-based information on ADHD symptoms? To address this question, ten items from the ADHD index, which is part of the Conners-3 scales (Lidzba et al., 2013) were selected for purposes of illustration. These 10 items differentiated best between children diagnosed with ADHD and children without an ADHD diagnosis. To evaluate the proximity of reported ADHD symptoms to behavioral descriptions, items from the ADHD index were matched to DSM-5 criteria (see Table 3). Although the behavioral descriptions in the diagnostic manuals are more detailed, the ADHD index items directly match the DSM-5 criteria. Therefore, the rating scale items are directly inspired by behavioral descriptions and can be considered a behavior-based assessment method.

Table 3
ADHD Index Items from Conners Scales and Corresponding DSM-5 Criteria

ADHD index	DSM-5 criteria
Fidgeting.	Often fidgets with or taps hands or feet or squirms in seat.
Does not seem to listen to what is being said to him/her.	Often does not seem to listen when spoken to directly (e.g., mind seems elsewhere, even in the absence of any obvious distraction).
Doesn't pay attention to details; makes careless mistakes.	Often fails to give close attention to details or makes careless mistakes in schoolwork, at work, or during other activities (e.g., overlooks or misses details, work is inaccurate).
Inattentive, easily distracted.	Often has difficulty sustaining attention in tasks or play activities (e.g., has difficulty remaining focused during lectures, conversations, or lengthy reading).
Has trouble organizing tasks or activities.	Often has difficulty organizing tasks and activities (e.g., difficulty managing sequential tasks; difficulty keeping materials and belongings in order; messy, disorganized work; has poor time management; fails to meet deadlines).
Gives up easily on difficult tasks.	Often avoids, dislikes, or is reluctant to engage in tasks that require sustained mental effort (e.g., schoolwork or homework; for older adolescents and adults, preparing reports, completing forms, reviewing lengthy papers).

Fidgets or squirms in the seat.	Often fidgets with or taps hands or feet or squirms in seat.
Restless or overactive.	Is often "on the go," acting as if "driven by a motor" (e.g., is unable to be or uncomfortable being still for extended time, as in restaurants, meetings; may be experienced by others as being restless or difficult to keep up with).
Is easily distracted by sights or sounds.	Is often easily distracted by extraneous stimuli (for older adolescents and adults, may include unrelated thoughts).
Interrupts others (for example, butts into conversations or games)	Often interrupts or intrudes on others (e.g., butts into conversations, games, or activities; may start using other people's things without asking or receiving permission; for adolescents and adults, may intrude into or take over what)

Note. ADHD Index parental report (10 item scale from the Conners rating scales, 3rd edition; (Conners, 2011), and DSM-5 criteria which match best with regard to content.

How *objective* is the obtained information on ADHD symptoms? To answer the question of objectivity regarding reported ADHD symptoms, three important points are explained in more detail. First, findings from intervention evaluations are pointed out. Second, findings on informant discrepancies are pointed out, and third, psychological processes that influence the answers to questionnaire questions are explained.

First, evaluating intervention effects is complicated. The relevant outcome measure for interventions is in most cases reported ADHD symptom severity. It is important to distinguish expectancy effects in changes observed between pre and post intervention measurement points from real changes in symptom severity. To reach that goal, raters of ADHD symptoms should be blinded to the fact that a child underwent any kind of intervention. This is difficult to achieve in practice, because caregivers or teachers are usually involved in organizing the study or the child tells them about taking part in the intervention. In meta-analyses that evaluate the effectiveness of interventions, raters of ADHD symptoms severity are usually differentiated into *raters most proximal* to the therapeutic setting, and *probably blinded raters*. Probably blinded raters are usually parents who report on ADHD symptoms, in cases where the intervention takes place at school, and teachers or clinicians in cases where the intervention takes place at home. If available, independent behavioral observations are taken as probably blinded ratings. Comparing the intervention effects reported by most proximal raters and probably blinded raters reveals huge differences. A comprehensive meta-analysis on non-pharmacological interventions in ADHD reveals significant

effects for measures using most proximal raters, but no significant effects for measures using probably blinded raters (Sonuga-Barke et al., 2013). In the same vein, raters most proximal to the therapeutic setting find a significant decrease in ADHD symptoms after cognitive trainings. However, the effects shrink substantially for measures using probably blinded raters (Cortese et al., 2015). In the field of neurofeedback as well, a recent meta-analysis revealed significant effects for raters most proximal to the therapeutic setting, but interventions effects that fall to an insignificant level for probably blinded raters (Cortese et al., 2016). Although the classification into probably blinded and most proximal raters is not straightforward and has been criticized (Daley et al., 2014), the differences in intervention effects highlight the limited objectivity of reported ADHD symptoms. The proximity to the therapeutic setting influences the evaluation of ADHD symptoms. Thus, reported ADHD symptoms can be considered to be highly influenced by the informant.

Second, findings on informant discrepancies in child and adolescent mental health are a well-documented phenomenon. In general, across societies, children and adolescents report more mental health problems themselves than their parents do about them (Rescorla et al., 2013). Cross-informant correspondence for externalizing disorders was estimated to be low to medium in a meta-analysis (r = .30, De Los Reyes et al., 2015). In the same vein, the accuracy of classifying children as with or without an ADHD diagnosis using teacher and parental reports is poor, and the association between teacher and parental reported ADHD symptoms is moderate (Mitsis, McKay, Schulz, Newcorn, & Halperin, 2000). Furthermore, the correspondence between informants within one setting (home or school) is much higher than the correspondence between informants across settings (Burns et al., 2014).The informant discrepancies highlighted here imply that reported ADHD symptoms are influenced by the informant to a substantial degree and therefore cannot be considered objective.

Third, the psychological processes that influence answers to questionnaire items in order are explained briefly, to evaluate the objectivity of reported ADHD symptoms (Schwarz, Knäuper, Oyserman, & Stich, 2009). To this end, it is helpful to become aware of the tasks a respondent to items has to complete in order to respond to an item. First, the respondent has to understand the item. Beyond literal understanding, the respondent has to understand what the receiver of the questionnaire, for instance, the health practitioner, wants to know and make assumptions about that intention. The formation of this assumption is important for the selection of meaningful answers. The

item “fails to give attention to details, makes careless mistakes” (Lidzba et al., 2013), for example, cannot be answered meaningfully without the assumption that the receiver of the questionnaire seeks to obtain information about inattentiveness beyond that to be expected by normal developmental processes, not the question of whether a first-grade student fails to be attentive to the spelling of every word he or she writes. Furthermore, the response alternatives of questionnaire items require that the respondent match the recalled frequency or intensity of a behavior of concern to the available response alternatives. The response alternatives provide anchor points for the respondent’s assumptions about the frequency and intensity of the concerned behavior the receiver is interested in. For example, a dichotomous response alternative (yes/no) for the item “fails to give attention to details, makes careless mistakes” (Lidzba et al., 2013) might lead to the assumption that the researcher is interested in really strong inattentive behavior, whereas response alternatives like “rarely”, “sometimes”, “often” might target rather mild inattentive behavior. Moreover, the context in which the questionnaire is given to the respondent provides important indications relevant to the formation of assumptions on what the receiver is interested in. The results of a questionnaire can be used to select children with difficulties for a free extracurricular training course, or in an online survey for epidemiological research. Answers to the same item will probably differ between the two contexts. Finally, representations of relatively frequent behavior in autobiographical memory are not detailed. In addition, memories are not categorized by type of behavior but rather according to time and place. The respondent has to rely on estimates in selecting autobiographical episodes that correspond to the item (Schwarz et al., 2009; Sudman, Bradburn, & Schwarz, 1996). All in all, this review of the tasks a respondent has to complete in order to answer a questionnaire item illustrates the influence an informant has on the way an item is answered. Therefore, ADHD symptoms as reported using questionnaire items are highly influenced by the respondent and are not objective.

In sum, evidence from intervention evaluations, informant discrepancies and psychological processes suggests that reports are a valuable but not objective source of information on ADHD symptoms. Reported ADHD symptoms are very behavior-based, meaning that they adhere closely to relevant behavioral descriptions. The following paragraphs look one step deeper, focusing on the diagnostic utility of cognitive conditions associated with ADHD symptoms.

Assessment of ADHD symptoms on the cognitive level

In addition to reports, information on ADHD symptoms can be obtained on a cognitive level. Cognition refers to information processing. Consequently, methods that gather information on ADHD symptoms on a cognitive level target information processing functions. Psychological theories indicate that alterations in certain information processing functions lead to the emergence of ADHD symptoms (see 1.1.2, Barkley, 1997b; Sonuga-Barke, 2002).

Do cognitive tasks provide *behavior-based* information on ADHD symptoms? The question regarding the proximity of information obtained on the cognitive level to behavioral descriptions of ADHD symptoms relates to the question of ecological validity. What does information on information processing functions tell us about actual behavior as described in the diagnostic criteria? To address this question, (1) the association with reported or observed ADHD symptoms or (2) the discriminative or predictive power can be investigated. An early study found low to medium ecological validity for laboratory measures of ADHD symptoms. One of the conclusions of that study is that measures should be improved and combined to overcome problems with ecological validity (Barkley, 1991). A more recent meta-analysis showed strong evidence for group differences between children with and without ADHD in relevant cognitive tasks (Willcutt et al., 2005). However, this meta-analysis revealed that an altered pattern of cognitive functions is present only in some but not all children with ADHD, and altered cognitive functions can be found in children without ADHD symptoms too. An altered pattern of cognitive functioning lacks universality among children with ADHD. Thus, problems with ecological validity in cognitive tasks (Barkley, 1991) cannot be diminished by improved tasks, as they result from the absence of deviant cognitive functions in some children with ADHD. Contrary to that finding, another study directly addressing the ecological validity of stop-signal and choice-delay tasks showed that a combined measure based on both tasks (Solanto et al., 2001) discriminated very well between children with and without ADHD. Moreover, results of the choice-delay task correlated with both reported and observed ADHD symptoms, while the stop-signal task correlated with observed ADHD symptoms. However, the discriminative power of cognitive tasks with regard to ADHD was not confirmed in a study combining multiple neuropsychological tests (Doyle, Biederman, Seidman, Weber, & Faraone, 2000). The combination of tests did not provide sufficient diagnostic utility either. However, associations between ADHD symptoms and reaction time variability, inhibitory control and delay aversion were confirmed in a community

sample (Wåhlstedt, 2009), underpinning the notion of ADHD as a dimensional construct. Cognitive tasks play a special role in preschool aged children, because behavioral rating scales are less reliable for them than for older children (Merkt, Siniatchkin, & Petermann, 2016). The most consistent associations have been found between ADHD symptoms and measures of delay aversion in preschool aged children (Mahone & Pritchard, 2013; Merkt et al., 2016). In the same vein, delay and inhibition among preschool aged children could predict ADHD symptoms in third grade (Campbell & Von Stauffenberg, 2009). Although cognitive tasks in preschool aged children predict ADHD symptoms, they do not possess diagnostic utility.

In sum, cognitive tasks as for instance the stop-signal paradigm (Alderson et al., 2007) or delay aversion tasks (Kuntsi, Stevenson, Oosterlaan, & Sonuga-Barke, 2001) provide information on specific cognitive functions but only very vague information related to behavioral descriptions of ADHD symptoms in the diagnostic manuals (APA, 2013; WHO, 1993). Therefore, methods that rely on cognition have to be evaluated as not very behavior-based and quite distal to behavioral descriptions of ADHD symptoms.

How *objective* is the obtained information on ADHD symptoms? To answer the question of objectivity regarding information obtained by cognitive tasks, it is important to consider potential influences of the informant (child) on the cognitive tasks. Informant influences on the results of cognitive tasks through motivation, task compliance and physiological conditions (such as tiredness) are conceivable. The analysis process for cognitive tasks usually contains procedures that exclude material suggesting inappropriate response behavior. Therefore, informants might be able to worsen their performance intentionally, but their influence on information about ADHD symptoms obtained via cognitive tasks is limited. Cognitive tasks can be considered objective measures. Turning away from cognitive functions, etiological approaches for ADHD assessment refer to neurobiological functioning. For this reason neurobiological measures are reviewed with regard to their proximity to behavioral descriptions and objectivity.

Assessment of ADHD symptoms on the neurobiological level

Information about ADHD symptoms on the neurobiological level here refers to differences between children with and without ADHD in the genome and brain structure and function. The attempt to describe disorders like ADHD on the biological level is mirrored in the search for biomarkers. A biomarker is an objectively measurable feature

that serves as an indicator of normal or pathological biological processes. A biomarker must identify ADHD cases accurately (sensitivity and specificity > .80) and must be reliable and valid. A comprehensive review on biomarkers for ADHD revealed that no biomarkers have been found to date (Thome et al., 2012). First attempts to use neurobiological measures in the ADHD diagnostic process (Müller, Candrian, & Kropotov, 2011) demonstrate the attractiveness of objective measures for ADHD. However, the context of the current evidence does not support the implementation of neurological measures in the ADHD diagnostic process. Difficulties in the search for biomarkers led to the development of the construct of endophenotypes. Endophenotypes try to directly link a biological mechanism with a behavioral function. In contrast to biomarkers, the focus here lies on the relation between the biological mechanism and behavior. Endophenotypes do not have to meet classification criteria (sensitivity and specificity), because the concept tries to account for the etiologic heterogeneity of ADHD and psychiatric disorders in general (Crosbie, Pérusse, Barr, & Schachar, 2008; Zobel & Maier, 2004). Thus, theoretically, several endophenotypes could cause the same clinical manifestation, for instance ADHD, as well as the clinical manifestations of other disorders.

Do neurobiological measures provide *behavior-based* information on ADHD symptoms? The fact that biomarkers are not sufficiently developed to predict diagnoses shows that information on neurobiological measures is quite distal to the clinical manifestation of ADHD. Neurobiological measures are important for capturing altered neurobiological and neuropsychological functioning but do not provide diagnostic utility. Considering the heterogeneity of the disorder, it is unlikely that one biomarker for ADHD can be isolated (Thome et al., 2012). The concept of endophenotypes seeks to advance genetic research by searching for links between genes and behavior. By definition, endophenotype research is non-clinical and does not try to find neurobiological correlates of clinical manifestations (Crosbie et al., 2008). Endophenotypes relate to concrete behavior, but do not focus on clinical manifestations such as ADHD. Thus, endophenotypes are also quite distal to behavioral descriptions of ADHD symptoms.

How *objective* is the obtained information on ADHD symptoms? One important reason for the efforts put into the search for biomarkers is that information obtained on the neurobiological level is objective (Müller et al., 2011; Thome et al., 2012). The informant, usually the concerned child, has no influence on his or her genome and only

limited influence on neurobiological measures. Thus, information obtained on the neurobiological level can be considered a source of objective information on ADHD symptoms.

1.2.2 Need for a behavior-based and objective assessment approach

The diagnostic process relies on the clinical manifestation of ADHD symptoms (see 1.1.3). This supports the notion of ADHD as a "useful clinical construct" (Sonuga-Barke, 2002, p. 29) rather than a disorder with a clearly delimitable psychological entity. Thus, behavioral descriptions of ADHD symptoms in the diagnostic manuals (see Table 1and 2, APA, 2015; WHO, 2011) are the anchor point for ADHD symptoms. The previous section reviewed sources of information on ADHD symptoms with regard to their (1) proximity to behavioral descriptions of ADHD symptoms and (2) the objectivity of the obtained information. Figure 1 shows the results of the review. Reported ADHD symptoms can be considered behavior-based and are in close proximity to behavioral descriptions of ADHD symptoms. However, reported ADHD symptoms are influenced by the informant to a great extent and cannot be considered objective sources of ADHD symptoms. Cognitive measures cannot provide information about the behavioral expression of ADHD symptoms. Therefore, cognitive measures must be considered distal to the behavioral descriptions of ADHD symptoms in the diagnostic manuals. However, informants have minimal influence on the result of cognitive measures, meaning that they can be considered objective. In the same vein, neurobiological measures are only loosely associated with behavioral descriptions of ADHD symptoms and are therefore considered not behavior-based and distal to behavioral descriptions. However, neurobiological measures cannot be influenced by informants and are very objective.

In sum, the review of assessment methods has shown that different assessment methods possess different qualities and provide information on ADHD symptoms from different perspectives. Concretely, the review reveals that reported symptoms and cognitive and neurobiological measures involve a trade-off between objectivity and proximity to behavior. This trade-off implies that there is a lack of information that is both objective and behavior-based. The following paragraph intends to show that observation has the potential to provide *behavior-based and objective* information on ADHD symptoms (Volpe, DiPerna, Hintze, & Shapiro, 2005).

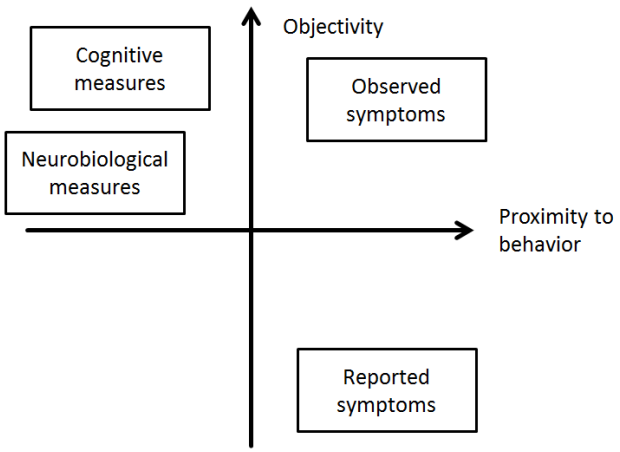

Figure 1 Categorization of information sources on ADHD symptoms on the axes proximity to behavior and objectivity (self-developed)

Behavioral observation

Behavioral observation can be defined as a measure that intends to capture a person's naturally occurring stream of behavior (Eid, Gollwitzer, & Schmitt, 2015). Naturally occurring behavior is the central interest of any psychological assessment, including assessments of ADHD symptoms. Moreover, behavior is the anchor point of ADHD symptoms and the only source of information that can be directly accessed. All information aside from observed behavior is more prone to interpretation bias (Furr & Funder, 2007). For instance, cognitive measures interpret finger-tapping on keyboards and reports interpret ticked boxes on questionnaires (Baumeister, Vohs, & Funder, 2007).

Why does observation have the potential to provide *behavior-based* information? Observation extracts information from behavior inductively. This means that the information of interest is extracted post-hoc, after the behavior has occurred. By contrast, both reports and cognitive and neurobiological measures often specify in advance which information should be provided by the informant via questionnaire items or computer tasks. This approach can be considered rather deductive. Although observation can also occur within a researcher-specified framework, it is much less intrusive and therefore closer to naturally occurring behavior.

Why does observation have the potential to provide *objective* information on ADHD symptoms? Ratings of observed behavior resemble reported ADHD symptoms and are not objective per se. Observation as an assessment method has developed three tools which allow objective, meaning unbiased by the informant, information to be extracted. First, the people who judge the observed behavior (hereafter: raters) have to

be *blinded* to the construct of interest. For instance, if ADHD symptoms are being observed, raters should not know whether or not a given child previously received an ADHD diagnosis. Second, a *training for raters* should ensure that all raters have the same conception of the construct of interest and teach them rating rules. The aim of this training is for all raters to evaluate the construct of interest in the same way. Third, *multiple raters* must observe and evaluate the behavior of interest *independently* of one another. *Agreement between these multiple raters* is an indicator for the objectivity of the information obtained by observation: The more similarly the trained and blinded raters evaluate the behavior of interest independently of one another, the less influenced the evaluation is by the rater him/herself (Wirtz & Caspar, 2002). In sum, behavioral observation is an assessment method that is most proximal to the behavior of interest and has the potential to be objective.

However, these advantages of behavioral observation as an assessment method are accompanied by obstacles. Behavioral observation is a very rich assessment format, which implies both great flexibility and great challenges. Behavioral observation is costly and takes more effort to implement than reports or questionnaires. In most cases, a psychometrically sound rating system has to be developped. Independent raters have to be selected and these raters have to be trained until inter-rater agreement reaches a sufficient level of inter-rater reliability (Furr & Funder, 2007). Thus, before gaining insight into the behavior of interest, many methodological questions have to be answered. Some typical questions are summarized here for purposes of illustration. First, (1) the level of abstraction of the behavioral observation has to be set. Evaluations can be made on the construct level or on the level of behavioral indicators for the construct. An example of this is the decision whether raters should evaluate a child's level of hyperactivity or the frequency with which he/she leaves his/her seat. This decision depends primarily on the amount of effort one is willing to put into rater training. Second, (2) the pattern of observation is an important feature. Raters can evaluate either the frequency or duration of a behavior of interest, for example. Both practical reasons, for instance the availability of observation software, and theoretical reasons, for instance the occurrence rate of a behavior of interest, should guide this decision. Third, (3) one has to decide on an appropriate response format for the observation. Dichotomous response formats (occurred/did not occur) or intensity scales are possible. Forth, the level of (4) standardization of the observation setting has to be set. Standardized settings allow the experimenter or diagnostician to have more

influence, thus making it easier to link inferences to the setting, whereas observation in more natural settings possesses greater ecological validity. Fifth, (5) the degree to which the observer is involved is important. The observer can take part in the observed situation (participatory observation), observe the situation without participating, or evaluate video recordings of the situation without being physically present and thus probably remain more blinded to contextual factors.

In sum, observation is a flexible assessment method and possesses many modifiable parameters. However, one differentiation between observation and other assessment methods is important in most cases: Observation provides information on states, whereas reported information on ADHD symptoms are traits in most cases. This differentiation is explained in the following paragraph in more detail.

Distinction between state and trait

In addition to the implementation challenges associated with behavioral observation, the type of information obtained via this method differs from reported information. Early research work on the identification of ADHD symptoms advised against the use of behavioral observation because information from behavioral observation is usually based on only one occasion. Reported information collapses observations over a longer time frame and in a number of different situations (Barkley, 1997a). This is why reported information asks about traits, meaning general behavioral dispositions. Behavioral observation assesses states, meaning locally and temporally specific behavior. Traits are defined as personal dispositions which are relatively independent from specific situations. States, however, describe the present activity and are specific for a certain time and situation (Allport & Odbert, 1936). This has important implications. State information is always embedded in a certain context, whereas trait information seeks to capture information that is independent of any specific context.

In sum, behavioral observation, if not conducted on multiple occasions over a longer time frame, captures states, locally and temporally specific behavior, rather than general behavioral dispositions.

Review of observation studies of ADHD symptoms

The existing literature describes different implementations of the behavioral observation of ADHD symptoms. The following paragraphs describe some example applications in the classroom environment as well as some existing standardized measures.

Ecologically valid observational studies observe ADHD symptoms in classroom contexts. An overview of the behavioral codes used in the reviewed studies entailing observation of ADHD symptoms is provided in Table 4. First, studies that implemented *observation in simulated classroom situations* are reviewed. A study investigating peer interactions revealed that children with ADHD did indeed exhibit more off-task behavior compared to their healthy peers (Cunningham & Siegel, 1987). A second study observed the level of hyperactivity in hyperactive and healthy children in formal and informal simulated classroom settings. Hyperactive children showed more hyperactive behavior in formal and informal simulated classroom settings (Jacob, O'Leary, & Rosenblad, 1978). A third study in a simulated classroom developed an observation system to differentiate aggressive from hyperactive boys in a free play, restricted play, and restricted academic setting. More than two-thirds of the participants could be successfully classified as hyperactive, aggressive, hyperactive and aggressive or healthy. Time on task was particularly successful at differentiating between aggressive and hyperactive boys (Roberts, 1990). The reviewed literature on behavioral observations of ADHD symptoms in simulated classroom situations reveals great variability in rating categories. Time on task is a widely used measure in observational studies in classrooms.

This holds true equally for the *observation of ADHD symptoms in natural classroom environments*. A sophisticated observation study looked for differences in ADHD symptoms among children with and without ADHD during idle time and non-idle time in classrooms. As expected, activity and noisiness were higher among children with ADHD, especially during non-idle time (Imeraj et al., 2016). The goal of a second observation study to observe children with and without ADHD in three different natural classroom environments (regular lesson, regular lesson with interaction, non-instructional context). The behavioral categories deployed were moderately to strongly associated with teacher reports on ADHD symptoms (Lauth, Heubeck, & Mackowiak, 2006). A third observational study in natural classroom environments focused on the variability of children's attention during academic assignments. In line with the authors' hypotheses, children with ADHD switched between visually attentive and inattentive states more frequently than healthy controls (Rapport et al., 2009). A fourth study developed a direct observation protocol for school-aged children with ADHD in natural classrooms. This protocol can be reliably applied after approximately 30 training observations (Steiner et al., 2013). A fifth study developed observation categories for

natural classroom behavior aimed at differentiating hyperactive from healthy children. The observation categories interference and off-task classified children correctly as hyperactive or healthy with a sensitivity of about 80% (Abikoff, Gittelman, & Klein, 1980). In line with the mentioned observational studies in natural classrooms, a meta-analysis on the observation of inattentiveness in classrooms used on- and off-task behavior to operationalize inattentiveness in classrooms (Kofler et al., 2008).

In addition to studies on behavioral observation of ADHS symptoms, a few *standardized published rating systems for ADHD symptoms for use in practice* exist which are relevant for ADHD behavior. The Classroom Assessment Scoring System (CLASS, Pianta, La Paro, & Hamre, 2006) assesses emotional support, classroom organization and general instructional support as categories of classroom management. The InCLASS is the equivalent for observing an individual child (Downer, Booren, Lima, Luckner, & Pianta, 2010). It assesses teacher interaction, peer interaction and task orientation. Thus, it captures behavior relevant for ADHD. In the same vein, the German questionnaire on judging in-class behavior (German: Fragebogen zur Verhaltensbeurteilung im Unterricht, FVU, Breuer, Rettig, & Döpfner, 2009) can be used to assess working behavior at school independently of a particular disorder. It is not intended to assess ADHD symptoms in particular, yet its items are very close to descriptions of ADHD symptoms. It provides information on the scales “attention problems” and “lacking compliance”. Validation studies exist, but a standardization study has not been conducted yet (Breuer et al., 2009). A sophisticated observation protocol exists for diagnosing disruptive behavior (Wakschlag, Briggs-Gowan, et al., 2008; Wakschlag, Hill, et al., 2008), which has some utility for diagnosing ADHD in preschoolers (Bunte, Schoemaker, Hessen, Van Der Heijden, & Matthys, 2013). Standardized observation systems for practice do not exist specifically for ADHD symptoms.

All in all, the categories in rating systems for ADHD symptoms differ a lot. In most cases, the categories are aligned with the specific research aims rather than the behavioral descriptions of ADHD symptoms. Further effort has to be put into developing rating items for ADHD symptoms in practically and theoretically meaningful situations that are close to the behavioral descriptions.

Table 4
Overview of Behavioral Coding in Observational Studies in Classrooms entailing ADHD Symptoms

Study	Situation	Behavioral codes
Cunningham & Siegel, 1987	Simulated classroom	– For peer interactions: positive interaction, controls, solitary activity, positive response, complies, controlling responses, ignores, observes – On-task behavior: looking at or working on the drawing or math task
Jacob et al., 1978	Simulated classroom	– Solicitation: attempts to initiate interactions with the teacher, – Aggression: physical or verbal attacks, refusal or resistance to obeying the teacher's commands – Change of position: was coded as change in location of at least two steps – Daydreaming: noninvolvement with the task – Weird sounds: non-verbal vocal sounds aside from language communication
Roberts, 1990	Simulated classroom	– Proportion of time on task, fidgeting, out of seat, vocalizing, number of task shifts – Number of times crosses squares drawn on the floor (for assessing activity).
Imeraj et al., 2016	Natural classroom	– Activity: problematic or unproblematic – Nonsocial vocalization: noisy or not noisy – Social behavior: not disruptive or disruptive
Lauth et al., 2006	Natural classroom	– Off-task: child is actively disruptive or passive and inattentive – On-task behavior was coded when the child showed expected behavior, inconspicuous behavior, self-initiated activity (e.g. raising hand) and other-initiated activity (answering a question).
Rapport et al., 2009	Natural classroom	– On-task behavior, defined as visual fixation on task-relevant stimuli
Steiner et al., 2013	Natural classroom	– On-task: active engaged time (writing, reading, raising hand) or passive engaged time (listening) – Off-task: motor (movements not related to assigned task), verbal (vocalization not related to assigned task), passive (not attending to assigned task)
Abikoff et al., 1980	Natural classroom	– Interference, solicitation, off task, minor motor movement, gross motor-all, gross motor-standing, gross motor-vigorous, noncompliance, out of chair, aggression, verbal aggression to children, verbal aggression to teacher

1.2.3 Context-dependent information on ADHD symptoms

According to official guidelines, the main sources of information on ADHD symptoms are reports from multiple informants and clinical exploration (Bundesärztekammer, 2005). Rating scales are used to produce reported ADHD symptoms. In most cases, items ask for a child's general disposition to be inattentive, hyperactive or impulsive. As described in the previous section, reported ADHD symptoms in most cases provide information on traits, or general behavioral dispositions. Informants filling out rating scales are asked to estimate the frequency of a certain behavior, e.g. within the last six months. Conners rating scales use the response format "not at all", "a little", "very much", and "exactly" for items like "Is inattentive and easily distractible" (Lidzba et al., 2013). By contrast, temporally and locally specific behavior is referred to as a state (Allport & Odbert, 1936). The fact that the main sources of information on ADHD symptoms refer to traits has important implications. Because traits collapse information on ADHD symptoms over a longer time frame and in various situations, trait ADHD symptoms contain reduced information on situational variability (Steyer, Ferring, & Schmitt, 1992). Situational variability of behavior means that the occurrence of a behavior is context-dependent. There is an longstanding controversy about the validity of traits in light of the situational specificity of behavior (Anastasi, 1983). This controversy is especially important for the assessment of ADHD symptoms. Current *psychiatric nosology*, *cross-situational and multi-informant discrepancies*, and *ADHD theory* all highlight the importance of the context-dependent, meaning situation-specific, measurement of ADHD symptoms.

Psychiatric nosology of ADHD

Psychiatric nosology relies on clinical manifestations. Behavioral descriptions of ADHD symptoms in most cases refer to contexts (see Table1 and 2, APA, 2015; WHO, 2011). This is particularly apparent in an example item from the DSM-5: "Often leaves seat in situations when remaining seated is expected (for instance, leaves his or her place in the classroom, in the office or other workplace, or in other situations that require remaining in place)"(APA, 2015, p. 78). The behavioral description itself ("leaving seat") is accompanied with concrete conditions referring to the context in which this behavior takes place. In addition to a rather vague condition ("when

remaining seated is expected"), concrete examples of relevant contexts are mentioned ("in the classroom, in the office or other workplace"). Many of the DSM-5 criteria are worded in accordance with this pattern (APA, 2015). ICD-10 research criteria distinguish whether each item applies to the home, school or nursery setting (WHO, 1993). Interestingly, widely used rating scales often forgo context descriptions and examples for the sake of simplicity (Conners, 2011; Lidzba et al., 2013). Juxtaposing DSM-5 criteria and rating scales (for an example, see Table 3) reveals that contextual information is often dropped in rating scales. Thus, information on the context-dependent occurrence of ADHD symptoms as required by the diagnostic manuals (APA, 2015; WHO, 1993) can rarely be captured by rating scales. Moreover, answers to rating scale items are usually aggregated into a mean value and are not subject to context-dependent analysis. This aggregation routine erases possible information on context dependence. Therefore, unstructured clinical exploration is the only source of information on context-dependent assessment. Furthermore, in contrast to the emphasis on the context of behavior in the diagnostic manuals, both DSM-5 and ICD-10 demand that symptoms occur pervasively across settings (APA, 2015; WHO, 1993).

In sum, the current nosology of ADHD symptoms specifies concrete contexts for behavioral descriptions. Thus, ADHD symptoms can be considered to be context-dependent. However, current questionnaire items do not capture information of the context of ADHD symptoms.

Cross-situational and multi-informant discrepancies

After having looked at the context-dependent definition of ADHD symptoms in the diagnostic manuals, this section takes a closer look at related empirical findings. In the previous section (see 1.2.1), informant discrepancies in reported ADHD symptoms were discussed in the context of the objectivity of reports. This section reviews discrepancies with an eye to the information on context inherent in them. Substantial discrepancies between informants in rating scale evaluations are a stable finding in child and adolescent mental health and ADHD research (Gomez, Burns, Walsh, & de Moura, 2003; Mitsis et al., 2000; Rescorla et al., 2013). In the same vein, the association between information on ADHD symptoms drawn from two different situations (for instance home and school) is considerably weaker than information on ADHD symptoms drawn from the same situation (Burns et al., 2014). The empirical evidence for cross-situational and multi-informant discrepancies in ADHD symptoms raises the question of how to evaluate them. One explanatory approach is to declare the

discrepancies to be measurement error. Another explanatory approach is to view ADHD symptoms as context-dependent. Different informants (usually teachers and parents) experience the child in different situations and use this as their basis of evaluation when answering rating items. Prominent researchers argue that taking evaluation discrepancies into consideration improves the validity of the assessment (De Los Reyes, 2011; De Los Reyes et al., 2014; Dirks, De Los Reyes, Briggs-Gowan, Cella, & Wakschlag, 2012). An important reason for this assumption is that evaluation discrepancies contain information on the context in which symptoms occur. This notion is underpinned by the fact that evaluation discrepancies can be reliably measured and are stable over time (Burns et al., 2014; De Los Reyes, 2011). Statistically, the integration of information from multiple informants requires latent modeling. In practice, the integration of judgment discrepancies is confronted with huge obstacles (Achenbach, 2011).

In sum, empirical findings on cross-situational and multi-informant discrepancies in rating scales imply that the occurrence of ADHD symptoms is context dependent. However, widely used rating scales and analysis techniques cannot capture this context dependence.

Development of theoretical approaches of ADHD

Another reason for the context dependence of ADHD symptoms is provided by psychological theories on ADHD. In the previous section, psychological theories of ADHD were reviewed regarding their etiological approach (see 1.1.2). Here, ADHD theory is reviewed with regard to the context of ADHD symptoms. Executive dysfunction theory assumes that a general deficit in behavioral inhibition is the reason for the occurrence of ADHD symptoms (Barkley, 1997b). Due to the fact that a dysfunction in behavioral inhibition can be measured for some but not all children with ADHD (Willcutt et al., 2005), pathway theories were developed to account for the psychological heterogeneity of children with ADHD and have introduced delay aversion into the theoretical framework as a context-dependent feature (Sonuga-Barke, 2002). State regulation theory, which is most current, states that an impaired adaptation to current needs is a core deficit in ADHD (Van der Meere, 2005). The emphasis on adaptation processes and the reference to current needs implies a strong context dependence of ADHD symptoms.

In sum, the developmental progression of psychological theories shows that theoretical approaches accounting for the context-dependent occurrence of ADHD symptoms have become more important over time.

All in in all, psychiatric nosology, cross-situational and multi-informant discrepancies and ADHD theories imply that the occurrence of ADHD symptoms is context-dependent. Rating scales, the most frequently implemented assessment method which in research and practice, do not capture information on the context in which ADHD symptoms occur. The following section introduces observation as an assessment instrument that is capable of capturing the context of ADHD symptoms.

1.2.4 Need for a context-dependent assessment approach

The previous section pointed out that the occurrence of ADHD symptoms is context-dependent. However, symptom reports using rating scales, the most widely used assessment method for ADHD symptoms, usually do not capture information on the context of ADHD symptoms. Observation always takes place in a concrete situation. Information on the observation situation enables behavioral observation to deliver context-dependent information. Information from behavioral observation is temporally and locally specific and provides information on states unless observation takes place on multiple occasions (Allport & Odbert, 1936). A review on the importance of context-dependent measures in child and adolescent mental health confirms that "observational measures provide a significant amount of contextual information, both at the setting and situation level" (Dirks et al., 2012, p. 560). An important drawback of the state ADHD symptoms provided by behavioral observation is the fact that observed ADHD symptoms do not represent a general disposition. This means that generalizability is questionable (Barkley, 1997a). An important approach to overcoming this drawback is to observe ADHD symptoms in meaningful situations. Observed behavior in meaningful situations is likely to generalize to situations similar to the observation situation. Which criterion should be used to select meaningful situations? Situations which are affected by impairment through symptoms should be targeted. Diagnostic manuals require ADHD symptoms to cause impairment in social, academic or occupational life (APA, 2015; WHO, 2011). This means that information on the context of ADHD symptoms is needed; otherwise, impairment in important contexts for ADHD symptoms like social and academic functioning cannot be assessed. Important contexts

which are associated with considerable impairment in children with ADHD symptoms are the family context (Deault, 2010; Johnston & Mash, 2001), delay contexts (Patros et al., 2016; Pauli-Pott & Becker, 2011) and the school context (Daley & Birchwood, 2010; Frazier, Youngstrom, Glutting, & Watkins, 2007).

In sum, ADHD symptoms are highly context-dependent. Evidence of this has been provided from the perspectives of psychiatric nosology, cross-situational and multi-informant discrepancies and ADHD theory. Reported symptoms do not capture information on context. Moreover, current assessment methods do not provide both objective and behavior-based information on ADHD symptoms. Therefore, the present dissertation investigates the potential of observation to provide behavior-based, objective and context-dependent information on ADHD symptoms. Three empirical studies have been conducted to that end; the following section describes the specific research aims.

.

2 RESEARCH AIMS

This dissertation follows the overarching goal of investigating the potential of observation of ADHD symptoms as a behavior-based, objective and context-dependent assessment instrument. The review of assessment methods for ADHD symptoms via reports and on the cognitive and neurobiological level revealed that none of these assessment methods provides information on ADHD symptoms which are behavior-based, objective and context-dependent (see 1.1). The basis of observation is naturally occurring behavior; therefore it is most proximal to behavior and very behavior-based. Moreover, observation as an assessment method provides techniques to guarantee objectivity, meaning the independence of the information on ADHD symptoms from the observer (see 1.2.2). Finally, naturally occurring behavior, the basis of observation, always takes place in a specific context. The potentials of observation as a behavior-based, objective and context-dependent assessment method are measured by three parameters in the empirical studies of the present dissertation. First, the objectivity of observation is investigated by a comprehensive assessment of inter-rater reliability. Second, the validity of observations is investigated by means of association analyses between observed and reported symptoms. Third, the contexts of observations in the empirical studies have been selected according to their relevance for functional impairment through ADHD symptoms and to theoretical considerations. The family context and the school context are pivotal living contexts of school-aged children and often affected by impairment in children with ADHD (Frazier et al., 2007; Johnston & Mash, 2001). Thus, observation techniques are applied to the family context (*Study 1*) and to ADHD symptoms in a classroom context (*Study 3*). From a theoretical perspective, the delay aversive motivational style of children with ADHD symptoms has been studied before with the delay of gratification paradigm (Mischel, 1996; Patros et al., 2016). Therefore, ADHD symptoms in a delay of gratification task are observed in *Study 2*. This selection of context allows to measure meaningful outcome variables (*Study 2*: delay of gratification performance and *Study 3*: performance in a math test) and to relate them to observed ADHD symptoms. In sum, the potentials of observation are investigated with the following parameters (1) inter-rater reliability, (2) association analyses between reported and observed ADHD symptoms and (3) association analyses between observed ADHD symptoms and meaningful outcome variables (*Study 2*: delay

of gratification performance and *Study 3*: performance in a math test). The research aims of the three studies are described in more detail in the following:

Study 1 uses observation techniques to assess the family context. Concretely, it investigates the utility of a Five-Minute Speech Sample (FMSS, Magana, Jenkins, & Miklowitz, 1986) to assess Expressed Emotions (EE) as a measure for family context (Graf Schimmelmann et al., 2003) in German school-aged children. The research questions are:

1. Does the FMSS assess EE reliably?
2. Are EE scales associated with reported ADHD symptoms?

Study 2 involves behavioral observation of German school-aged children with and without ADHD during a delay of gratification task (Mischel, 1996). The aim of the study is to investigate the influence of attention orientation, observed activity and impulsivity and an ADHD diagnosis on performance in a delay of gratification task. The research questions are:

1. Is observed attention orientation during the task a predictor of performance?
2. Do children with an ADHD diagnosis perform worse compared to children without an ADHD diagnosis?
3. Do observed ADHD symptoms during the task influence performance?

Study 3 investigates the utility and feasibility of a behavioral observation protocol for ADHD symptoms in a simulated classroom situation as one important context affected by impairment (Frazier et al., 2007; Kofler et al., 2008). To achieve that goal, German school aged children were observed during a math test and a competitive card game via video recordings of every single child. The research questions are:

1. Can inattention, hyperactivity and impulsivity be reliably assessed?
2. How are observed and reported ADHD symptoms associated with one another?
3. Do children who exhibit more ADHD symptoms during a math task perform worse in that task?

3 EMPIRICAL STUDIES

3.1 Study 1

3.1.1 Introduction

The etiology of ADHD is multifactorial and based on an interaction of psychosocial and biological factors (Banaschewski et al., 2004; Faraone & Biederman, 1998). While evidence for the heritability of ADHD is especially strong (h = .80, Faraone et al., 2005), investigations into the psychosocial environmental factors associated with ADHD are less developed (Johnston & Mash, 2001). Beyond etiological questions, the investigation of psychosocial environmental factors associated with ADHD is of major importance when it comes to gaining insights into functional impairments associated with ADHD symptoms (APA, 2015; Gordon, 2006). One of the most prominent psychosocial environments children experience regularly and which could be affected by impairment due to children's ADHD symptoms is daily life in the family.

ADHD symptoms in the family context

Review articles that have explored the family context of children with ADHD are mainly based on correlational studies and report a great heterogeneity of findings (Johnston & Mash, 2001; Deault, 2010). Conceptualizations of family context are broad and range from marital relationships to observations of parent-child interaction to parents' self-report measures on family functioning. Importantly, the association between comorbid symptoms of conduct and oppositional defiant disorder and unfavorable family contexts seems to be stronger than the association between ADHD symptoms and unfavorable family contexts (Deault, 2010). Therefore, whether the association between family context and ADHD symptoms can be fully explained by comorbid symptoms or is in fact specific to ADHD symptoms is a subject of discussion. If the broad concept of family context is narrowed down to parenting stress, the picture becomes more explicit. A meta-analysis investigated the association between parenting stress and ADHD symptoms and compared the strength of association between clinical and healthy controls and parenting stress. Parenting stress was defined as stress that arises from perceived incapability to deal with parenting demands. The results showed

that parents of children with ADHD experience more stress than parents of healthy controls, but no more stress then parents of clinical controls (Theule, Wiener, Tannock, & Jenkins, 2013). Child psychopathology seems to be associated with parenting stress independently of the child's specific disorder. In sum, the assessment of family context in relation to child psychopathology faces the challenge of finding a conceptually limited construct that captures the family context relevant to child psychopathology.

Expressed Emotion

EE is a construct which captures the emotional climate of the home environment, which is relevant for psychiatry patients (Peris & Baker, 2000). The EE construct was developed to capture the interaction between mental disorders and family variables (Rutter & Brown, 1966). EE provides a suitable operationalization of family context for investigating of the association between child psychopathology and the family context. It developed out of schizophrenia relapse research and the observation that relapse is associated with characteristics of the family context (Brown, Monck, Carstairs, & Wing, 1962). The original assessment took place in a three-hour semi-structured interview (Rutter & Brown, 1966). The most prominent interview for this purpose is the Camberwell Family Interview, which is audio-recorded and rated on the scales critical comments, hostility, positive remarks, emotional overinvolvement and warmth (Van Humbeeck, Van Audenhove, De Hert, Pieters, & Storms, 2002). A more ecological method is the FMSS (Magana, Jenkins, & Miklowitz, 1986). It consists of a five-minute audio recording of one relative who speaks about the patient and his/her relationship to him/her. The great advantage of this measure is the low vulnerability to socially desirable answers and the enhanced objectivity compared to self-report questionnaires. Instead of conscious beliefs about parenting and parenting stress, which are captured by questionnaires, EE instead captures emotions and attitudes (Peris & Miklowitz, 2015). The EE was originally coded in terms of the components emotional overinvolvement and criticism (Magana et al., 1986). The measure has been adapted for use with child psychiatry patients and their parents as respondents (Daley, Sonuga-Barke, & Thompson, 2003; Schuh, 2015). Empirical evidence has revealed inconsistent associations between the emotional overinvolvement subscale and patient functioning during childhood (Daley et al., 2003; Graf Schimmelmann et al., 2003; Peris & Miklowitz, 2015). Thus, for the assessment of EE in parents responding for their child, EE consists of the original component criticism as well as scales for the parent's initial statement, quality of relationship and frequency counts for critical and positive

comments. A fifth coding criterion, warmth, is also assessed in childhood samples. A worse family environment is assumed in cases where parents express a negative initial statement, report a lower quality of relationship, less warmth, and make more critical comments and fewer positive ones. Importantly, the applied subset of scales shows substantial variability in different studies. Although some studies have tried to find disorder-specific components of EE, evidence suggests that its association to psychopathology is not specific to certain disorders (Asarnow, Tompson, Woo, & Cantwell, 2001; Butzlaff & Hooley, 1998; Graf Schimmelmann et al., 2003; Stubbe, Zahner, Goldstein, & Leckman, 1993). Current evidence suggests that EE in general is unstable over time. Correlation coefficients of EE ranged between r = .14 - .15 for a time interval of one year and between r =.-.14 and .08 for a time interval of six years (Peris & Baker, 2000; Richards et al., 2014), which implies that EE is changeable. EE status seems to be more driven by child factors than parental factors like parental psychopathology (Cartwright et al., 2011; Psychogiou, Daley, Thompson, & Sonuga-Barke, 2007). The utility of EE as a measure for family context among children exhibiting ADHD symptoms in particular is examined in the next paragraph.

Expressed Emotion and ADHD symptoms

The association between ADHD symptoms and EE has been shown in preschoolers (Daley et al., 2003; Schloß et al., 2015) and schoolchildren (Christiansen, Oades, Psychogiou, Hauffa, & Sonuga-Barke, 2010; Peris & Hinshaw, 2003; Richards et al., 2014). In preschool-aged children, EE discriminates between ADHD and non-ADHD children (Daley et al., 2003) and correlates with dimensional parental reported ADHD symptoms (Schloß et al., 2015). In school-aged children, EE scales are significantly higher among children diagnosed with ADHD than healthy children (Christiansen et al., 2010), and the EE scale warmth was found to be significantly associated with dimensional parental reported ADHD symptoms, although criticism was not (Richards et al., 2014). However, another published study did not find differences between children with ADHD and healthy controls (Asarnow et al., 2001), and two other studies showed that the association between ADHD symptoms and EE is better explained by comorbid conduct problems (Cartwright et al., 2011; Psychogiou et al., 2007). A sophisticated longitudinal study showed that persistent high EE-criticism can be found particularly in children with ADHD that do not exhibit an age-appropriate decline in hyperactivity-impulsivity (Biederman, 2000; Musser, Karalunas, Dieckmann, Peris, & Nigg, 2016). Parental EE was obtained for 208 children at two time points one

year apart. Thus, expressed emotion could play an important role in the developmental outcomes of children with ADHD (Musser et al., 2016).

Generalizability aspects of the association between EE and ADHD symptoms

First, a German adaptation of the original English coding manual for EE in children has only been applied to a preschool sample (Schloß et al., 2015) and in the context of a training program for schoolchildren with conduct problems (Schuh, 2015). The applicability of the FMSS to German speakers has been shown in studies with relatives of adult schizophrenic patients (Leeb et al., 1991). As language aspects play an important role in the coding process, the applicability of the German version of the FMSS manual for schoolchildren should be investigated and developed further. Second, most of the evidence for the association between ADHD symptoms and EE in schoolchildren relies on clinical samples (Asarnow et al., 2001; Cartwright et al., 2011; Christiansen et al., 2010). The generalizability of the association to subclinical samples is important to account for ADHD as a dimensional construct and prove the utility of FMSS in subclinical community samples (Stubbe et al., 1993).

3.1.2 Present Study

The aim of the present study is twofold. First, the reliability of EE as assessed by the German FMSS coding manual is tested. Second, the association between ADHD symptoms and EE in German school-aged children is investigated. In particular, we hypothesize that more ADHD symptoms in parental and self-reports are associated with (1) a more negative initial statement, (2) less warmth, (3) a more negative relationship, and (4) more critical and fewer positive comments.

3.1.3 Method

Procedure

The study was part of a larger project on attention at school. Recruitment took place via local schools, local child and adolescent psychotherapy practitioners and e-mails to university affiliates in a German university town. Information on the study was sent to interested parents. After parents and children gave informed and written consent to take part in the study, appointments for one telephone interview with the child (10 min) and one with the parent (30 min) were made. The author of this study conducted

all telephone interviews. All data used in the present study were obtained in these telephone interviews. Participation in the entire project also included a videotaped group session with the children at the university. The local ethics committee approved the study and the local educational authorities permitted recruitment at schools.

Participants

Thirty-eight children and adolescents took part in the study, as did one parent of each student. Due to technical issues, five recordings of FMSS were unusable; therefore, 33 children (19 female, M_{age} = 10.65 years, SD_{age} = 1.34, n = 6 diagnosed with ADHD according to parental report) as well as one parent each (82% mothers) were included in the analyses. Eight participants had one sibling that took part in the study, so four parent respondents answered for two children. About one half of the children attended primary school (55%) and the other half (45%) secondary school. Six parents (18 %) reported that their child is diagnosed with ADHD, n = 26 parents (79 %) reported that their child had never received an ADHD diagnosis. One parent reported that their child was currently in the diagnostic process for ADHD. Four of the six children with an ADHD diagnosis were taking medication to treat their ADHD during the time of the study (i.e. 1 child - amphetamine; 3 children - methylphenidate). Four parents reported in an open-ended question format that their child had been diagnosed with a mental disorder other than ADHD (1 child - emotional disorder; 1 child – unstated, 1 child highly gifted, 1 child in the diagnostic process for dyslexia).

ADHD symptoms

ADHD symptoms were assessed using both parental reports and self-reports. The German translation of the DSM-5 criteria for Attention-Deficit/Hyperactivity Disorder was used (ADHD DSM) to assess inattention and hyperactivity-impulsivity (APA, 2015). In the instructions, participants were asked to judge the frequency with which the statement read aloud by the interviewer occurred in the last six months. The response format was a 4-point Likert scale ranging from 0 (*not at all*) to 3 (*very much*). Each criterion was adapted for parental and self-reporting. For example, the criterion "often fails to give close attention to details" (APA, 2015, p. 77) was changed to "your child often fails to give close attention to details [...]" (parental report) and "I often fail to give close attention to details [...]" (self-report). All items were analogously adapted to fit this pattern. Nine items refer to inattention, five to hyperactivity and four to impulsivity (18 items in total). For the present study, the internal consistency of these 18 ADHD DSM items was sufficient for both the parental report (Cronbach's α = .94)

and the self-report (Cronbach's $\alpha = .88$). After the assessment of the ADHD DSM content scales (inattention and hyperactivity-impulsivity), parents were asked to decide whether or not the symptoms caused impairment in social life or academic functioning (yes/no)[1], in line with the D-criterion of the DSM-5 (APA, 2015). Because the DSM-5 criteria are not commonly used to assess the severity of symptoms, the ADHD Index of the German version of the Conners-3 rating scale was also used in self- and parental report versions (Lidzba et al., 2013). The ADHD Index contains 10 items that best differentiate between children with and without an ADHD diagnosis. The items on the ADHD Index refer to the child's behavior in the last month. Responses were given on a 4-point Likert scale ranging from 0 (*not at all*) to 3 (*very much*). The validity and reliability of the ADHD Index have been proven (Lidzba et al., 2013). The intercorrelation for ADHD DSM mean scores and ADHD Index mean scores was very high for self-report; $r = .88$, $p < .001$, $n = 33$; as well as for parental report; $r = .92$, $p < .001$, $n = 33$ (both Pearson's correlation coefficients). Mean scores for all ADHD DSM and ADHD Index items were calculated and used for further analyses.

Expressed Emotion

The FMSS (Magana et al., 1986) was applied in the present study to assess EE. At the beginning of the telephone interview, the interviewer asked the parent to "tell what kind of a person [name of the child] is and how the two of you get along together". The parent was told that the interviewer would not interrupt for five minutes. The answer was saved as an audio file[2]. Two rating manuals served as a basis for coding the FMSS in the present study. The first one was the English manual for coding EE from the Preschool FMSS (Daley et al., 2003). The second one is the German translation of this manual, which has been adapted and tested for the use with older children (Schuh, 2015). Table 5 gives an overview of the rating scales used in the present study. The rating scales in the present study remained the same as in the original adaptation for preschool-aged children, with two major exceptions (Daley et al., 2003). First, the inter-rater reliability of the warmth scale was especially low in previous studies: (Schuh,

[1] Wording of the item (translated from German): (1) The behaviors that you just rated for your child are present to an extent which is not appropriate for the developmental level of your child and negatively affects social and academic activities.

[2] An adapter connected the receiver of a landline phone, a laptop computer, and the landline phone. The voice of the parent was recorded via a common recording program.

2015: Cohen's κ = .54; Schloß et al., 2015: Cohen's κ = .57). The description of the original main rating criterion of warmth –"spontaneity" is very inexplicit. Therefore, it was changed to a more explicit criterion "affection and appreciation". Second, tone of voice as an indicator for warmth is not explicitly described in the manuals that served as the basis for the present study. Moreover, in a pilot study (n = 7), the author of the present study did not succeed in finding a more explicit description of tone of voice. Thus, tone of voice was dropped from the rating criteria of these scales in the present study. Tone of voice originally had the function of preventing ironic comments from being counted as critical or positive comments respectively. A caveat on ironic statements was therefore added to the description of positive and negative comments.

Table 5
Description of EE Scales

	Main rating criteria	Coding
Initial Statement	– Description of the child – Description of the relationship to the child	Global rating: positive – 1 neutral – 2 negative – 3
Warmth	– Affection and appreciation – Concern and empathy	Global rating: high – 1 moderate – 2 low – 3
Relationship	– Description of joint activities and evaluation of these activities – Direct statement about the relationship towards the child	Global rating: positive – 1 neutral – 2 negative – 3
Critical Comments	– Statements on misconduct, negative characteristics and destructiveness of the child (summarized by topic)	Frequency count
Positive Comments	– Praise, appreciation and description of positive characteristics of the child (summarized by topic)	Frequency count

Note. EE – Expressed Emotion

Rating procedure

Two undergraduate psychology students blind to the severity of ADHD symptoms rated all of the speech samples (fully-crossed design). Sample speech material presented during training was not included in the present study. The raters read the manual before the training session. The training session lasted approximately three hours. The raters were presented with audio examples for every coding possibility (1/2/3) for each scale with global ratings (initial statement, warmth and relationship) as well as one example of a critical and one example of a positive comment. Caveats for each scale were discussed on the basis of the presented examples. After that, two training speech samples were rated by both raters and inconsistencies were discussed. During the actual ratings, the raters used written transcriptions as well as audio files. The speech samples for the present study were rated in two cycles. First, the raters rated three randomly chosen speech samples from the present study. After this, the raters met to discuss and clarify problems in the rating procedure. Finally, the raters rated the remaining 30 speech samples in a randomized order.

3.1.4 Results

Inter-rater reliability

The first aim of the present study was to test the reliability of EE as assessed by FMSS after the described changes in the coding manual. Association analyses between EE and psychopathology have been done using either a composite score of high/low EE or the constituent subscales (Magana et al., 1986). A composite score leads to a loss of information and restriction of variance; thus, the analyses in the present study refer to the EE scales. Various coefficients were chosen for performing comprehensive reliability analyses. Krippendorff's α coefficient is the ratio between the disagreement observed within each compared unit and the disagreement expected by chance based on the responses given in all units (Honour, 2016; Krippendorff, 1970). The ICC can be broadly defined as the ratio of the variance of interest to the sum of the variance of interest plus error (Shrout & Fleiss, 1979). The ICC (Hallgreen, 2012; Shrout & Fleiss, 1979) and Krippendorff's α (Hayes & Krippendorff, 2007; Krippendorff, 1970) were used to assess inter-rater reliability because both are applicable to ordinally and intervally scaled data. The ICC was calculated using the "icc" function of the "irr" package (Gamer, Lemon, Fellows, & Sing, 2012) for the statistical software R (R Development Core Team, 2012). The "model" was set to "twoway", because all raters rated all of the material. The "type" was set to absolute "agreement" instead of consistency in rank-order. The ICC was calculated once with "unit" set to "single" (ICC_{single}) and once with "unit" set to "average" ($ICC_{average}$). ICC_{single} reflects the reliability of the ratings of a random rater who received the same training as the raters in the present study. Therefore, the ICC_{single} value can be interpreted as the reliability which is generalizable to any single random rater. The $ICC_{average}$ reflects the reliability of the average values of all raters (here two). As the average values of the two raters were used to test the association between ADHD and EE, the $ICC_{average}$ values are relevant for the present study but not generalizable to any single random rater (Shrout & Fleiss, 1979). Higher ICC values indicate better inter-rater reliability; acceptable ICC values range between .60 and 1.00 (Hallgreen, 2012). Krippendorff's α (α_{Kripp}) reflects a form of reliability comparable to the ICC_{single}, that is, generalizable to any single random rater who received a comparable training to that in the present study (Hayes & Krippendorff, 2007). It was calculated using the "kripp.alpha" function from the "irr" package (Gamer et al., 2012) for the statistical software R (R Development Core Team,

2012). The "method" was adjusted to the scale level, "ordinal" or "interval" respectively. Acceptable values for α_{Kripp} range between .70 and 1.00 (Hayes & Krippendorff, 2007). All inter-rater reliability coefficients are depicted in Table 6. The results showed unacceptable ICC_{single} values for warmth and relationship. The $ICC_{average}$ value was unacceptable only for warmth. Unacceptable values of α_{Kripp} were found for initial statement and warmth as well as critical and positive comments. The subsequent association analyses are calculated with mean values of both raters. Therefore, the $ICC_{average}$ values are relevant for the subsequent analyses. Warmth yielded unacceptable values on $ICC_{average}$ and therefore was excluded from further analyses.

Table 6
Inter-rater Reliability of the EE Scales

	ICC_{single}	$ICC_{average}$	α_{Kripp}
Initial Statement	.55	.71	.52
Warmth	.31	.47	.31
Relationship	.87	.93	.87
Critical Comments	.63	.78	.62
Positive Comments	.67	.80	.64

Note. ICC_{single} – intra-class correlation coefficient for one random rater; $ICC_{average}$ – intra-class correlation coefficient for the mean values of two raters; α_{Kripp} – Krippendorff's α; EE – Expressed Emotion

Association between ADHD symptoms and Expressed Emotion

Descriptive analyses for ADHD symptom severity and the EE scales are depicted in Table 7. Except for one answer to one item in the ADHD Index self-report, the data were complete. For the case with the missing answer, the mean score refers to nine instead of 10 items. The second aim was to investigate the association between ADHD symptoms in parent and self-reports and the EE scales. The particular expectations are that ADHD symptoms are associated with a more negative initial statement, a more negative relationship, and more critical comments and fewer positive ones. Spearman's rank correlation coefficient was selected for association analyses because the global EE scales (initial statement, warmth, and relationship) are ordinally scaled and require non-parametric coefficients. To increase comparability, the same coefficient was used for all analyses. Spearman's rank correlation coefficient ρ was

calculated for ADHD DSM and ADHD Index mean scores in parental and self-report as well as the EE scales. The results are depicted in Table 8. Contrary to expectations, the correlation between self-reported ADHD symptoms and initial statement was not significantly different from zero, but the association between the latter and parental reported ADHD symptoms was. Also in contrast to expectations, the results show no significant correlation between ADHD symptoms and relationship. Consistent with expectations, more critical comments and fewer positive comments were associated with a greater severity of ADHD symptoms in parental and self-report for both ADHD measures (ADHD DSM, ADHD Index).

Table 7
Descriptive Statistics for EE Scales and ADHD Symptoms

EE scales		ADHD symptoms		
	%			*M(SD)*
Initial Statement	1.0 – 21.21% 1.5 – 24.24% 2.0 – 48.48% 2.5 – 3.03 % 3.0 – 3.03%	ADHD (self-report)	DSM	0.97 (0.51)
Relationship	1.0 – 33.30% 1.5 – 6.06% 2.0 – 60.06%	ADHD (self-report)	Index	0.94 (0.65)
	M(SD)			
Critical Comments	2.45 (1.68)	ADHD (parental report)	DSM	0.94 (0.65)
Positive Comments	4.77 (2.14)	ADHD (parental report)	Index	1.02 (0.72)

Note. Response format for ADHD DSM and ADHD Index is a four point Likert scale (0 – not at all; to 3 – very much); response format for EE scales: critical and positive comments are frequency counts, initial statement and relationship ordinal rating scales (positive – 1;neutral – 2; negative – 3)

Table 8
Association Between EE Scales and ADHD Symptoms

	Parental report		Self-report	
	ADHD DSM	ADHD Index	ADHD DSM	ADHD Index
Initial Statement	.42 *	.50 **	.20	.16
Relationship	.06	.07	.03	.12
Critical Comments	.66 **	.70 **	.61 **	.60 **
Positive Comments	- .57 **	- .61 **	- .36 *	- .39 *

Note. Spearman's correlation coefficient ρ; ADHD DSM – mean score of the ADHD DSM criteria (APA, 2015); ADHD Index - mean score of the ADHD Index (Conners 3; Lidzba et al., 2013);
∗ $p < .05$; ** $p < .01$

Exploratory analyses

In addition to investigating the association between ADHD symptom severity and unfavorable family contexts, EE could shed light on impairments resulting from ADHD symptoms (Gordon, 2006). Ten of 33 parents (30 %) indicated that ADHD symptoms caused impairment in social life or academic functioning. Given the strong positive correlation between the EE scale critical comments and ADHD symptom severity as well as its strong negative correlation with the EE scale positive comments, the question arises as to whether the number of critical and positive comments can predict the probability of belonging to the group of parents who reported impairment due to symptoms. To answer this question, a logistic regression analysis was performed, with the outcome variable defined as impairment due to symptoms (impaired – 0; not impaired – 1) and critical and positive comments serving as predictors (both standardized to $M = 0$ and $SD = 1$ before being entered into the model). Critical and positive comments enhanced the predictability significantly; χ^2 (2) = 23.42, $p < .001$; and effect size coefficients are sufficient; Cox and Snell $R^2 = .40$, Nagelkerke's $R^2 = .57$, McFadden's $R^2 = .42$. A Hosmer and Lemeshow goodness of fit test confirms that the model fits the data; $\chi^2(8) = 11.80$, $p = .16$. More positive comments ($\beta = 1.01$, $SD = 0.55$, $z = 2.00$, $p = .047$) and fewer critical comments ($\beta = -1.84$, $SD = 0.82$, $z = -2.23$, $p = .03$) predicted the probability of belonging to the group of parents who did not indicate impairment due to ADHD symptoms significantly. Although the question of

impairment due to symptoms was asked very broadly and not specifically in terms of the family context, the results of this exploratory analysis support the notion that the EE scales critical and positive comments can serve as a valid measure of impairment due to ADHD symptoms.

3.1.5 Discussion

The aim of the present study was to investigate the reliability of the FMSS to assess EE and the association between the severity of ADHD symptoms and EE scales in a German sample of school-aged children, oversampling for children with a reported ADHD diagnosis (n = 6). In particular, the hypotheses are that more ADHD symptoms in parental and self-reports are associated with a more negative initial statement, less warmth, a more negative relationship, and more critical and fewer positive comments. Analyses revealed that inter-rater reliability was sufficient for the EE scales relationship as well as critical and positive comments when considering the reliability of one random trained rater. If the mean values of two raters are taken into account, initial statement also reached sufficient inter-rater reliability. Warmth could not be reliably measured in the present study. Only parental report measures of the severity of ADHD symptoms showed significant associations with the EE scale initial statement. The EE scale relationship did not show any association to ADHD symptom severity. The investigation of the association between the EE scales and the severity of ADHD symptoms revealed that critical and positive comments were significantly correlated with ADHD symptoms measured using both parental and self-reports and for both measures used.

Generalizable aspects of the association between EE and ADHD

The aim of the adaptations to the manual in the present study was to enhance explicitness and therefore utility and inter-rater reliability. Tone of voice was dropped completely as a rating criterion. This was done because earlier work had already described the inclusion of tone of voice as a coding criterion in measuring family context as ambiguous, as intuitive and realistic on the one hand but possibly inexplicit and unreliable on the other hand (Rutter & Brown, 1966). Tone of voice and spontaneity as indicators for warmth as well as critical and positive comments were excluded from the rating manual to enhance its comprehensibility and explicitness. These alterations did not result in a sufficiently reliable measurement of warmth. This

suggests that it is hardly possible to measure warmth using explicit coding criteria. However, critical and positive comments yielded excellent inter-rater reliability. The explicit mention of ironic statements seems to be sufficient, and tone of voice does not appear to be an indispensable coding criterion for positive and critical comments.

The present study found no consistent association between ADHD symptom severity and the EE scales in a subclinical school-aged sample. However, the association between critical and positive comments and ADHD symptom severity seems to be robust for samples with reduced symptom severity as well. The difference in associations between ADHD symptom severity and the EE scales raises the question of whether all EE scales measure the same construct. In the present study, the internal consistency of EE, including the items initial statement, relationship and critical and positive comments, was insufficient (Cronbach's $\alpha = .56$). The internal consistency of an EE composite score in a comparable study was similarly insufficient (Cronbach's $\alpha = .62$). Construct validity has been shown in pre-school samples as associations with observed behavior during free play on sensitivity and general maternal play behavior (Daley et al., 2003; Schloß et al., 2015). However, the construct validity for schoolchildren is unclear. Thus, future work must include a definition, explanation of scales and importantly, measures of construct validity (i.e. association of EE to other measures of parenting).

EE as a measure for impairment due to symptoms

One important reason to investigate the association between ADHD symptom severity and EE scales is to gain information on how these symptoms are associated with impairment in the familial environment (APA, 2015; Gordon, 2006). Based on the reliability and association found here, we divided the sample into a group whose parents indicated impairment through symptoms and a group of parents who indicated no impairment through symptoms. Exploratory analyses showed that fewer critical and more positive comments significantly predict the probability of belonging to the group of parents who indicated impairment through symptoms. This result should be verified using more suitable designs in future studies. In addition to a bigger sample size, more precise questions on impairment through symptoms in the family context should be included. Nevertheless, it demonstrates that the EE scales show associations not only with ADHD symptoms but also with perceived impairment due to ADHD symptoms, which can be seen as a validation of the EE scales. Moreover, this result points to the

potential utility of the EE scales within the diagnostic process and intervention planning.

Limitations

First, the sample size in the present study was too small and heterogeneous to use multivariate analysis strategies. Therefore, possible confounding variables of the association between ADHD symptom severity and the EE scales could not be taken into account. In particular, the influence of symptoms of conduct disorders on the association between ADHD symptoms and EE (Richards et al., 2014) should have been taken into account. Second, the danger of circularity in the associations has to be mentioned. The EE scale critical comments, for instance, captures "statements on misconduct, negative characteristics and destructivity of the child" (rating manual). According to that definition, and primed by the context of a study about attention at school, parents might have mentioned ADHD behavior in the speech sample. Mentioning ADHD behavior, which fits the definition of critical comments, would count as a critical comment. Thus, there is a danger that the EE scales to some extent reflect the intensity of ADHD symptoms.

Implications for research

The most important implication for further studies that can be derived from the present study is the need for further development of the EE construct and the manual for coding EE from the FMSS. In previous studies, raters have been trained by research labs (University of California Los Angeles: Musser et al., 2016; Peris & Hinshaw, 2003; University of Nottingham: Schloß et al., 2015; Schuh, 2015). This fact together with the insufficient inter-rater reliability in the present study implies that the EE construct is not sufficiently developed to be used solely on the basis of a manual. Moreover, the variability in the percentages of parents labeled as high EE in ADHD samples across studies is great (29% - 62 %; Peris & Miklowitz, 2015). Reliability and comparability among studies are negatively affected by these facts. In addition to the investigation of EE as a risk factor and EE in association with psychopathology, the development of the EE construct should include studies on construct validity. In this vein, EE should be put in relation to widely used self-report measures on parenting (Maguin, Nochajski, De Wit, & Safyer, 2015) or observational measures on parenting (Bertram et al., 2008). Development of a more precise manual and validity studies could help lay the basis for using EE in practice.

A promising approach for a more objective and economical way to analyze speech samples is computerized linguistic analysis. With this approach, semantic categories that distinguish children according to their level of ADHD symptoms have been identified in a preschool sample (Perez, Turner, Fisher, Lockwood, & Daley, 2014). Importantly, computerized linguistic analysis is data-driven and cannot improve and develop the EE construct. A construct validation is needed before application of linguistic analyses.

Implications for practice

Beyond the assessment of ADHD symptoms during a clinical diagnostic exploration process, starting points for an individually tailored intervention have to be figured out. The results of the present study show that EE as measured by FMSS could be an appropriate instrument to decide whether the family context is a suitable target for an intervention.

3.1.6 Conclusion

The overarching aim of the present study was to explore the utility of EE as measured by FMSS to assess family context as an important psychosocial environment of children with ADHD symptoms. The present study revealed difficulties in the reliable measurement of the EE scales initial statement and warmth. Therefore, the results of the present study suggest that conceptual work on the EE construct is needed to improve the reliable measurement and comparability of studies. Critical and positive comments showed a consistently strong association to the severity of ADHD symptoms. Thus, EE measured by FMSS is a promising tool to capture the family context relevant for children with ADHD symptoms.

Study 1 applied an observation technique to the family context, a meaningful context for children with ADHD symptoms because it is often affected from functional impairment. *Study 2* narrows down the focus from the context of ADHD symptoms to the observation of ADHD symptoms itself. A delay of gratification task has been chosen for the observation of ADHD symptoms. The delay of gratification situation provides a lab-based situation, which allows for a very standardized observation. Moreover, delay of gratification performance plays an important role for children with ADHD symptoms.

3.2 Study 2

3.2.1 Introduction

ADHD is a frequently diagnosed psychiatric disorder (prevalence ~ 5-7%) in children and adolescents, with onset before the age of twelve (APA, 2015; Willcutt, 2012). Although a categorical view of ADHD (i.e., diagnosed vs. healthy) is useful in clinical practice, ADHD itself is often considered a dimensional construct (Coghill & Sonuga-Barke, 2012).

Nature of ADHD symptoms

ADHD symptoms are divided into two symptom groups: inattention and hyperactivity-impulsivity (APA, 2015). In the DSM-5, diagnostic criteria for both symptom groups are concrete behavioral examples of inattention, activity, or impulsivity exceeding age-appropriate levels in specified contexts (APA, 2015). In addition to this trait component, a situation-dependent state component also exists. The appearance of state ADHD symptoms is dependent on the demands of the situation (Antrop et al., 2006; Imeraj et al., 2016). Over and above the situational dependence of symptom emergence, intensive longitudinal investigations have also revealed that ADHD symptoms have a general fluctuating character (Schmid, Stadler, Dirk, Fiege, & Gawrilow, 2016). Taken together, beyond the trait component, ADHD symptoms comprise a fluctuating, situational dependent and dimensionally spread component. Therefore, an investigation of behavior related to ADHD should always consider a state component above and beyond the usual trait component of ADHD.

ADHD and delay of gratification

Delay aversion is one of multiple pathways that have been proposed as causes of ADHD symptoms (Sonuga-Barke, 2002). According to this model, a preference for immediacy and subsequent experience of failure in delay situations cause delay aversion in children with ADHD (Sonuga-Barke, 2003). To overcome the aversive experience of delay, children react impulsively, hyperactively and inattentively (Sonuga-Barke, 2003). Empirical evidence shows an association between ADHD symptoms and performance in delay of gratification tasks. Delay of gratification ability in these studies has been mainly measured using (computerized) delay of gratification tasks (Merkt et al., 2016; Pauli-Pott & Becker, 2011). Even after the preschool age, which is the traditional age

for investigating delay of gratification abilities, performance in delay of gratification tasks still differs between schoolchildren with and without ADHD (Patros et al., 2016). From a theoretical perspective, the association between ADHD and performance in delay of gratification tasks has been mostly explained by delay aversion as a motivational style (Antrop et al., 2006; Pauli-Pott & Becker, 2011). Despite this suitable theoretical underpinning, what role ADHD symptoms play over and above more general factors that influence performance in delay of gratification tasks remains an open question.

Delay of gratification

An essential part of volitional self-control is delay of gratification, meaning the ability to effectively delay a smaller reward for the sake of a larger but delayed reward (Mischel, 1996). In a delay of gratification task, a standardized situation (e.g., one marshmallow now vs. two marshmallows later) is created that allows for the investigation of delay of gratification ability as well as factors influencing this ability. The original delay of gratification task with a single waiting period and a treat as reward can be labeled as a *waiting task* (Mischel, 1996). In contrast, *choice task* paradigms often use computerized tasks with token economies (e.g., collected points can be exchanged for money). These tasks usually use the number of times a larger, later reward is selected rather than a smaller, sooner reward in multiple trials as an outcome measure. The ability measured by choice tasks that incorporate only a hypothetical delay rather than an actual delay is referred to as delay discounting (Neubauer, Gawrilow, & Hasselhorn, 2011; Patros et al., 2016; Reynolds & Schiffbauer, 2005). Performance in waiting tasks at preschool age has been shown to predict academic achievement and social competence ten years later (Shoda, Mischel, & Peake, 1990). The predictive validity of performance in the delay of gratification task is substantial and supports the relevance of the paradigm. Thus, factors facilitating a child's decision to wait have been investigated. One crucial factor influencing performance seems to be the way in which the reward is mentally represented (Mischel, 1996; Mischel & Baker, 1975; Mischel, Ebbesen, & Zeiss, 1972; Mischel & Moore, 1973). A conceptualization of mental representations is provided by the *hot and cool system* (Metcalfe & Mischel, 1999). Abstract, informative, and iconic mental representations of a reward foster self-control, in this case waiting for the larger but delayed gratification. Mental representations that highlight the arousing, motivating, and consummatory features of a reward impede self-control. One operationalization of mental representations is the

investigation of spontaneous attention orientation during a delay situation. Attention orientation towards the reward indicates a hot representation, whereas attention orientation away from the reward indicates a cool representation. In the same vein, less attention orientation towards the delayed reward and more orientation elsewhere have been found to be associated with better performance in delay of gratification tasks (Eigsti et al., 2006; Manfra, Davis, Ducenne, & Winsler, 2014; Neuenschwander & Blair, 2017; Rodriguez, Mischel, & Shoda, 1989; Vaughn, Kopp, Krakow, Johnson, & Schwartz, 1986)

ADHD symptoms and mental representation of the reward

The evidence for performance differences in delay of gratification tasks between school-aged children with and without ADHD symptoms stems mainly from choice tasks (Patros et al., 2016). Assessing delay of gratification ability with choice tasks has important implications. Attention orientation during delay time is one possible operationalization of the mental representation of the reward. Obstacles to the measurement of attention orientation in choice tasks are very short delay periods (e.g., 30 sec) and the absence of the actual reward (Kuntsi et al., 2001). Choice tasks seem to be designed to capture choice preferences and the initial value of the reward, whereas waiting tasks capture the ability to sustain that choice (Reynolds & Schiffbauer, 2005). Moreover, making inferences about hot or cool mental representation from attention orientation in choice tasks is probably different than in waiting tasks and not directly possible. Without information on attention orientation, it is hardly possible to answer the following question: Do ADHD symptoms (state and trait) explain performance differences in delay of gratification waiting tasks over and above attention orientation? Therefore, we describe the assumed impact of ADHD symptoms on attention orientation in a delay of gratification waiting task in the paragraphs that follow.

Impulsivity is the ADHD symptom that has been discussed as the causal factor of lower performance in delay of gratification tasks (Patros et al., 2016; Reynolds & Schiffbauer, 2005). The description of impulsivity stresses excessive talking, interrupting of others, and not waiting one’s turn (APA, 2015). That fits with mental representations consisting predominantly of the hot features of a stimulus for the state and trait components of impulsivity. This makes lower performance in delay of gratification tasks plausible. Inattention is characterized by a lack of attention to details, a lack of sustained attention, easy distractibility, forgetfulness, and poor organization (APA, 2015). In a delay of gratification task, there are no situational demands or

challenges to attention. Therefore, it is not possible to observe a lack of attention paid to situational demands, i.e. inattention, during the delay situation. On the one hand, higher trait inattention could help a person maintain a cool representation (i.e., not concentrating on the reward). On the other hand, it might also work against a cool representation of the reward (i.e., causing problems keeping sustained attention on an object other than the reward). In the same vein, (hyper)activity in the delay situation could facilitate a cool representation of the reward by intentionally or unintentionally distracting one from the reward by way of body movements. On the other hand, more (hyper)activity during the delay situation could prevent goal-directed distraction from the reward, resulting in a hot representation of the reward and thus worse performance.

3.2.2 Present Study

The aim of the present study is to investigate the influence of attention orientation, state ADHD symptoms (activity and impulsivity) and trait ADHD symptoms (diagnosis) on performance in a delay of gratification waiting task in schoolchildren with and without ADHD. We expect less attention orientation towards the reward and more attention orientation towards another object to be a significant predictor of better performance in a delay of gratification waiting task. Furthermore, we hypothesize an ADHD diagnosis to be a predictor of worse performance in a delay of gratification waiting task. We also assume impulsive behavior during the task to be a predictor of worse performance in a delay of gratification task. Finally, we expect that activity can affect performance positively and negatively. Thus, we do not expect activity to be a significant predictor of performance in a delay of gratification waiting task.

3.2.3 Method

Participants

The current study is part of the GIDeCA project (IDEA Center Indivdual Development and Adaptive Education, 2011). Part of the data used in this study has been published before (Reinelt, Wirth, Rauch, & Gawrilow, 2014). All children in that project participated in a delay of gratification waiting task that was video recorded. Inclusion criteria for the ADHD group were (1) an ADHD diagnosis on a structured

interview for psychiatric disorders in children (Kinder-DIPS; Schneider, Unnewehr, & Margraf, 2009; based on diagnostic criteria from DSM-IV TR; APA, 2003) and (2) an ADHD diagnosis by an external pediatrician or psychiatrist. Inclusion criteria for the non-ADHD group were (1) no ADHD diagnosis on a structured interview for psychiatric disorders in children (Kinder-DIPS; Schneider et al., 2009; based on diagnostic criteria from DSM-IV TR;APA, 2003) and (2) no ADHD diagnosis by an external pediatrician or psychiatrist. Thirty-one children met the inclusion criteria for the ADHD group and 55 children met the inclusion criteria for the non-ADHD group. The video material from some children was unusable due to the following reasons: (1) incomplete material, (2) items the child took to the waiting situation or (3) an incorrect number of chocolate bars (rewards) present during the waiting situation. After the exclusion of the children with unusable video material, 26 children remained in the ADHD group and 35 children in the non-ADHD group. Thus, the total sample selected for analysis consisted of 61 children. Information on sex, medication status for ADHD treatment, age (min = 7.33 years; max = 13.67 years) and IQ (culture fair intelligence test CFT 20-R;Weiß, 2006) for the entire sample, the ADHD group and the non-ADHD group are displayed in Table 9. In cases where children received medication for ADHD treatment, parents were asked to omit the medication on the day of assessment. Parents gave written and informed consent to their child's participation in the study. The study was approved by the local ethics committees.

Table 9
Descriptive Statistics for the Sample

	Complete Sample	ADHD	Non-ADHD	Did not Wait	Waited
	N = 61	*n* = 26	*n* = 35	*n* =10	*n* = 51
	n	*n*	*n*	*N*	*n*
Male	47	21	26	39	10
Medicated	13	13	0	12	1
	M(SD)	*M(SD)*	*M(SD)*	*M(SD)*	*M(SD)*
Age	10.40 (1.58)	10.84 (1.68)	10.08 (1.45)	10.41 (1.58)	10.35 (1.70)
IQ	109.45 (14.45)	103.76 (14.31)	113.51 (13.31)	108.74 (14.87)	113.00 (12.19)
Waiting Time	23.06 (5.34)	24.12 (3.62)	22.27 (6.26)	13.18 (7.76)	25.00 (0.00)
Activity (Rating)	.18 (0.13)	.19 (0.14)	.16 (0.13)	.25 (0.17)	.16 (0.12)
Impulsivity (Rating)	.18 (0.19)	.21 (0.23)	.16 (0.15)	.28 (0.28)	.16 (0.16)
Goal-driven Attention Orientation (Rating)	.49 (0.16)	.51 (0.12)	.47 (0.18)	.36 (0.21)	.52 (0.14)
Stimulus-driven Attention Orientation (Rating)	.31 (0.16)	.32 (0.16)	.30 (0.15)	.43 (0.24)	.28 (0.12)

Note. One IQ value is missing for a child that was diagnosed with ADHD and waited for two chocolate bars.
12 of 13 children who were medicated took methylphenidate, one parent did not specify a drug name.

Procedure

Delay of gratification waiting task. Testing took place in a laboratory in a university building. Apart from furniture (i.e., table, chairs, and a separating wall), the room was empty, and the blinds on the windows were lowered during the task. The experimenter put a plate and a bell on the table and showed the child four different chocolate bars, asking which of the bars the child preferred. The experimenter asked the child to unwrap the preferred bar and put it on a plate and removed the other bars. Then the experimenter read the following instructions: "I will go and sit down behind the separating wall. You can choose to eat the chocolate bar immediately. Or you can choose to wait until I come back. In this case you would get two of the chocolate bars. If you do not want to wait any longer, you can ring the bell. Then, I will come back, the game is over and you won't get a second bar." After that, the experimenter put a second, still wrapped bar of the same preferred type next to the plate. Next, the experimenter repeated the rules, asked the child whether he or she had any remaining questions, and requested that the child hands over any watches or mobile phones to the experimenter. Then, the experimenter went behind the separating wall. If the child did not ring the bell, the experimenter came back after 25 min. Delay of gratification performance was operationalized as (a) waited or (b) did not wait for the two chocolate bars, and continuously as the amount of time a child waited before ringing the bell.

Control questions. After the delay of gratification waiting task, the experimenter asked the child the following questions: (1) "Did you try to distract yourself from the chocolate bar during the waiting situation?" (yes/no), (2) "When (how many hours ago) was your last meal?". We asked that question to check whether hunger, assessed by the amount of time since one's last meal, could explain the difference between children who managed to wait for the second chocolate bar and those who did not.

Rating scales for attention orientation, activity, and impulsivity. The assessment of attention orientation captured (1) attention paid to the reward on the one hand and (2) attention paid to any other object except from the reward on the other hand. Visible attention orientation to any other object was interpreted as driven by the goal of getting two chocolate bars; therefore, it was labeled goal-driven attention orientation. Goal-driven attention orientation (e.g., Colombo, 2001) is defined as internally, volitionally directed attention orientation. Here, we define goal-driven attention orientation in the delay of gratification situation as attention orientation towards and engagement with any object that is not related to the reward (i.e., bell and chocolate bar). In the language of

the hot/cool system, goal-driven attention orientation fosters cool representations of the reward. Visible attention towards the reward is interpreted as attention driven by the reward; therefore, it is labeled stimulus-driven attention orientation. Stimulus-driven attention orientation (Alvarez & Freides, 2004) is defined as reflexive and automatic. Here, we define stimulus-driven attention orientation in the delay of gratification situation as attention orientation towards the bell or chocolate bar. In terms of the hot/cool system, stimulus-driven attention orientation fosters hot representations of the reward.

Activity and impulsivity during the delay time were measured on the basis of the description of ADHD symptoms in the DSM-5 (APA, 2015) and the German version of the Conners 3 Rating Scales (Lidzba et al., 2013). We transformed hyperactivity items into activity items because hyperactivity refers to activity that is inappropriate for the demands of a given situation. However, situational demands were not specified for the delay of gratification waiting task. For the same reason – a lack of situational demands to be attentive to – inattention could not be assessed. In order to be included, items had to fulfill the following criteria: (1) No event relation within the item. Thus, items like "blurts out answers before the question has ended" were excluded. (2) Assessable without temporal contingencies. Thus, items like "*often* talks excessively" were excluded. (3) Assessable in the specific waiting situation without information beyond sound and picture from the video. Thus, items like "difficulties playing quietly during leisure time" were excluded. Additionally, we examined the video material to check whether the selected items could be assumed to vary between participants. All rating items used in the present study are displayed in Table 10.

Preparation of video ratings. First, the appropriate time period of the video was selected, starting from the moment the experimenter went behind the separating wall and ending with the return of the experimenter. Each video was cut into 2 min clips. In cases where three or more clips existed, the first and the last clip were excluded from the analyses. Thus, we only included video material showing the child during the waiting time in 2 min clips and excluded video material showing parts of the interaction with the experimenter or clips shorter than 2 min. Every item was judged in every two minute clip. The scales goal-driven and stimulus-driven attention orientation both consist of one item each on visual, auditory, tactile, and olfactory-gustatory attention orientation. We designed all attention orientation and activity items as events that either (1) occurred or (0) did not occur during a given 2 min clip. On the other hand, the

impulsivity items (e.g., "seems to be restless") do not describe single occurrences of behavior, but rather a general tendency over a period of time. Therefore, they were scored depending on the behavior observed over the entire period of the clip (i.e., (1) if the behavior was present during the whole clip and (0) otherwise).

Table 10
Video Rating Items Delay of Gratification Task

	Waited (n = 12)	Did not wait (n = 49)
Activity	rocks the trunk	rocks the upper part of the body
	fidgets with hands or feet	fidgets with hands or feet
	squirms in the chair	squirms in the chair
	gets up from the chair	gets up from the chair
	climbing around	climbing around
Impulsivity	seems to be restless and uneasy	seems to be restless and uneasy
	seems to have difficulties	seems to have difficulties
	pulling his/herself together	pulling his/herself together
Goal-driven Attention Orientation	looks at something (Except bell/chocolate bar)	**looks towards a thing (Except bell/chocolate bar)**
	touches something (Except bell/chocolate bar)	**takes something in his/her hand (Except bell/chocolate bar)**
	tastes or smells something (Except bell/chocolate bar)	tastes or smells something (Except bell/chocolate bar)
	makes noise (Except with bell/chocolate bar)	makes noise (Except with bell/chocolate bar)
Stimulus-driven Attention Orientation	looks at the chocolate bars/bell	looks at the chocolate bars/bell
	touches the chocolate bars/bell	touches the chocolate bars/bell
	tastes or smells chocolate bars/bell	tastes or smells chocolate bars/bell
	makes noises with chocolate bar/bell	makes noises with chocolate bar/bell

Note. Items changed in rating the participants who waited compared to those who did not wait are in bold. In the rating of the participants who did not wait, we included two randomly selected children who waited.

Rating procedure in the delay of gratification waiting task. The sequence of participants was randomized and thus different for each rater, whereas the sequence of clips of each child remained chronological. Four raters (undergraduate psychology students) rated all the video material from participants who did not wait plus two

randomly selected children who waited, with three of them continuing to rate the remaining material. Rater watched the videos with a standard media player and headphones. Items were filled in in form of paper-pencil questionnaires. Three training sessions took place. First, the raters received written descriptions and video examples of all items using video material that was excluded from the present study. An 'occurred' and a 'did not occur' video example were presented for each item. Subsequently, a practice rating was conducted and discrepancies were discussed. In the second training session, the descriptions of the scales were discussed again and another practice rating was conducted. In the third training session, the adapted items were discussed, and we reviewed the descriptions of the rating scales. After this first rating step, feasibility and inter-rater reliability were checked using the video material from participants who did not wait plus two randomly selected children who waited, thus allowing us to evaluate the rating scales that we developed. We made slight adaptations to two items after this first rating (see Table 10). All raters were blind with regard to the ADHD status of the participants and rated the videos independently.

Aggregation of ratings. Hypothesis testing in the present study took place using one value per scale for each participant. To reflect the intensity on the four developed scales (activity, impulsivity, stimulus-driven attention orientation, goal-driven attention orientation), each participant's average score over all clips and all items of a given scale was calculated.

Inter-rater reliability. Because we rated the participants' video material in two cycles, we calculated inter-rater reliability separately as well (see Table 10). We selected the ICC (A, 2) as appropriate for the present study (Hallgreen, 2012; Shrout & Fleiss, 1979). We used the "icc" function of the "irr" package (Gamer et al., 2012) of the statistical software R (R Development Core Team, 2012). The "model", "type" and "unit" can be specified in this function. Our rating design was fully-crossed, meaning the same set of raters evaluated all the material in both rating cycles. Therefore, we selected a two-way model that accounts for systematic deviations due to specific raters. For our purposes, inter-rater reliability was defined as absolute agreement instead of similarity in rank-order (consistency). For that reason, the ICC type was agreement. The mean values from all raters were used for hypothesis testing (see above). Thus, we used the unit average instead of single. Higher ICC values indicate better inter-rater reliability and acceptable ICC values range between .60 and 1.00 (Hallgreen, 2012).

ICC values for the present study are listed in Table 11. All values indicate excellent inter-rater reliability.

Table 11
Inter-Rater Reliability of Video Ratings in the Delay of Gratification Task

	$ICC_{average}$ Did not wait (n = 12)	$ICC_{average}$ Waited (n = 49)
Activity	.95	.95
Impulsivity	.88	.85
Goal-driven Attention Orientation	.92	.95
Stimulus-driven Attention Orientation	.96	.96

Note. $ICC_{average}$ – Intraclass correlation coefficient, In the rating of the participants who did not wait, we included two randomly selected children who waited. Three raters judged the participants who did not wait; four the participants who waited.

3.2.4 Results

General performance in the delay of gratification waiting task

Ten (13.36%) of the 61 children did not wait for the second chocolate bar. Two (3.28%) of the children who did not wait had an ADHD diagnosis. The mean waiting time of the entire sample was 23.06 min (*SD* = 5.34). The children who did not wait for the second bar had an average waiting time of 13.18 min (*SD* = 7.76). In subsequent analyses, we operationalized performance on the delay of gratification waiting task as a categorical variable (waited vs. did not wait). Conclusions on factors that increase or decrease the probability of waiting or not waiting seemed more meaningful because of the generally high probability of waiting. However, analyses yielded the same results when performed with waiting time as the dependent measure.

Control questions

Among the participants who waited, 13 children (25.5 %) reported that they did not try to distract themselves from the chocolate bar during the waiting situation. Among the participants who did not wait, two (28.6 %, one answer missing) of the children reported that they did not try to distract themselves from the chocolate bar during the waiting situation. Children who waited had their last meal on average 4.36 h ago (*n* = 46, five answers missing, *SD* = 3.21) and children who did not wait reported having had their last meal on average 5.56h ago (*n* = 8, two answers missing,

SD = 5.51). A two-sided Wilcoxon rank sum test for unpaired samples did not reveal a significant difference between those two (W = 201.5, p = .68, Cohen's d = .33).

Statistical Hypotheses Testing

We performed a logistic regression analysis with the outcome variable "waiting" (0 – did not wait; 1 – waited) and the predictors goal-driven attention orientation, stimulus-driven attention orientation, activity, impulsivity and diagnosis (0 – non-ADHD group; 1 – ADHD group). As age (in years), sex (1 – male, 0 - female), IQ (Patros et al., 2016; Rodriguez et al., 1989) and medication status (0 – not medicated, 1 – medicated) may influence performance in delay of gratification tasks, we entered those variables into the model as control variables. The scale for the video-rated predictors (goal-driven attention orientation, stimulus-driven attention orientation, activity, impulsivity) ranged from 0 (behavior not at all present) to 1 (behavior totally present). To check for multicolliniarity, we calculated variance inflation factors (VIF) for all predictors. No VIF value exceeded 10; thus, we supposed no multicolliniarity in our predictor set that would endanger the interpretability of the model (O'Brien, 2007). To improve the interpretability of the model, we standardized all continuous predictors to have M = 0 and SD = 1. One IQ score was missing. To be able to use the other values for that person, we imputed the missing IQ score with the mean value of the remaining IQ score. Results of the final logistic regression model are depicted in Table 12. A likelihood-ratio test for the predictor set of the entire model compared to a model without predictors was significant ($\chi^2(9)$ = 30.28, p < .001), indicating that the set of predictors explained the probability of belonging to the group of children who waited significantly better than the probability of having waited in general. Effect size coefficients were sufficient: Cox and Snell R^2 = .39, Nagelkerke's R^2 = .66, McFadden's R^2 = .56. Hosmer and Lemeshow's goodness of fit test confirmed that the model fits the data well ($\chi^2(8)$ = 7.85, p = .45).

We expected less stimulus-driven attention orientation and more goal-driven attention orientation to be a significant predictor of better performance in the delay of gratification waiting task. In line with this hypothesis, more goal-driven attention orientation increased the probability of waiting (β = 2.59, SD = 0.10, z = 2.56, p = .01) and less stimulus-driven attention orientation increased the probability of waiting (β = -1.82, SD = 0.89, z = -2.04, p = .04). Odds ratios show that an increase of one unit on the predictor goal-driven attention orientation increased the probability of waiting 13.27 times. The odds ratio of stimulus-driven attention orientation is 0.16. That means an

increase of one unit on the predictor stimulus-driven attention orientation decreased the probability of waiting by a factor of 6.20 (1 divided by 0.16 equals 6.20). We expected an ADHD diagnosis to be a predictor of worse performance in the delay of gratification waiting task. The model revealed that an ADHD diagnosis was not a significant predictor of waiting (β = 1.73, *SD* = 1.78, *z* = 0.97, *p* = .33). We expected impulsive behavior to be a predictor of performance on the delay of gratification waiting task. Contrary to our expectations, impulsivity was not a significant predictor of waiting (β = 0.39, *SD* = 0.88, *z* = 0.45, *p* = .65). We did not expect activity to be a significant predictor of performance on the delay of gratification waiting task. Contrary to our expectations, activity during the delay situation decreased the probability of waiting significantly (β = -2.64, *SD* = 1.26, *z* = -2.09, *p* = .04). The odds ratio of activity was 0.07. That means an increase of one unit on the predictor activity decreased the probability of waiting by a factor of 13.98 (1 divided by 0.07 equals 13.98). None of the control variables (age, medication status, sex, or IQ) predicted waiting significantly

Table 12
Results of the Logistic Regression Analysis for the Outcome Waiting Status and VIF for Predictors (Multicolliniarity)

Predictor	β	*SE* β	*z*	*p*		e^{β}	VIF
Constant	2.79	1.75	1.60	.11			
Goal-driven Attention Orientation	2.59	0.10	2.60	.01	**	13.27	4.43
Stimulus-driven Attention Orientation	-1.82	0.90	-2.04	.04	*	0.16	4.49
Activity	-2.64	1.26	-2.09	.04	*	0.07	8.27
Impulsivity	0.39	0.88	0.45	.65		1.48	5.11
Diagnosed	1.73	1.78	0.98	.33		5.62	2.54
Age	-0.33	0.71	-0.46	.64		0.72	2.07
Medicated	-2.10	2.55	-0.82	.41		0.12	3.06
Male	1.04	1.53	0.68	.48		2.83	1.70
IQ	-1.65	1.14	-1.45	.15		0.19	3.72

Note. Outcome variable waiting status is coded as 1 - waited; 0 – did not wait;.β – regression coefficient (ln of the odds ratios); *SE* β standard error of the regression coefficient; e^{β} – odds ratio of the predictor variables; VIF – Variance inflation factor; * < .05; ** < .01

3.2.5 Discussion

The aim of the present study was to investigate the influence of ADHD symptoms and attention orientation on performance in a delay of gratification task in schoolchildren with and without ADHD. In line with our expectations, less stimulus-driven attention orientation and more goal-driven attention orientation predicted waiting in the delay of gratification waiting task. Contrary to our expectations, an ADHD diagnosis and state impulsive behavior were not significant predictors of waiting. However, less activity during the delay of gratification waiting task turned out to be a significant predictor of waiting.

ADHD symptoms in a delay of gratification waiting task

In our study, children with an ADHD diagnosis did not perform worse in the delay of gratification waiting task than children without an ADHD diagnosis. Our finding is in line with previous studies that have not found general deficits in delay of gratification tasks among school-aged children with ADHD (Gawrilow, Gollwitzer, & Oettingen, 2011; Sjöwall, Roth, Lindqvist, & Thorell, 2013) Furthermore, impulsive behavior during the delay time was not a significant predictor of performance on the delay of gratification waiting task. Although deciding in favor of immediate gratification certainly resembles impulsive behavior (Patros et al., 2016; Reynolds & Schiffbauer, 2005), assessments of impulsive behavior during the waiting time in the present study ("seems to be restless and uneasy," "seems to have difficulties pulling him/herself together") remained mainly on the interpretative level. That is because directly observable impulsive behavior requires an event: for instance, a person has to talk for someone else to be able to interrupt her or him. During the delay situation, ringing the bell was the only directly observable impulsive behavior. The assessment of impulsive behavior during the waiting time in the present study probably instead captured behavior resulting from successfully suppressing the urge to ring the bell. That could be the reason why impulsive behavior in the present study was not a significant predictor of performance on the delay of gratification task. In contrast to impulsivity, direct observation of activity during the waiting task was possible, and the item content is comparable to the behavioral descriptors of hyperactivity in the DSM-5(APA, 2015). Children who were more active during the delay situation had a smaller probability of waiting than children who were less active. This supports the notion that activity during the delay situation prevents goal-directed distraction from the reward, resulting in less

cool and more hot representations of the reward and consequently in worse performance. Taken together, the consideration of state as well as trait ADHD symptoms yielded incongruent results, which supports the combined state and trait view of ADHD symptoms we took in the present study.

Attention orientation in a delay of gratification waiting task

In line with previous findings and our expectations, attention orientation was a strong predictor of performance on the task (Eigsti et al., 2006; Manfra et al., 2014; Rodriguez et al., 1989; Vaughn et al., 1986). Although the general probability of waiting was very high, goal-driven and stimulus-driven attention orientation strongly influenced the probability of waiting. Interestingly, the proportion of children who reported that they did not try to distract themselves from the reward was approximately the same in the groups of children who waited (25.5%) and who did not wait (28.7%). Apparently, the objective behavioral observation we conducted in the present study is needed. Importantly, we only rated audible and visible behavior in order to assess attention orientation, such as "touches the chocolate bars/bell". Inferences regarding whether the behavior was stimulus-driven or goal-driven remain subject to interpretation. Our results confirm that the most important factor influencing performance on a delay of gratification task is the mental representation of the reward (Mischel, 1996); the assessment of spontaneous attention orientation seems to be a suitable operationalization of mental representation.

Strengths and limitations

The present study makes several contributions to the investigation of delay of gratification performance in children with different levels of ADHD symptoms. The objective behavioral observation during a delay of gratification waiting task guarantees that the information on behavior during the task is free from social desirability and recall biases. In the current study we used state (impulsivity and activity in video ratings) as well as trait (ADHD diagnosis) components of ADHD symptoms. This comprehensive approach was crucial to the results of the study. The differential results for state and trait components should encourage future studies to consider them both. Our approach of investigating the impact of ADHD symptoms on measures of self-control from the perspective of how the actual behavioral expression of ADHD symptoms influences outcome variables with regard to self-controlled behavior is promising.

ADHD symptoms have a heterogeneous cognitive foundation (Sonuga-Barke, 2002). Therefore, the DSM-5 relies on behavioral descriptors in the diagnostic process (Tannock, 2013). These behavioral descriptors stress the importance of investigations of the actual behavioral expression of ADHD symptoms in order to gain ecologically valid and generalizable results. The present study fits well with that approach.

The generalizability of the study is limited. The number of children who managed to wait for the second chocolate bar was very high in the current sample. The original age of investigation for the paradigm used in this study is preschool age; variance in waiting time decreases with age (Ayduk et al., 2000; Rodriguez et al., 1989). Thus, the investigation of factors that enhance or decrease the probability of waiting is limited due to this restricted variance. Furthermore, the complete sample size is very small. A recent meta-analysis revealed for school-aged children a medium effect for a group comparison on performance in delay of gratification tasks between children with and children without ADHD(d = .46; Patros et al., 2016). A post-hoc power-analysis for a one-tailed group comparison with n_{ADHD} = 26 and $n_{non\text{-}ADHD}$ = 35, α = .05 and the effect size d = .46 revealed a power of .54 for this group comparison in the present study. Although this calculation is not directly transferable to the analyses we performed in the present study, it gives hints, that the sample size was too small to detect an effect of an ADHD diagnosis on performance in a delay of gratification task. We did not use comprehensive manipulation checks, such as self-reported cravings, during waiting time. This type of information could improve our understanding of the performance patterns in the study, because cravings lie at the root of the incentive to behave in a self-regulated way within the delay of gratification waiting task implemented in this study. Although the difference in the amount of time past after the last meal for children who waited and children who did not wait was not statistically different, the difference exceeded one hour which is substantial for the age group in the present study. Furthermore, children who are usually medicated in the context of ADHD treatment were asked to omit medication on the day of assessment. However, no distinctions were made with regard to type of medication. For this reason, some children might have been medicated with extended-release preparations, meaning that the effect of medication was probably still present during the testing session.

Implications for further research

In contrast to most existing studies on ADHD symptoms and delay of gratification beyond preschool age, the present study uses a waiting task rather than a

choice task (Patros et al., 2016). One study has tested the ecological validity of choice tasks (Solanto et al., 2001). However, the surprisingly good performance of children diagnosed with ADHD in the present study raises the question of whether convergent validity for both tasks can be assumed. Choice tasks usually contain a far more abstract conceptualization of the reward, which could affect mental representations of it (Metcalfe & Mischel, 1999). Future studies should focus on this question. Moreover, in light of the surprising good performance of children with ADHD in this study, whether self-regulatory strategies (e.g. attention orientation) work differently for children with different preconditions, such as an ADHD diagnosis (Neuenschwander & Blair, 2017; Sturge-Apple et al., 2016), should be investigated.

3.2.6 Conclusion

The main finding of the present study is that attention orientation explains performance in a delay of gratification waiting task over and above an ADHD diagnosis or impulsive behavior during the waiting situation. Furthermore, children who exhibit more activity during the waiting situation are also more likely to exhibit worse performance on the delay of gratification waiting task. Thus, considering situational driven state components and more general trait components of ADHD symptoms as assessed in a behavior-based delay of gratification setting revealed no general performance deficits in delay of gratification tasks among children with ADHD symptoms.

The focus of *Study 2* was the observation of ADHD symptoms in a delay of gratification task and the relation to performance in a delay of gratification task. The delay of gratification task provided a lab-based setting for the observation of ADHD symptoms. In *Study 3*, the observation of ADHD symptoms takes place in a more ecological valid context, the classroom. ADHD symptoms frequently lead to disturbances in the classroom context, and academic achievement is significantly lower in children with more ADHD symptoms (Frazier et al., 2007). Therefore, *Study 3* targets the classroom context for the observation of ADHD symptoms.

3.3 Study 3

3.3.1 Introduction

Assessment of ADHD symptoms in the school context

Schoolchildren with ADHD are described as inattentive, hyperactive, and impulsive (APA, 2015). Inattention can lead to problems finishing tasks, careless mistakes, and losing things necessary for schoolwork. Hyperactive behavior can be described as fidgeting around, leaving one's place in the classroom or acting "as if driven by a motor," for instance. Examples of impulsive behavior are blurting out answers or talking excessively when one is not supposed to. All these behaviors lead to severe interruptions in the classroom. They not only disturb the affected children's classmates, teachers, and the teaching process but also hinder the student's own learning.

Thus, ADHD symptoms exhibit their effects at school, causing social and academic impairments. Empirical evidence confirms that impairments in peer relationships (Bagwell, Molina, Pelham, & Hoza, 2001; Thorell, Sjöwall, Diamatopoulou, Rydell, & Bohlin, 2016), task inappropriate behavior (Kofler et al., 2008) and problems in academic achievement (DuPaul, Reid, Anastopoulos, & Power, 2014) are associated with ADHD symptoms. The describe how functional impairment not only affects children with an ADHD diagnosis, but increases steadily with symptom severity, even in the subclinical range (Norén Selinus et al., 2016). Behavioral observation is an assessment instrument that can include context directly, in contrast to reports. Hence, the direct observation of behavior within the school context might be a useful complement to routine questionnaire assessments. This is because it may allow for the measurement of the core ADHD symptoms of inattention, hyperactivity, and impulsivity in a real-life situation that is often difficult for schoolchildren showing these ADHD symptoms. Thus, the present study explores the potential of behavioral observation of ADHD symptoms in a simulated classroom situation.

Observation of ADHD symptoms in the classroom

Reports of ADHD symptoms relate to the experiences a parent, teacher or the concerned child him- or herself have had regarding the child's ADHD. In contrst to symptom reporting, direct observation directly evaluates target behavior in a certain

situation (Furr & Funder, 2007). Questionnaires collapse the occurrence of ADHD symptoms in many contexts into a general behavioral disposition independent of the contextof interest. Behavioral observation has the potential to target the classroom as a relevant context. The following paragraph contains a review of studies that have applied behavioral observation of ADHD symptoms in a classroom context. A meta-analysis on differences in inattentive behavior in the classroom between children with and without ADHD reveals that children with ADHD show more off-task behavior compared to their healthy peers across classroom settings, observation techniques, and age groups (Kofler et al., 2008). Children with ADHD do not only show more off-task behavior, but oscillate more between off- and on-task behavior in natural classrooms (Rapport et al., 2009). This means that task inapproproate behavior in children with ADHD fluctuates and is not always present to the same extent. Some studies found that children with ADHD tend to exhibit symptoms in unstructured classrooms more than structured ones (Imeraj et al., 2016; Roberts, 1990), while others did not find an effect of classroom type on ADHD symptoms (Lauth et al., 2006). Direct observation of ADHD symptoms should consider both structured and unstructured classroom types. In most cases, the behavioral coding for the observation of ADHD symptoms in the cited studies is on-task vs. off-task behavior. This operationalization is quite distal to the behavioral descriptions of ADHD symptoms in the diagnostic manual. On-task and off-task coding is used in behavioral observation studies unrelated to ADHD as well (Godwin et al., 2016). In the present study, behavioral codes for ADHD symptoms are aligned with behavioral descriptions in the diagnostic manuals.

Methodologically, the direct observation of ADHD symptoms in classrooms faces two challenges: the *selection of the setting*, and the *behavioral coding of ADHD symptoms*. First, the *selection of the setting* for behavioral observation includes a trade-off between ecological validity and standardization. The advantage of natural classroom environments is their ecological validity. Observational studies that incorporate the type of the natural classroom environment usually distinguish between structured (nonidle time or regular lesson) and unstructured (idle time or noninstructional time) classroom environments (Imeraj et al., 2016; Lauth et al., 2006). Simulated classroom environments establish relevant classroom situations artificially, and ADHD symptoms are observed in this context. The ecological validity is lower, but the higher standardization level allows for greater influence on the setting and therefore more precise statements on behavioral differences in a specific environment. Many more

facets of behavior can be observed and coded than in natural classroom environments (Cunningham & Siegel, 1987; Jacob et al., 1978; Roberts, 1990). For instance, a simulated classroom oberservation study determined that the most pronounced behavioral differences between hyperactive and aggressive boys arose in a school setting. Other settings did not reveal such differences (Roberts, 1990). In the present study, a simulated classroom setting was chosen to keep the comparability between observations high, thus allowing for an investigation of the psychometric properties of new rating items.

Second, *behavioral coding of ADHD symptoms* is not trivial in many cases. The descriptions of diagnostic criteria for ADHD are usually linked to specific sitations, for instance, getting up from one's seat when remaining seated is expected (APA, 2015). It is hardly possible to recreate all the situations delineated in the description of symptoms. Moreover, ADHD symptoms are expressed as traits; these trait ADHD symptoms have to be transformed into observable states before they can be directly observed. Traits can be defined as general behavioral dispositions, whereas states describe locally and temporally situated behavior (Allport & Odbert, 1936). Thus, observed and reported ADHD symptoms shed light on the construct from different perspectives. In the present study, rating items were developed to closely align with the behavioral descriptions in the diagnostic manuals in order to keep the information obtained from observation close to those behavioral descriptions.

ADHD symptoms and academic achievement

Another important reason for behavioral observation of ADHD symptoms relates to deficits in academic performance among children with ADHD symptoms, which are well documented in the literature (Barry, Lyman, & Klinger, 2002; Dave Daley & Birchwood, 2010; Frazier et al., 2007; Loe & Feldman, 2007; Scholtens, Rydell, & Yang-Wallentin, 2013; Tymms & Merrell, 2011). A meta-analysis revealed that an overall comparison of academic achievement among children with and without ADHD shows a medium to large effect, with children with ADHD performing worse (d =.71; Frazier et al., 2007). This finding generalizes to a community sample (Merrell & Tymms, 2001). Interestingly, the achievement gap between children with high scores and low scores on inattention is much bigger than the gap between children with high and low scores on hyperactivity and impulsivity (Gaub & Carlson, 1997; Merrell & Tymms, 2001). Thus, inattention seems to hamper academic achievement to a greater extent than hyperactivity and impulsivity. Narrowing the focus to mathematical

abilities, a strong negative association between ADHD symptoms and mathematical ability could be found. Here again, the association between inattentive symptoms and mathematical abilities is stronger than the association between hyperactive and impulsive symptoms and mathematical abilities (Tosto, Momi, Asherson, & Malki, 2015). The strong negative association between academic achievement and ADHD symptoms, and particularly between inattention and academic achievement, is of great importance not only due to the favorable outcomes related to academic achievement itself but also because diagnostic criteria (APA, 2015) require ADHD symptoms to cause impairment, for instance in academic functioning (DuPaul et al., 2014; Gordon et al., 2006). As described above, a differential perspective reveals that not all symptom components of ADHD contribute equally to the achievement gap. Impulsive behavior, when seen as cognitive engaged behavior, has been discussed as having a beneficial effect on academic achievement, and hyperactivity no effect at all (Tymms & Merrell, 2011).

In sum, low academic achievement is a pivotal impairment that comes along with ADHD symptoms. This finding relies mainly on trait ADHD symptoms. Behavioral observation can provide information on symptom severity during the development process of academic achievement. Observation adds the perspective of state ADHD symptoms. Thus, behavioral observation can provide a more direct link between symptom severity and the achievement process.

Association between observed and reported ADHD symptoms

The gold standard for diagnosing ADHD symptoms includes first of all a report of symptoms obtained via questionnaires filled out by multiple informants and an in-depth clinical exploration (Bundesärztekammer, 2005). For validation, observed ADHD symptoms have to be related to measures of reported ADHD symptoms. Which associations are to be expected between reported and observed ADHD symptoms? Studies have shown that children's observed behavior during tests is typically associated to a medium degree with test scores and parental reports (Glutting, Youngstrom, & Watkins, 1996; Gordon, DiNiro, Mettelman, & Tallmadge, 1989; McConaughy et al., 2010). Correlation coefficients between test sessions or classroom observations and reported ADHD symptoms range between r = .07–.39 (McConaughy et al., 2010). This evidence implies that behavioral observation does not replace information from self- and parental reports but should be seen as a complement in the assessment process. In line with that interpretation, a multitrait-multimethod study

showed that classroom and test observation explained unique variance in parent- and teacher-reported hyperactivity and impulsivity in 6- to 12-year-old children (McConaughy et al., 2010). In sum, reported and observed ADHD symptoms seem to measure related but different aspects of ADHD symptoms. Thus, reported and observed ADHD symptoms should be associated to a low to medium degree

3.3.2 Present Study

The overarching research aim of the present study is to investigate the utility and feasibility of a behavioral observation protocol for ADHD symptoms in a simulated classroom situation. To achieve that goal, schoolchildren were observed during a math test and a competitive card game by way of individual video recordings. The first expectation is that inattention, hyperactivity and impulsivity can be observed reliably during the simulated classroom situation. The second expectation is that reported and observed ADHD symptoms are associated to a low to medium degree (cf. Cohen, 1988). The third expectation is that children with fewer ADHD symptoms (reported and observed) show significantly better performance in a math test.

3.3.3 Method

Participants

Thirty-eight children and one parent per child took part in the present study. Video recordings of three children were not available; therefore, 35 children (42.86% male, M_{age} = 10.67 years, SD_{age} = 1.36 years), 30 mothers, and five fathers were included in the analyses. Eight children had one sibling participating in the study; thus, four sibling pairs took part in total. For siblings, the same parent reported ADHD symptoms for both children.

The majority of the participants (54.29 %) attended elementary school (German: Grundschule), 8.57 % a vocational track secondary school (German: Haupt- und Realschule), and 37.14 % an academic track secondary school (German: Gymnasium). Grade levels were distributed as follows: 3rd grade: 20 %; 4th grade: 34.29 %; 5th grade: 22.86 %; 6th grade: 8.57 %; 7th grade: 8.57 % and 8th grade: 5.71 %. Six parents reported that their child had received an ADHD diagnosis in the past. The exact wording of the positive answers to the open-ended question about previous diagnoses

were as follows: “ADHD” (3x); “attention deficit disorder” (1x) and “ADHD combined with tick disorder” (1x); “currently in the diagnostic process for ADHD” (1x). Three children were medicated in the context of ADHD treatment. In an open-ended response format, four parents indicated that their child had been diagnosed with a psychiatric disorder other than ADHD: “emotional disorder” (1x); “in psychotherapeutic treatment” (1x); “specific reading disorder” (1x); “highly gifted” (1x). One participant was one minute late for the math test, and therefore was excluded from analyses that involved performance in the math test (N = 34). The study was approved by the local ethics committee.

Procedure

Recruitment took place via local schools, local child and adolescent psychotherapy practitioners, and circular e-mails to university affiliates in a German university city. Information on the study was sent to interested parents. After parents and children gave their written informed consent to take part in the study, appointments for one telephone interview with the child (10 min) and one telephone interview with the parent (30 min) were made. Apart from the data used in the present study, the parent telephone interview contained a speech sample (5 min), questions on the child’s impairment in daily life due to symptoms, and an assessment of activity using actigraphs.

For the simulated classroom situation, appointments were made with five to seven children at a time. Parents were not involved in the simulated classroom situation. All groups were conducted by the same experimenter. Each group contained one of the children whose parents had indicated an ADHD diagnosis. Siblings did not take part in the same group. Children received a 7 € book voucher as a reward for participation.

Reported ADHD symptoms

ADHD symptoms were assessed via both self-report and parental report using two different instruments during the telephone interview. Both parents and children answered the (1) DSM ADHD, an adaption of the German translation of the criteria for Attention-Deficit/Hyperactivity in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5, APA, 2015) and the (2) ADHD Index, a screening instrument based on the Conners 3 rating scales (Lidzba et al., 2013).

ADHD DSM

In the instructions for the ADHD DSM, participants were asked to evaluate the frequency of each statement read aloud by the interviewer in the last six months. The

response format was a 4-point Likert scale ranging from 0 (*not at all*) to 3 (*very much*). Each criterion was adapted for parental and self-report. The criterion "often fails to give close attention to details" (APA, 2015, p. 77) for instance, was changed to "your child often fails to give close attention to details" (parental report) and "I often fail to give close attention to details" (self-report). All items were adapted in an analogous manner. Nine items referred to inattention, five to hyperactivity and four to impulsivity (18 items in total). Mean scores of the three subscales (DSM inattention, DSM hyperactivity, DSM impulsivity) as well as a total score (DSM ADHD) were used in analyses.

ADHD index

Parental and self-report versions of the ADHD index of the German version of the Conners 3 rating scales was also used as a supplement, because the DSM-5 criteria are not typically used to assess symptom severity (Lidzba et al., 2013). The ADHD index contains 10 items that differentiated best between children with and without an ADHD diagnosis. Items on the ADHD index refer to the child's behavior in the last month. Responses were given on a 4-point Likert scale ranging from 0 (*not at all*) to 3 (*very much*). The validity and reliability of the ADHD index has been proven (Lidzba et al., 2013).

To check the internal consistency, Cronbach's α was calculated for all scales (see Table 13). Internal consistency was sufficient for all scales (α = .72 - .94) except the self-reported DSM hyperactivity scale (α = .64). Thus, the self-reported DSM hyperactivity scale was excluded from further analyses.

Table 13
Descriptive Statistics and Internal Consistency of Scales on ADHD Symptoms

	Parental report			Self-report		
	M(*SD*)	$M_t(SD_t)$	α	*M*(*SD*)	$M_t(SD_t)$	α
DSM ADHD	0.99 (0.65)	.33 (.22)	.94	0.94 (0.49)	.31 (.16)	.87
DSM Inattention	1.18 (0.75)	.39 (.25)	.92	1.05 (0.49)	.35 (.16)	.73
DSM Hyperactivity	0.63 (0.58)	.21 (.19)	.79	0.68 (0.52)	.23 (.17)	.64
DSM Impulsivity	1.00 (0.77)	.33 (.26)	.81	1.02 (0.69)	.34 (.23)	.72
ADHD Index	1.04 (0.72)	.35 (.24)	.93	0.88 (0.62)	.29 (.21)	.87

Note. Response format: 4-point Likert scale 0 (not at all) to 3 (very much);
N = 35; α - Cronbach's alpha internal consistency
M_t – mean transformed to a scale from 0 to 1; SD_t – standard deviation transformed to a scale from 0 to 1

Simulated classroom situation

All behavioral observations were conducted using a standardized behavioral observation protocol, rater training, and an investigation of inter-rater reliability (Furr & Funder, 2007; Steiner et al., 2013). The following criteria were used to create the setting for the simulated classroom situation.

Criteria for the simulated classroom situation. (1) Children should be observed in small groups of 5 to 8 children. (2) The observation procedure should provide a video-recorded frontal whole body view of each seated child. (3) The simulated classroom situation should resemble a common classroom situation. (4) During the simulated classroom situation, one aspect of academic achievement should be assessed. (5) A situational setting should be selected that demands children's attention and requires them to sit still and not be impulsive. The setting was piloted twice, once with young adults and once with children.

Procedure of the simulated classroom situation. The following simulated classroom setting meeting all the aforementioned criteria was created: Five to seven children were invited to a university conference room with tables, chairs and a white

board arranged in a manner comparable to a classroom, with tables in rows. Separate video cameras recorded each child. Hero3 white edition mini-cameras (5cm*3cm*2cm) from the brand GoPro[3] were attached to each table leg using an adjustable goose-neck holder. The film's angle captured the seated child's entire body, with the table board in the middle. All cameras were turned off and on via a remote control. One structured and one more unstructured situation were selected because previous studies using standardized observations have found differential effects of structured and unstructured situations on ADHD symptoms (Imeraj et al., 2016; Jacob et al., 1978; Roberts, 1990). For the structured situation, a 15-min *math test* completely individually and in silence and was chosen. For the unstructured situation, a *competitive card game* comparable to dominoes (Ubongo; Rejchtman, 2011) lasting 9 to 13min was played.

During that *competitive card game*, each child received nine playing cards in each of six rounds. The task in each round was to place seven of their nine cards next to each other, with the rule that exactly two identical symbols form the juncture between two cards. The number of symbols on the cards increased with each round and made the task more difficult. The game was competitive with respect to time. The child who succeeded first in placing the cards together and shouting a signal word won the round.

The *performance in the math test* was used as an operationalization and approximation for academic achievement. An adaption of a standardized test (concentration and performance test; KLT-R Düker & Lienert, 2001) was used in the present study. In its original form, the test consists of mathematical problems with three one-digit numbers (addition and subtraction, example item: 8-2+5 = ?). The task is to calculate the sum of the solutions to two problems, taking certain rules into account. This task poses high cognitive demands on participants. The math test in the present study should impose more repetitive and medium cognitive demands on the children. Moreover, the test should be solvable for children independently of their grade level and type of school they attend. Therefore, the task was adapted in the following way: Children were asked to solve math problems with three one-digit numbers (addition and subtraction) in a row. The solution was never negative and between 1 and 20. Three hundred forty-four math problems in eight rows were presented.

[3] https://de.gopro.com/ (retrieved March 3rd, 2017)

The *math test* and the *competitive card game* demanded that participants sit still. However, the *math test* and the *competitive card game* demanded attention and the control of impulsive behavior to different degrees. While the *math test* emphasized sustained attention to the test, the *competitive* element in the card game probably triggered more impulsive behavior. The math test requires participants to focus on the text for 15 min without distraction, whereas the competitive card game requires participants to refrain from shouting a signal word until they have placed seven cards next to each other. The video recordings lasted 15 min for the *math test* and 9 – 13 min for the *competitive card game.* The complete simulated classroom situation lasted 50 – 60 min. A schedule with the sequence is depicted in Table 14. At the end of each simulated classroom situation, children filled out an 18-item questionnaire on ADHD symptoms during the situation. Most of the children had comprehension problems and asked what several items meant while they were filling out the questionnaire. Moreover, some children had difficulties with independent reading and the experimenter had to read the items aloud to the children. Because of these problems, the questionnaire is not part of the analyses.

Table 14
Schedule of the Simulated Classroom Situation

	Content	Duration (min)
Welcome	The experimenter and every participant introduce themselves (name, grade in school, hobby) Explanation of group rules by experimenter We let the others finish what they want to say We raise our hand if we want to say something We listen to the experimenter	10
Instructions	Instructions are read out to the group. Solutions to examples are discussed together. Time for questions.	5
Math test	*Video recording*	15
Instructions	Instructions are read out to the group and examples are discussed together.	
Competitive card game	*Video recording*	9-13
Questionnaire ADHD symptoms	Explanation and completion of questionnaire (18 items)	10
Farewell	Distribution of book vouchers as reward	5

Video rating procedure

Rating scales for inattention, hyperactivity and impulsivity. The criteria for inattention, hyperactivity and impulsivity in the DSM-5 (APA, 2015) and the items from the Conners 3 scales (Lidzba et al., 2013) served as a basis for item development in the video rating. The following guidelines were used to transform the criteria into rating scales: (1) The video rating items have to be assessable without reference to a certain event. (2) Moreover, video rating items have to be assessable without temporal dependence. (3) The video rating items have to be assessable without information beyond the visuals and audio from the video recording. (4) The video rating items have to be dichotomous (occurred/did not occur). (5) The video rating items have to apply to both the math test and the competitive card game situation. Each DSM-5 criterion was evaluated in light of the listed criteria. Two video rating items each for inattention, hyperactivity and impulsivity were created after piloting. Table 15 shows video rating items, a short explanation of coding criteria and corresponding trait items.

Table 15
Video Rating Items, Coding Criteria and Corresponding Trait Items

Video rating item	Coding criteria	Corresponding trait item
Inattention	"Occurred"	
Does not pay attention to task/game.	Posture and viewing direction not directed to task/game for more than 30 sec.	DSM-5 Often fails to give close attention to details or makes careless mistakes in schoolwork, at work, or during other activities (e.g., overlooks or misses details, work is inaccurate).
Is easily distracted by extraneous stimuli	Distracted by a visual or auditory stimulus.	DSM-5 Is often easily distracted by extraneous stimuli (for older adolescents and adults, may include unrelated thoughts).
Hyperactivity		
Often fidgets with or taps hands or feet or squirms in seat.	Fast, repetitive movements of external extremities	DSM-5 Often fidgets with or taps hands or feet or squirms in seat.
Is constantly on the move.	Constant enduring movements of whole body (not external extremities)	Conners 3 Is constantly on the move.
Impulsivity		
Makes noises/starts interaction intentionally	Interaction or noise that exceeds unintentional noise or interactions required to do the tasks	DSM-5 Often talks excessively.
Has difficulties pulling him/herself together	Problems withholding an impulse. Signs: immediate noise, complaining, not following the rules or taking part very unwillingly.	DSM-5 Often interrupts or intrudes on others (e.g., butts into conversations, games, or activities; may start using other people's things without asking or receiving permission; for adolescents and adults, may intrude into or take over what others are doing).

Note. DSM-5 (APA, 2015); Conners 3 scales (Lidzba et al., 2013)

Rating procedure. Each video was cut into 20 sec clips. After excluding the last clips, 45 or 46 clips for the math test and about 30 clips for the competitive card game (depending on the length of the game) existed for each child. Each item was evaluated once for every clip. Paper-and-pencil questionnaires were used for the video rating items. The material was rated according to the following pattern: First, six children were randomly chosen from the sample of 35 children and rated by three trained raters (Cycle 1). Second, the remaining video material from 29 children was distributed among the three raters (main cycle, two raters rated ten children, and one rater rated nine children). Third, every rater additionally rated four children, i.e. two children that each of the other two raters had rated before. This meant that six additional children were rated by all three raters (Cycle 2). The three raters were undergraduate psychology students. Table 16 gives an overview of the rating cycles.

Table 16
Overview of Rating Cycles

	Rater 1	Rater 2	Rater 3	Children with multiple evaluations	Total number of children rated
Cycle 1		6		6	6
Main cycle	9	10	10		29
					0
Cycle 2	4	4	4	6	

Note. Number of children rated per rater and cycle.

Rater training. The training took place in two sessions. In the first session before the rating, the raters received written descriptions of the rating procedure as well as examples for each item and for the math test and the competitive card game. They were taught when items "occurred" and "did not occur". Video examples from the two pilot situations for "occurred" and "did not occur" were presented and discussed. Thereafter, all raters rated the video material for Cycle 1. After Cycle 1, a second training session took place. The inter-rater reliability coefficients for Cycle 1 were computed and problems were discussed. This discussion resulted in slight adaptations to the description of the following items: For the item "Does not pay attention to task/game," a definition of being attentive during each phase of the competitive card game (playing, checking the winner of the round, instructions, and waiting) was added. A statement was added to the item "Is easily distracted by extraneous stimuli" that the extraneous stimulus must be visible or audible for the rater. For the item "is constantly on the

move," a statement was added that the movement has to be enduring and not limited to fine motor movements. For the item "has difficulties pulling him/herself together," a reference to displeasure was deleted and a concrete description of withholding an immediate impulse was stressed. Thereafter, the remaining video material was rated by the three raters (main cycle and Cycle 2).

Aggregation of ratings. Due to the described rating procedure, ratings from three raters were available for the video material of six children in Cycle 1 and six children in Cycle 2. For the remaining children (n = 23), one rating from one rater was available. For the calculation of inter-rater reliability coefficients, the ratings from Cycles 1 and 2 were aggregated to form mean values of all the ratings of each item for each child. This aggregation was done once for each child's complete video material, once for the math test and once for the competitive card game. For the calculation of association analyses, a random rating was selected from the cases in which three ratings for one item existed (Cycles 1 and 2). After that selection, the ratings were again aggregated into one value per child for each item, once for the complete video material, once for the math test, and once for the competitive card game. The result of the aggregation was an interval scaled variable (range 0-1) for each video rating item.

Inter-rater reliability coefficients. Two reliability coefficients, Krippendorff's α (Hayes & Krippendorff, 2007; Krippendorff, 1970) and the ICC (Hallgreen, 2012; Shrout & Fleiss, 1979), were selected. The calculation of two different coefficients was advisable in this case due to the wide variety of coefficients (Hayes & Krippendorff, 2007). Data entry took place twice, with an agreement of above 99.9%. The Kirppendorff's α coefficient is the ratio between the disagreement observed within each compared unit and the disagreement expected by chance based on the responses given in all units (Honour, 2016; Krippendorff, 1970). Krippendorff's α is flexible with regard to the level of measurement (nominal, ordinal or interval). In the version for interval data used in the present calculation, reliability is defined as absolute agreement between the ratings, meaning that systematic variance between raters is taken into account. Krippendorff's α calculates the reliability of the ratings of a random rater. Thus, it should be applied in cases when subsequent analyses are done using one assessment per unit (Hallgreen, 2012; Krippendorff, 1970). The ICC can be defined broadly as the ratio of the variance of interest over the sum of the variance of interest plus error (Shrout & Fleiss, 1979). Many different versions exist. The model, type and unit can be defined. In the present study, the model was set to "two-way" because the same set of raters rated

the same material, the type was set to "agreement" to account for the systematic variance between the raters, and unit was set to "single" to calculate the reliability of one random rater. (Hallgreen, 2012; Shrout & Fleiss, 1979). Krippendorff's α and ICC_{single} were calculated with the functions "icc" and "kripp.alpha," respectively, from the "irr" package (Gamer et al., 2012) using the statistical software R (R Development Core Team, 2012). The values of Krippendorff's α and ICC_{single} range from .00 (no agreement) to 1.00 (absolute agreement). Values above .60 for ICC_{single} are considered acceptable (Hallgreen, 2012). Values above .70 for Krippendorff's α are considered acceptable (Hayes & Krippendorff, 2007).

3.3.4 Results

Inter-rater reliability

The first expectation was that inattention, hyperactivity and impulsivity can be observed reliably during the simulated classroom situation. The design of the present study incorporated two rating cycles (Cycle 1 and Cycle 2). Inter-rater reliability coefficients were calculated separately for each cycle, because the rating manual was slightly adapted after Cycle 1, as described above. Separate coefficients were calculated for the entire video material, the math test, and the competitive card game. The reliability criterion for the selection of an item for further analyses was that one of the two inter-rater reliability coefficients reached an acceptable value in both rating cycles. All inter-rater reliability coefficients are depicted in Table 17. According to this criterion for the complete video material, the following items did not reach sufficient reliability: The item "is constantly on the move" had to be excluded from all analyses. The impulsivity items had to be excluded from the analyses relating to the math test. The inattention items had to be excluded from the analyses relating to the competitive card game.

Table 17
Inter-Rater Reliability of Video Rating Items

	Complete video material				Math test				Competitive card game			
	Cycle 1		Cycle 2		Cycle 1		Cycle 2		Cycle 1		Cycle 2	
	α_{Kripp}	ICC_{single}	α_{Kripp}	ICC_{single}	α_{Kripp}	ICC_{single}	α_{Kripp}	ICC_{single}	α_{Kripp}	ICC_{single}	α_{Kripp}	ICC_{single}
Inattention	.69	**.74**	**.77**	**.81**	.66	**.71**	**.90**	**.92**	.20	.27	.17	.34
Does not pay attention to task/game.												
Is easily distracted by extraneous stimuli.	**.77**	**.81**	**.80**	**.83**	**.82**	**.84**	**.77**	**.80**	.44	.50	**.73**	**.77**
Hyperactivity	.59	**.65**	**.97**	**.97**	.59	**.65**	**.95**	**.95**	.63	**.69**	**.90**	**.91**
Often fidgets with or taps hands or feet or squirms in seat.												
Is constantly on the move.	.43	.53	.20	.31	.59	**.65**	.22	.32	.21	.32	-.00	.14
Impulsivity	**.85**	**.87**	**.89**	**.90**	1	-	.31	.36	**.83**	**.85**	**.99**	**.99**
Makes noises/starts interaction intentionally.												
Has difficulties pulling him/herself together.	**.78**	**.81**	.61	**.66**	1	-	.29	.35	**.79**	**.82**	.68	**.73**

Note. α_{Kripp} – Krippendorff's α; ICC_{single} – Intraclass Correlation Coefficient; Acceptable values are in bold. 3 raters rated 6 children in each cycle; aggregated to one value per child per item; ICC_{single} for impulsivity items in Cycle 1 could not be computed because impulsivity did not occur at all.

Association between observed and reported ADHD symptoms

Descriptive statistics for the parental and self-reported ADHD symptom scales are displayed in Table 13 and descriptive statistics for video rating items (observed ADHD symptoms) are depicted in Table 18. In order to increase the comparability of observed and reported ADHD symptoms, descriptive statistics for reported ADHD symptoms are reported once on the original scale (range 0-3) and once transformed to the same scale (range 0-1) as the observed ADHD symptoms.

Table 18
Descriptive Statistics of Video Rating Items

	Complete video material	Math test	Competitive card game
	M(*SD*)	*M(SD)*	*M(SD)*
Inattention:			
Does not pay attention to task/game.	.32 (.16)	.46 (.27)	
Is easily distracted by extraneous stimuli.	.14 (.15)	.16 (.19)	
Hyperactivity			
Often fidgets with or taps hands or feet or squirms in seat.	.12 (.14)	.15 (.17)	.09 (.13)
Impulsivity			
Makes noises/starts interaction intentionally	.09 (.10)		.17 (.19)
Has difficulties pulling his/herself together	.05 (.07)		.08 (.12)

Note. Items range between 0 and 1, Aggregated to one value per participant (N = 35)

The second expectation was that reported and observed ADHD symptoms are associated to a low to medium degree (cf. Cohen, 1988). A visual check of whether the video rating items were normally distributed was performed using quantile-quantile plots created by the "qqnorm" function from the "stats" package implemented in the statistical software R (R Core Team and contributors worldwide, n.d.). A normal distribution could not be assumed for any of the video rating items. Therefore, the rank correlation coefficient Spearman's ρ was selected for association analyses. Tables 19, 20 and 21 depict correlation analyses between video rating items and parental and self-reports for the complete video material, the math test and the competitive card game, respectively.

For the complete video material, analyses show significant associations between video rating items and reported ADHD items, particularly for hyperactivity and impulsivity video rating items but not with regard to inattention video rating items.

Comparing the associations between self-reports and parental reports, respectively, to the video rating items shows few deviations. In general, the associations between self-reported ADHD symptoms and video rating items were smaller.

For the math test situation, analyses showed significant associations between the video rating inattention item “Is easily distracted by extraneous stimuli” and the DSM ADHD total score as well as reported impulsivity items. The video rating hyperactivity item “Often fidgets with or taps hands or feet or squirms in seat” was associated with the DSM ADHD total score. Interestingly, the video rating item “Does not pay attention” was not significantly associated with any of the reported ADHD symptoms. Here again, self-reported items are associated to video rating items to a similar extent and in the same direction like ADHD symptoms in parental report, but associations are in most cases little smaller.

For the competitive card game situation, pronounced significant associations between reported ADHD symptoms and impulsivity video rating items were found. This holds true for the associations between impulsivity video rating items and reported inattention, impulsivity and the DSM ADHD total score. No significant correlations could be found between the hyperactivity video rating item and reported ADHD symptoms. Associations involving self-reports and parental reports were comparable, with the associations involving self-reports a little smaller in most cases. Importantly, the association between reported and observed impulsivity was the only association where the correlation coefficient exceeded .5

Table 19
Complete Video Material (Math Test and Game) Association Between ADHD Symptoms in Video Rating and Reported ADHD Symptoms (Spearman's rank correlation coefficient)

	DSM ADHD		DSM Inattention		DSM Hyperactivity	DSM Impulsivity		ADHD Index	
Video rating items	Self-report	Parental report	Self-report	Parental report	Parental report	Self-report	Parental report	Self-report	Parental report
Inattention									
Does not pay attention to task/game.	.31	.29	.27	.26	.21	.28	.24	.20	.21
Is easily distracted by extraneous stimuli	.17	.19	.11	.17	-.01	.28	.24	.01	.13
Hyperactivity:									
Often fidgets with or taps hands or feet or squirms in seat.	.21	.38 *	.22	.33 *	.40 *	.12	.31	.10	.34 *
Impulsivity									
Makes noises/starts interaction intentionally	.40 *	.38 *	.24	.39 *	.14	.51 **	.43 **	.26	.32
Has difficulties pulling him/herself together	0.35 *	.46 **	.15	.45 **	.22	.40 *	.55 **	.15	.36 *

Note. $N = 35$; DSM ADHD – mean score of 18 adapted DSM-5 criteria; DSM Inattention – mean score of adapted DSM-5 inattention criteria; DSM Hyperactivity - mean score of adapted DSM-5 hyperactivity criteria; DSM Impulsivity - mean score of adapted DSM-5 impulsivity criteria; ADHD Index – mean score of ADHD index of Conners 3 rating scales (Lidzba et al., 2013)
* $p < .05$; ** $p < .01$

Table 20
Math Test Association Between ADHD Symptoms in Video Rating and Reported ADHD Symptoms (Spearman's rank correlation coefficient)

	DSM ADHD		DSM Inattention		DSM Hyperactivity	DSM Impulsivity		ADHD Index	
Video rating items	Self-Report	Parental report	Self-Report	Parental report	Parental report	Self-Report	Parental report	Self-Report	Parental report
Inattention:									
Does not pay attention to task/game.	.26	.25	.22	.27	.17	.22	.14	.15	.18
Is easily distracted by extraneous stimuli	.25	.35 *	.18	.31	.16	.40 *	.39 *	.13	.29
Hyperactivity:									
Often fidgets with or taps hands or feet or squirms in seat.	.16	.34 *	.17	.31	.30	.09	.31	.01	.32

Note. $N = 35$; DSM ADHD – mean score of 18 adapted DSM-5 criteria; DSM Inattention – mean score of adapted DSM-5 inattention criteria; DSM Hyperactivity - mean score of adapted DSM-5 hyperactivity criteria; DSM Impulsivity - mean score of adapted DSM-5 impulsivity criteria; ADHD Index – mean score of ADHD index of Conners 3 rating scales (Lidzba et al., 2013)
* $p < .05$; ** $p < .01$

Table 21
Competitive Card Game Association Between ADHD Symptoms in Video Rating and Reported ADHD Symptoms (Spearman's rank correlation coefficient)

	DSM ADHD		DSM Inattention		DSM Hyperactivity	DSM Impulsivity		ADHD Index	
Video rating	Self-Report	Parental report	Self-Report	Parental report	Parental report	Self-Report	Parental report	Self-Report	Parental report
Hyperactivity:									
Often fidgets with or taps hands or feet or squirms in seat.	.17	.31	.17	.29	.28	.11	.26	.22	.24
Impulsivity									
Makes noises/starts interaction intentionally	.43 **	.39 *	.27	.41 *	.13	.52 **	.43 **	.30	.33
Has difficulties pulling him/herself together	.35 *	.48**	.16	.48 **	.22	.36 *	.55 **	.17	.39 *

Note. N = 35; DSM ADHD – mean score of 18 adapted DSM-5 criteria; DSM Inattention – mean score of adapted DSM-5 inattention criteria; DSM Hyperactivity - mean score of adapted DSM-5 hyperactivity criteria; DSM Impulsivity - mean score of adapted DSM-5 impulsivity criteria; ADHD Index – mean score of ADHD index of Conners 3 rating scales (Lidzba et al., 2013)
* $p < .05$; ** $p < .01$

Association between ADHD symptoms and mathematical performance

The third expectation was that children with fewer ADHD symptoms (reported and observed) would show significantly better performance in a math test. Performance was operationalized as the number of correctly solved problems. In order to exclude the variance accounted for by grade level, a semi-partial correlation coefficient (Spearman's ρ) was calculated. To this end, the function "spcor.test" from the "ppcor" package (Seongho, 2015) for the statistical software R (R Development Core Team, 2012) was used. The inattention video rating items were strongly negatively associated with performance ("Does not pay attention to task/game": ρ = -.49 **; "Is easily distracted by extraneous stimuli": ρ = -.50**). The hyperactivity video rating item was not significantly associated with performance (ρ = -.24). Neither did parental reported ADHD symptoms show significant associations with performance (DSM ADHD: ρ = -.22; DSM Inattention: ρ = -.23; DSM Hyperactivity: ρ = -.22; DSM Impulsivity: ρ = -.20, ADHD index: ρ = -.13).

3.3.5 Discussion

The overarching aim of the present study was to investigate the utility and feasibility of an observation protocol for ADHD symptoms in a simulated classroom situation. The results showed that the observation of inattention during the math test and impulsivity during the competitive card game had satisfactory reliability. One hyperactivity video rating item "Often fidgets with or taps hands or feet or squirms in seat" could be observed reliably for both situations and the complete video material. The other hyperactivity item "is constantly on the move" did not reach sufficient inter-rater reliability at all. Association analyses between reported and observed ADHD symptoms showed the most pronounced associations for hyperactivity and impulsivity and partly for inattention during the math test. Importantly, the only correlation coefficient that exceeded .50 was the association between reported and observed impulsivity during the competitive card game. Furthermore, the more problems that were solved in the math test, the less inattention could be observed. In contrast to observed state ADHD symptoms, no reported ADHD symptoms showed significant associations with performance.

Observation of ADHD symptoms in the classroom

The first expectation was that inattention, hyperactivity and impulsivity can be observed reliably during a simulated classroom situation. We sought to investigate the reliability of behavioral evaluations based on an observation protocol for ADHD symptoms in a simulated classroom situation. The observation protocol for ADHD symptoms used in the present study was developed to include the school context as one pivotal context in the assessment of ADHD symptoms. The reliability analyses showed that incorporating one situation requiring sustained attention (math test) and one situation with a competitive component (competitive card game) was important for assessing inattention as well as hyperactivity and impulsivity. A standardized setting was used in the present study to exclude contextual factors and conduct an initial investigation of psychometric properties. The simulated classroom situation used in the present study was a novel situation for the participants. Novel situations are known to diminish the emergence of ADHD symptoms (APA, 2015). This might be the reason why observed impulsivity and hyperactivity were below .20 on a scale from 0 to 1. Occurrences of the two inattention video-rating items differed (M = .32 and M = .14, scale 0 to 1), but were not very high either. To overcome the problem of lack of ecological validity due to the standardized settings, the reliability of the video-rating items should be reinvestigated in natural classrooms.

Association between observed and reported ADHD symptoms

The second expectation was that reported and observed ADHD symptoms would be associated to a low to medium degree (cf. Cohen, 1988). The correlation coefficients between observed and reported ADHD symptoms ranged between .10 and .50 and can be considered small to large associations (Cohen, 1988). The basis upon which parents evaluated their child's ADHD symptoms in a school context was somewhat unclear, because parents usually do not experience children in that context. On the one hand, parents might have inferred from the behavior they tended to witness at home. On the other hand, parents might have relied on feedback they had received about their child's behavior at school. One previous study found that parental reported ADHD symptoms at school were more highly correlated with their own reports of their child's behavior at home than with teacher reported ADHD symptoms at school (Mitsis et al., 2000). Some associations involving observed impulsivity had large coefficients. One reason for this might be that impulsive-disruptive behavior is much more salient than inattentive behavior, resulting in more equal ratings across settings. As expected, the remaining associations were small to medium.

ADHD symptoms and academic achievement

The third expectation was that children with fewer ADHD symptoms (reported and observed) would perform significantly better on a math test. A math test was integrated into the present observation protocol to gain insights on ADHD symptoms during the development process of academic performance. The results showed that observed inattention is strongly negatively associated to math performance. This finding is in line with previous studies, which stress the importance of inattention for lower academic achievement (Gaub & Carlson, 1997; Merrell & Tymms, 2001; Tosto et al., 2015). In this study, reported (parental) ADHD symptoms were not significantly associated with performance. It is not surprising that state inattention (observed) is more strongly associated with performance than trait inattention (reported). The differences in the associations between reported and observed inattention implies that the observational approach taken in the present study contributes important information to the investigation of academic achievement and ADHD symptoms.

Strengths and limitations

The generalizability of the results of the present study is limited. First, the sample size did not allow for multivariate analysis and a group comparison of children with and without ADHD. Second, a teacher report on ADHD symptoms would have been valuable, because one important goal of the present study was to include the school context in the assessment of ADHD symptoms. Teachers usually witness children in the classroom environment; thus, their ratings of ADHD symptoms are more likely to be based on the school environment than parents. The associations between observed ADHD symptoms and teacher reported ADHD symptoms are especially important. Third, the groups in the simulated classroom situation were not matched according to age and grade level, and the difficulty level of the math test was not adjusted to grade level. The composition of the groups probably influenced the children's behavior. Fourth, we used a simulated classroom situation to keep the context constant, thus increasing the comparability between observations. However, the situation was highly artificial and new to the children. This limits the utility of the protocol to a substantial extent. In order to maintain the advantage of standardized situations, observation should be transferred to similar situations in natural classrooms, such as "quiet work on a test." However, these limitations notwithstanding, the study also has several strengths. First, the training manual for the observation of ADHD symptoms is short and rater training lasts less than three hours. Thus, the time required

to apply the rating system developed here is reasonable, which could help facilitate its transfer to more natural observation settings and to practice. Second, the video recording set-up used in the present study delivers frontal whole-body pictures of the seated child. The basis for evaluation is far better than in participatory observation. Third, the assessment of inter-rater reliability in two cycles and a review process of the training materials improved the quality of the rating system and should facilitate its transfer to natural settings.

Implications

The present study has three major implications for further research. First, future studies should investigate the question of whether behavioral observation in classrooms adds incremental validity to the assessment of ADHD symptom severity. Studies on the incremental validity of an assessment method usually ask whether a certain questionnaire explains the existence of an ADHD diagnosis over and above another questionnaire. This approach bears the danger of circularity, because an ADHD diagnosis is made on the basis of questionnaire ratings, whereas behavioral observations have the potential to add additional information. Second, the inclusion of measures of academic achievement in behavioral observation probably ameliorated the assessment of impairment through ADHD symptoms (D-criterion) (APA, 2015; Gordon, 2006). Third, to differentiate the effect of an ADHD intervention from placebo effects, the raters of ADHD symptoms have to be blinded to whether the concerned child took part in an intervention. For practical reasons, this is very difficult to achieve because parents and teachers as raters of ADHD symptoms are usually involved in the study. Meta analyses of neurofeedback (Cortese et al., 2016), cognitive training (Cortese et al., 2015) and nonpharmacological treatments in general (Sonuga-Barke et al., 2013) reveal strongly diminished effects when information on ADHD symptoms from probably blinded raters are used. Behavioral observation in meaningful settings is an objective, hence blinded, assessment instrument. Thus, the observational approach taken in the present study bears great potential to assess the effectiveness of interventions.

3.3.6 Conclusion

The present study investigated the utility and feasibility of an observation protocol for ADHD symptoms in a simulated classroom situation. Results revealed that the developed rating system assesses ADHD symptoms reliably, with only reasonable

effort involved. Therefore, it can be an important tool for assessing ADHD symptoms in classroom settings and should be developed further for implementation in practice

4 GENERAL DISCUSSION

Behavioral descriptions of inattentive, hyperactive and impulsive behavior are the anchor point for the identification of ADHD symptoms. The overarching research aim of the present dissertation was to investigate the potential of observation to assess ADHD symptoms in a behavior-based, objective and context-dependent manner. To this end, three empirical studies have been conducted. The following section summarizes the results of these studies.

The aim of *Study 1* was to investigate the utility of a FMSS of parents for the assessment of family contexts relevant for school-aged children with ADHD symptoms. Five-minute audio recordings of parents' talk about their child and their relationship to their child were analyzed according to the subscales of Expressed Emotions: (Magana et al., 1986): valence of initial statement, warmth, relationship, critical comments and positive comments. The analysis revealed that reliability was sufficient for the relationship, critical comments, and positive comments subscales. A more negative initial statement was associated with more ADHD symptoms in the parental report. Moreover, more ADHD symptoms in parental and self-reports were significantly associated with more critical and less positive comments.

The aim of *Study 2* was to investigate the influence of attention orientation, observed activity and impulsivity, and an ADHD diagnosis on performance in a delay of gratification task in school-aged children with and without an ADHD diagnosis. Children who decide to wait for a bigger delayed reward rather than take a smaller immediate reward (Mischel, 1996) have been studied in terms of whether they succeed in waiting for the bigger reward and how. Video recordings taken during a delay of gratification waiting task were analyzed in terms of activity, impulsivity, stimulus-driven attention orientation (towards the reward), and goal-driven attention orientation (towards an object other than the reward). In line with the hypotheses, less stimulus-driven attention orientation and more goal-driven attention orientation predicted waiting in the delay of gratification task. Contrary to expectations, an ADHD diagnosis and impulsive behavior during the delay of gratification waiting task were not significant predictors of waiting. However, less activity during the delay of gratification increased the probability of waiting.

Study 3 investigated the utility and feasibility of a behavioral observation protocol for ADHD symptoms in a simulated classroom situation. Video recordings of

children during a math test and a competitive card game were analyzed with regard to inattention, hyperactivity and impulsivity. Observation of inattention was reliable for the math test only. Impulsivity could be reliably observed during the competitive card game only. The item "Often fidgets with or taps hands or feet or squirms in seat" could be observed reliably in both situations. As expected, reported and observed ADHD symptoms had low to medium associations with one another in general. The only correlation coefficient that exceeded .50 was the association between reported and observed impulsivity during the competitive card game. Moreover, the more observed inattention a child exhibited, the fewer problems he or she solved in the math test.

4.1 Observation: behavior-based, objective and context-dependent?

The aim of the present dissertation was to investigate the potential of observation to assess ADHD symptoms in a behavior-based, objectively and context-dependent manner. The following section reviews the results of the three empirical studies in terms of the (1) proximity to behavior, (2) objectivity and (3) context-dependence of observation as an assessment method.

Study 1 applied observation techniques to an audio-recorded speech sample of a parent taking about their child and their relationship towards their child. Therefore, the target of observation was the family context rather than ADHD symptoms. In this case, the question is: (1) How proximal to relevant behavior in families is the information obtained from the FMSS? Advocates of assessing actual behavior, rather reported behavior or behavior inferred from cognitive tasks, mention the analysis of natural language as an important sources of information about an informant's social attitudes (Baumeister et al., 2007; Pennebaker, Mehl, & Niederhoffer, 2003). The proximity of the information obtained to concrete behavior in the family context cannot be assessed because the study did not include a second measure of family context. However, it can be assumed that the analysis of natural language is more behavior-based than reported family context. One additional reason might be that rating scales pre-specify answers, while speech samples are more open. (2) How objective is the obtained information? In the FMSS, the parent has an influence on what he or she says. However, the assessment is less explicit than a questionnaire item, and thus less influenceable. Moreover, only three subscales (relationship, critical comments, and positive comments) reached

sufficient inter-rater reliability after about four hours of rater training. Thus, the current conceptualization of EE in the FMSS is not developed enough to deliver robust, objective, and reliable values on the family context. (3) How context-dependent is the information obtained from the FMSS? The FMSS provides information about the family context and not information about ADHD symptoms in a certain context. However, other studies have proven that the FMSS is a suitable measure for assessing the family context, which is relevant for children with ADHD symptoms over time (Musser et al., 2016).

In *Study 2,* children's activity, impulsivity, and attention orientation during a delay of gratification task was observed. (1) How behavior-based was the observation, and how proximal to behavioral descriptions of ADHD symptoms? Table 22 juxtaposes activity and impulsivity video rating items with corresponding DSM-5 criteria. Except for one activity item, the video rating activity items match the DSM-5 criteria well. The impulsivity video rating items are more abstract and do not capture impulsivity in a concrete manner as described in the DSM-5 criteria. To observe impulsivity as described in the behavioral descriptions, the observation setting should draw inspiration from situations described in the criteria, such as waiting in a line or at a restaurant. However, the observational coding in *Study 2* was proximal to the diagnostic criteria.

Table 22
ADHD Video Rating Items in the Delay of Gratification Task and Corresponding DSM-5 Criteria

	Video rating item	Corresponding DSM-5 criterion
Hyperactivity	Rocks the trunk	-
	Fidgets with hands or feet	Often fidgets with or taps hands or feet or squirms in seat.
	Squirms in the chair	Often fidgets with or taps hands or feet or squirms in seat.
	Gets up from the chair	Often leaves seat in situations when remaining seated is expected (e.g., leaves his or her place in the classroom, in the office or other workplace, or in other situations that require remaining in place).
	Climbing around	Often runs about or climbs in situations where it is inappropriate. (Note: In adolescents or adults, may be limited to feeling restless.)
Impulsivity	Seems to be	Is often "on the go," acting as if "driven by a

restless and uneasy	motor" (e.g., is unable to be or uncomfortable being still for extended time, as in restaurants, meetings; may be experienced by others as being restless or difficult to keep up with).
Seems to have difficulties pulling him/herself together	Often interrupts or intrudes on others (e.g., butts into conversations, games, or activities; may start using other people's things without asking or receiving permission; for adolescents and adults, may intrude into or take over what others are doing).

Note. Activity and impulsivity video rating items from *Study 2* and corresponding DSM-5 criteria (APA, 2015)

(2) How objective was the observation of ADHD symptoms? Four (15 %) or three (85%) evaluation were available for every observational code. Thus, the inter-rater reliability can be considered robust and the influence of the observer on the ratings negligible. This degree of agreement could be reached following a rater training of about five hours. However, the inter-rater reliability coefficient used in *Study 2* provides evidence for the reliability of observations using the mean values obtained by three or four observers. In practice, ratings from only one observer are used. For this reason, the reliability of the items applied here has to be reexamined for use in practice (Shrout & Fleiss, 1979). (3) How valuable is the information on context obtained in the observation? Due to the theoretically assumed delay-aversive motivational style (Sonuga-Barke, 2003) of children with ADHD symptoms, a delay situation should particularly provoke ADHD symptoms. Thus, a delay of gratification situation can be considered a critical situation for ADHD. The observed ADHD symptoms depend on a delay context, which is theoretically explained. Moreover, ADHD symptoms observed in the delay context had a different impact on performance in the delay of gratification task than ADHD symptoms as reflected by an ADHD diagnosis. Thus, the context-dependent observed symptoms added valuable information beyond reported ADHD symptoms (Imeraj et al., 2016).

Study 3 observed inattention, hyperactivity and impulsivity among school-aged children in a simulated classroom situation. (1) How behavior-based was the observation, and how proximal to behavioral descriptors of ADHD symptoms? Table 23 matches inattention, hyperactivity and impulsivity video rating items that could be reliably measured with corresponding DSM-5 criteria. With the exception of

impulsivity video rating items, the items align very well with DSM-5 criteria. The impulsivity items were specifically adapted to the classroom situation.

Table 23
ADHD Video Rating Items in the Classroom Situation and Corresponding DSM-5 Criteria

	Video rating item	Corresponding DSM-5 criterion
Inattention	Does not pay attention to task/game.	Often fails to give close attention to details or makes careless mistakes in schoolwork, at work, or during other activities (e.g., overlooks or misses details, work is inaccurate).
	Is easily distracted by extraneous stimuli	Is often easily distracted by extraneous stimuli (for older adolescents and adults, may include unrelated thoughts).
Hyperactivity	Often fidgets with or taps hands or feet or squirms in seat.	Often fidgets with or taps hands or feet or squirms in seat.
Impulsivity	Makes noises/starts interaction intentionally	Often talks excessively.
	Has difficulties pulling him/herself together	Often interrupts or intrudes on others (e.g., butts into conversations, games, or activities; may start using other people's things without asking or receiving permission; for adolescents and adults, may intrude into or take over what others are doing).

Note. Activity and impulsivity video rating items from *Study 3* and corresponding DSM-5 criteria (APA, 2015)

Similarly to *Study 2*, direct observation of impulsive behavior as described in the diagnostic manuals was difficult. Impulsive behavior is often described by way of single events with lower occurrence rates than for instance inattention. Thus, the information on ADHD symptoms obtained via behavioral observation in the classroom setting mapped well onto the behavioral descriptions of ADHD symptoms with the exception of impulsivity. (2) How objective was the observation of ADHD symptoms? To guarantee inter-rater reliability, 34% percent of the material was rated by three observers. The inter-rater reliability coefficients reflect the reliability of the evaluations of a rater who received a comparable training to the raters in the study. Therefore, the reliability is generalizable to one observer in practice. This level of agreement could be

reached following about four hours of rater training. (3) How valuable is the information on context obtained in the observation? In most cases, reported ADHD symptoms were associated with observed ADHD symptoms to a medium degree. Thus, the information obtained through observation probably goes beyond the reported ADHD symptoms. Moreover, observed inattention and performance in the math test showed a significant association, whereas reported ADHD symptoms were not significantly associated with performance. Thus, the context-embedded observed information on ADHD symptoms provided valuable information exceeding that obtained by symptom reports.

In sum, does observation meet the expectation of providing behavior-based, objective and context-dependent information on ADHD symptoms? First, natural occurring language (*Study 1*) or behavior (*Studies 2 and 3*) formed the basis of all the empirical studies. Thus, the proximity to natural occurring behavior was high in all three empirical studies. Second, the objectivity of the obtained information was ensured through the use of blinded raters, a comprehensive rater training, multiple raters (2-4), and the calculation of agreement coefficients. In *Study 2* (delay situation), observations were reliable. In *Study 1* (speech samples) and *Study 3* (classroom situation), inter-rater reliability was not sufficient for all rating items. Here, the rating procedure needs further development. Third, the assessment of ADHD symptoms within meaningful contexts (delay situation in *Study 2*, classroom in *Study 3*) provided information about associations with criterion variables (performance in a delay of gratification task in *Study 2*, performance in math test in *Study 3*) exceeding that available from reported ADHD symptoms. Taken together, observation is a time consuming and costly assessment method but does indeed meet the requirements of providing behavior-based, objective and context-dependent information on ADHD symptoms.

4.2 Limitations of empirical studies

Several factors limit the generalizability of the findings of the empirical studies included in the present dissertation. Looking to the future, the limitations discussed in the following section can be considered starting points for the development of new study designs. Specifically, this section focuses on the validity of observed ADHD symptoms, the failure to compare observation methods, the integration of information sources and the ecological validity of context-dependent assessment.

Validity of observed ADHD symptoms

The goal of the present dissertation was to investigate the potential of observation of ADHD symptoms as a behavior-based, objective and context-dependent assessment instrument. In order to estimate this potential, information from observation needs be related to information from other sources (convergent validity) and to relevant outcomes (criterion validity). If the convergent validity is low to medium, an in-depth investigation of criterion validity should clarify the association of observed ADHD symptoms with meaningful outcomes (for instance academic achievement) in order to justify the application of the assessment method. In *Study 1,* the observed family context, as an outcome variable, was related to reported ADHD symptoms. Hence, the criterion validity of the FMSS as a measure for family context was examined. However, the results of the FMSS have not been related to, for instance, reports on the family context. The convergent validity of the results of the FMSS has not yet been investigated and should be considered in the future. In *Study 2,* observed attention orientation, activity and impulsivity were related to the probability of waiting as a measure of performance in a delay of gratification task. The association between observation and performance in the delay of gratification task verifies the criterion validity of the observed ADHD symptoms. However, a more in-depth examination of the criterion validity is desirable. For instance, the association between observed ADHD symptoms and reported measures of delay aversion and teacher reports on behavior in naturally occurring delay situations in classrooms would deliver important information on the criterion validity of observed ADHD symptoms. Moreover, observed ADHD symptoms have not been related to reported ADHD symptoms directly. This information would have been valuable to assess the convergent validity of observed ADHD symptoms during the delay of gratification task. *Study 3* related observed ADHD symptoms (inattention, hyperactivity, and impulsivity) to reported ADHD symptoms and performance in a math task. The association between observed and reported ADHD symptoms examines the convergent validity of observed ADHD symptoms. The criterion validity was investigated through the association with performance in a math task. However, particularly because the association between observed and reported symptoms was mostly low to medium, a more in-depth investigation of the criterion validity of observed ADHD symptoms would have been valuable. For instance, the association between observed ADHD symptoms and grades,

teacher-reported ADHD symptoms and perceived impairment through symptoms should be considered in future studies.

Comparison of observation methods

As described in the theoretical background, (see 1.2.2) behavioral observation is a very flexible method. The empirical studies of the present dissertation varied the setting of behavioral observation. However, most other parameters stayed the same across all three studies. The level of abstraction of the behavioral codes in both studies that included observation of ADHD symptoms (*Study 2* and *Study 3*) remained on the level of behavioral indicators ("often fidgets with or taps hands or squirms in the chair", see *Study 3*) rather the construct level ("is hyperactive"). Insights from a comparison of behavioral codes at different levels of abstraction could improve the rating system. In addition, raters evaluated the occurrence of a certain behavior in a predefined time frame (*Study 2 and Study 3*) rather than the duration of the behavior of interest. However, duration of behavior is another important feature that can be assessed via behavioral observation and should be considered in future studies (Antrop, 2002). To facilitate rater training, the present dissertation deployed a dichotomous response format for the observation of ADHD symptoms (*Study 2 and Study 3*). It is much easier to train raters to make occurred/did not occur comparisons than to train them to use an intensity scale. Comparing response formats reflecting occurrence (dichotomous) and intensity would be valuable, particularly in light of the fact that ADHD symptoms are exhibited in different intensities (Coghill & Sonuga-Barke, 2012) and a dichotomous categorization is artificial. Moreover, all three empirical studies used recorded material (audio recordings and video recordings); the raters did not take part in the observed situation. Having raters participate in the setting increases the amount of information they see, which is an advantage compared to behavioral observation using video recordings. Future studies should consider having raters participate, especially in naturalistic, unstandardized setting which require an evaluation of the context. In sum, varying the parameters of behavioral observation (abstractness of observed constructs, duration vs. frequency, response format, level of participation) is needed to ensure that a sound rating system is developed (Wirtz & Caspar, 2002). Employing an observation software (Mangold International GmbH, 2017) would probably facilitate such a variation of the parameters in future studies.

Integration of information sources

Combining objectivity and proximity to behavioral descriptions of information on ADHD symptoms was one major motivation for the present dissertation. An assessment instrument is objective if every informant who uses the assessment instrument draws the same conclusions (Lienert & Raatz, 1994). Observation meets this requirement if several raters draw similar conclusions. The concordance between raters is measured by inter-rater reliability coefficients. In the empirical studies conducted in the present dissertation, observation of family context and ADHD symptoms proved to be an objective assessment instrument. Reported ADHD symptoms are biased by the informant of the questionnaire, which becomes clear when looking at informant discrepancies, for instance (Burns et al., 2014; Mitsis et al., 2000). However, it is important to critically question the requirement that information on ADHD symptoms be objective. For example, differences in parental and teacher reports could depict real behavioral differences between settings and provide an indication of where to implement interventions most effectively. For this reason, newer approaches suggest that the discrepancies in ADHD symptoms revealed by different sources of information provide meaningful information (Achenbach, 2011; De Los Reyes et al., 2014; Dirks et al., 2012). Following this line of interpretation, integrating reported, biased information with observed, objective information would be the gold standard. Integrating discrepant information from different sources is methodologically challenging and requires latent modeling (Achenbach, 2011). The present dissertation has associated but not integrated observed and reported information. Future studies should develop diagnostic algorithms that are capable of modeling meaningful differences between sources of information. Figure 2 depicts schematically how observation and reports together could provide a more conclusive picture than one source of information alone.

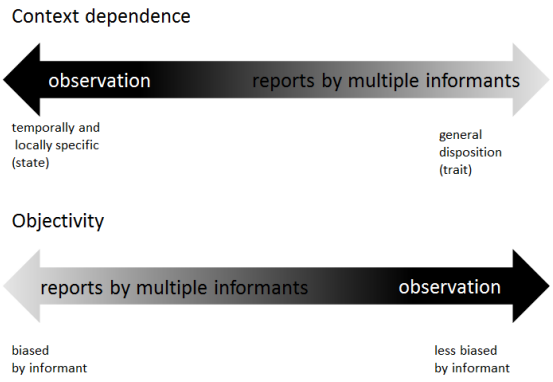

Figure 2 Visualization of the integration of assessment methods for ADHD symptoms (self-developed)

Ecological validity of context-dependent assessment

In the present dissertation, observation techniques were applied to a parent speech sample (*Study 1*), a delay situation in a laboratory (*Study 2*) and a simulated classroom situation (*Study 3*). The selection of these contexts is meaningful either because they are often mentioned in conjunction with the functional impairment associated with ADHD symptoms (*Studies 1* and *3*) or due to theoretical considerations (*Study 2*, Frazier et al., 2007; Sonuga-Barke, 2002). However, the information obtained in these contexts are states, and therefore associated with the specific context (Allport & Odbert, 1936). This information can likely be generalized to similar contexts, but its generalizability to daily life contexts (natural classrooms, naturally occurring delay situations) is questionable. Although the situation was designed to approximate natural classroom environments, it is not comparable to familiar surroundings. Moreover, the observations in the present study took place only once. From the outset, the problem of the generalizability of a single observation has raised concerns about the utility of observation in assessing ADHD symptoms (Barkley, 1997a). One method which has the power to combine context-dependent assessments and generalizability to daily life is ecologically momentary assessment (Mehl & Conner, 2013). Participants in an ecologically momentary assessment study provide information about a feature of interest within their daily life. Usually, repeated assessments take place within participants across hours or days. In most cases, portable devices (for instance smartphones) are used to capture participants' experiences in real life and real time (Mehl & Conner, 2013). The advantage is that information on natural contexts in daily life (for instance, current school subject) can be associated with a construct of interest (for instance, ADHD symptoms). This type of study accounts for the importance of the context of ADHD symptoms and assesses children's real life context. As an example, one study that applied ecologically momentary assessment investigated the temporal contingency of anger and mood in children with and without ADHD and their mothers. The moods of children with and without ADHD varied systematically with anger in mothers (Whalen et al., 2009). In sum, future studies should make an effort to assess ADHD symptoms in a context-dependent way using ecologically momentary assessment.

4.3 Implications for research and practice

The results of the empirical studies have important implications for research and practice.

Research

The present dissertation investigates the potential of behavioral observation for the assessment of ADHD symptoms. The clinical manifestation of ADHD, as described in the behavioral descriptions, is the anchor point for the identification of ADHD symptoms. Behavioral observation captures clinical manifestations of ADHD symptoms directly; for this reason, the advantages of behavioral observation have been discussed in the theoretical background. Reports, cognitive assessment methods and neurobiological mechanisms are not behavior-based, objective and context-dependent sources of information. An examination of behavioral observation revealed that this assessment method indeed fills the gap with regard to behavior-based, objective and context-dependent information on ADHD symptoms. Observation techniques should therefore be improved and further developed. However, relying on observable behavior and avoiding assessment methods which include interoception and neurobiological parameters carries the danger that the search for mechanisms and processes that lead to ADHD symptoms will be neglected. This critique resembles the caveat to research in the tradition of behaviorism (Skinner, 1966; Watson, 1994), which focuses solely on observable behavior. Therefore, the greatest challenge for future research in the assessment of child and adolescent psychopathology, including the assessment of ADHD symptoms, is to find algorithms and routines for integrating information from different sources (Achenbach, 2011). This integration should foster validity of diagnoses and in this way help to prevent overdiagnoses and misdiagnoses (Merten, Cwik, Margraf, & Schneider, 2017).

A newer attempt to integrate information on mental health from different levels is the Research Domain Criteria research framework[4], launched by the American National Institute of Mental Health (Insel et al., 2010; Peterson, 2015). The aim of this initiative is to revamp psychiatric nosology. The current diagnostic system is based on clinical manifestations in the form of valid and reliable diagnoses (APA, 2013; WHO,

[4] For more information: https://www.nimh.nih.gov/research-priorities/rdoc/index.shtml (retrieved, 27th March, 2017)

1993). As indicated in the theoretical background (see Chapter 1), the classification of ADHD according to clinical manifestations does not map well with findings from cognitive and neurobiological studies (Thome et al., 2012; Willcutt et al., 2005). The Research Domain Criteria framework seeks to develop an empiricism-driven psychiatric nosology on the basis of findings from genetic and neuroscientific research. The root of mental disorders is clearly seen in pathological genetic and neurological processes (Insel et al., 2010). An important caveat of this framework is that greater value is attached to genetic and neurological processes compared to reported or observed behavior. Instead of integration, the research efforts of the Research Domain Criteria framework could result in a reorganization along the lines of genetic and neuroscientific findings. Nevertheless, advancements within the Research Domain Criteria framework can clearly contribute to the integration of information on ADHD symptoms on different levels.

Practice

The present dissertation explored the utility of behavioral observation in the ADHD diagnostic process. The empirical studies revealed that observation as an assessment method provides important and unique information on ADHD symptoms. However, it is not possible to transfer the procedures used in the present studies directly to the diagnostic process. The analysis is too time consuming and not straightforward, particularly with regard to the cutting of video material and evaluation using paper-and-pencil procedures. In order to develop observation techniques specifically for ADHD symptoms which are usable in practice, it is worth considering diagnostic procedures for autism spectrum disorder (DGKJP & DGPPPN, 2016). The guidelines of the German professional associations say that behavioral observation is one of the main sources of diagnostic information along with clinical exploration and semi-structured interviews. In line with that specification, the Autism Diagnostic Observation Schedule provides an observation protocol system which can be adapted with regard to age and language proficiency (Bastiaansen et al., 2011; Lord et al., 2000). In this observation protocol, social occasions (called “presses”) are created in a one-to-one session with the diagnostician. The social occasions are “critical situations” for persons with autism spectrum disorder. These occasions are selected because reacting to the occasions requires a behavioral response which differentiates persons with autism spectrum disorder from healthy persons. For instance, for younger children, the diagnostician plays with a remote-controlled toy car and indicates that this car is interesting. A lack of

joint attention from the child and the diagnostician towards the car would be a sign of autistic behavior. Evaluations are directly entered into a questionnaire by the diagnostician (Lord et al., 2000). The development of an observational diagnostic protocol for ADHD symptoms in practice should follow a similar pattern as the Autism Diagnostic Observation Schedule. The present dissertation revealed that delay situations and classroom situations are possible critical situations which require a behavioral response that differentiates children with ADHD symptoms from children without ADHD symptoms. Critical and ecologically valid situations for ADHD are difficult to create because ADHD symptoms are primarily expressed in familiar surroundings (APA, 2015; Bundesärztekammer, 2005). However, the video recording set-up used in *Study 3* is easily transferrable to natural classrooms and allows for observations in natural surroundings. In sum, behavioral observation protocols need further development and should play a greater role in the diagnostic process of ADHD in practice.

Aside from the use of behavioral observation in diagnostic processes, therapeutic interventions can also profit from this method. One method that implements observation is the video-based method in educational counseling called marte meo (Bünder, 2017; Ervin, DuPaul, Kern, & Friman, 1998; Hawellek, 2012, 2014). This method relies on video recordings of critical situations in the daily life context of a client, such as dinner with the family. The counselor selects clips in the recording which show adaptive behavior, such as asking for something. The counselor watches this clip together with the client and points out the adaptive behavior. The clip is used to reinforce this behavior and serve as a role model for the client. Just as with the use of behavioral observation in the diagnostic process, the advantage of this method is that the advice is concrete and directly embedded in a relevant context. This context dependence allows for very individually tailored intervention approaches. For instance, a counselor could reinforce adaptive parenting behavior of parents with ADHD with this marte meo. This approach is much more embedded into the daily life context than a theoretical psychoeducational counseling session on adaptive parenting behavior for parents with children with ADHD.

4.4 Conclusion

The present dissertation investigated the potential of behavioral observation for the assessment of ADHD symptoms. The potential was measured by the parameters (1) inter-rater reliability, (2) association analyses between reported and observed symptoms and (3) association analyses between observed ADHD symptoms and meaningful outcome variables (*Study 2*: delay of gratification performance and *Study 3*: performance in a math test). The empirical studies in the present dissertation applied observation techniques to the family context of children with ADHD symptoms, ADHD symptoms during a delay of gratification task, and ADHD symptoms in a simulated classroom situation. (1) Sufficient inter-rater reliability could be reached in most cases for the observation of ADHD symptoms (*Study 2 and 3*) and the family context (*Study1*). The training of raters lasted around four hours respectively, which represents a reasonable cost-benefit ratio for the rich information provided by observation. (2) Association analyses between reported and observed symptoms revealed medium associations in most cases (*Study 3*) which implies, that observation provides information on ADHD symptoms which goes beyond information obtained by reported ADHD symptoms. (3) Association analyses between observed ADHD symptoms and meaningful outcome variables revealed (delay of gratification performance – *Study 2*; math performance – *Study 3*) differences from the associations involving reported ADHD symptoms. This finding strengthens the relevance for the inclusion of observation techniques in the diagnostic process of ADHD symptoms.

As expected, observation turned out to provide behavior-based, objective and context-dependent information. However, behavioral observation is an expensive assessment method in terms of personnel, time and information output. In order to save resources in the development of a rating scheme, the development of a training manual, the analysis of inter-rater reliability and the selection of a setting, standardized observation protocols for ADHD are urgently required. Moreover, future research should attempt to integrate discrepant information on ADHD symptoms, as for instance from observation and report. The development of standardized observation protocols as well as algorithms and routines for such an integration would represent a great advancement in the diagnostic process for ADHD symptoms.

German Summary (Zusammenfassung)

ADHS ist eine häufige kinder- und jugendpsychiatrische Störung (Willcutt, 2012), die sich durch Unaufmerksamkeit, Hyperaktivität und Impulsivität auszeichnet (APA, 2013; WHO, 1993). Die Identifikation von ADHS Symptomen basiert auf der klinischen Manifestation, die in den Verhaltensbeschreibungen in den diagnostischen Manualen festgehalten ist (APA, 2013; WHO, 1993). Das Auftreten von ADHS Symptoms ist kontextabhängig. Neben unstrukturierter klinischer Exploration, sind berichtete ADHS Symptome die Hauptinformationsquelle (Bundesärztekammer, 2005). Berichtete ADHS Symptome bilden die Verhaltensbeschreibungen von ADHS gut ab, und können darum als verhaltensnah angesehen werden. Berichtete ADHS Symptome werden allerdings stark durch den Informanten beeinflusst und sind darum nicht objektiv (Lienert & Raatz, 1994). Darüber hinaus drücken berichtete ADHS Symptome eine generelle Verhaltenstendenz aus, die sich nicht auf einen bestimmten Kontext bezieht. Berichtete ADHS Symptome können also keine Informationen liefern, die verhaltensnah, objektiv und kontextabhängig sind. Aus diesem Grund untersucht die vorliegende Dissertation Beobachtung als eine Erhebungsmethode, die diese Lücke füllen kann. Zu diesem Zweck wurden drei empirische Studien durchgeführt, die Beobachtungstechniken in Bezug auf den Familienkontext von Kindern mit ADHS Symptomen (Studie 1), auf ADHS Symptome in einer Belohnungsverzögerungssituation (Studie 2) und auf ADHS Symptome in einer simulierten Klassenraumsituation (Studie 3) anwendet.

Das Ziel von Studie 1 war es die Nützlichkeit von Expressed Emotion (EE) Skalen, gemessen mit der Fünf-Minuten Sprechprobe (FMSS) zu Erhebung des Familienkontext von Kindern mit ADHS Symptomen zu untersuchen. Zu diesem Zweck wurden die Interraterreliabilität und der Zusammenhang mit ADHS Symptomen berechnet. Dreiunddreißig Kinder (19 weiblich, $M(SD)_{Alter}$ = 10.65(1.34) Jahre, n = 6 mit einer ADHS Diagnose laut Elternbericht) und ein Elternteil nahmen an der Studie teil. Eltern und Kinder berichteten über die ADHS Symptomschwere der Kinder. Das FMSS wurde bei einem Elternteil innerhalb eines Telefoninterviews erhoben. Die Interraterreliabilität der EE Skalen Beziehung und kritische und positive Kommentare war ausreichend (ICC_{single} = .63 - .67). Die EE Skalen kritische und positive Kommentare zeigten konsistent für Selbstberichts und Elternberichtsmaße signifikante Zusammenhänge mit ADHS Symptomen. Die Ergebnisse zeigen, dass EE gemessen mit

FMSS ein vielversprechendes Instrument ist um den Familienkontext von Kindern mit ADHS Symptomen zu erfassen. Die Nützlichkeit von EE ist jedoch begrenzt, da das Konzept noch nicht ausreichend abgegrenzt ist und darum noch Weiterentwicklungsbedarf besteht.

Das Ziel von Studie 2 war es zu untersuchen, welche Variablen den Performanzunterschied zwischen Kindern mit ADHS und ihren gesunden Altersgenossen in einer Belohnungsverzögerungssituation beeinflussen. Konkret wurde untersucht, ob beobachtete und berichtete ADHS Symptome den Performanzunterschied in einer Belohnungsverzögerungssituation über Aufmerksamkeitsorientierung hinaus erklären. Einundsechzig Kinder (14 weiblich, $M(SD)_{\text{Alter}}$=10.40(1.58) Jahre, 26 diagnostiziert mit einer ADHS) nahmen an einer Belohnungsverzögerungsaufgabe teil, die gefilmt wurde. Die Videos wurden hinsichtlich Aufmerksamkeitsorientierung, Aktivität und Impulsivität beurteilt. Zehn Kinder warteten nicht auf die verzögerte Belohnung. Die Ergebnisse zeigen, dass Aufmerksamkeitsorientierung und beobachtete Aktivität während der Verzögerungssituation die Leistung in der Belohnugnsverzögerungsaufgabe über eine ADHS Diagnose und beobachtete Impulsivität hinaus erklären.

Das Ziel von Studie 3 war es die Nützlichkeit und Machbarkeit eines Beobachtungsprotokolls für ADHS Symptome in einer simulierten Klassenraumsituation zu untersuchen. Fünfunddreißig Kinder (20 weiblich, $M(SD)_{\text{Alter}}$ = 10.67(1.36) Jahre) nahmen an einer gefilmten simulierten Klassenraumsituation teil, die ein Mathetest und eine kompetitive Spielsituation beinhaltete. Die Kinder und ein Elternteil berichteten von der ADHD Symptomschwere der Kinder. Die Videos wurden hinsichtlich Unaufmerksamkeit, Hyperaktivität, und Impulsivität beurteilt. Die Interraterreliabilität der beobachteten Unaufmerksamkeit während des Mathetests und der beobachteten Impulsivität während der kompetitiven Spielsituation war zufriedenstellend. Die Interraterreliabilität von Hyperaktivitätsitems war teilweise zufriedenstellend. Die Werte der Zusammenhänge zwischen beobachteten und berichteten ADHS Symptomen drückten zumeist einen mittelgroßen Zusammenhang aus. Die beobachtete Unaufmerksamkeit hing stark negativ mit der Leistung im Mathetest zusammen. Die Weiterentwicklung des Beobachtungsprotokolls verspricht wertvolle Informationen über ADHS Symptome im Schulkontext zu liefern, die über die Informationen hinausgehen, die durch Fragebögen gewonnen werden können.

Die Ergebnisse der empirischen Studie zeigen, dass Beobachtung gemäß den Erwartungen verhaltensnahe, objektive und situationsspezifische Informationen zu ADHS Symptomen zur Verfügung stellt. Beobachtete ADHS Symptome hängen mit bedeutsamen Variablen zusammen und liefern Informationen, die über die berichteten ADHS Symptome hinausgehen. Beobachtung ist eine kostspielige und aufwändige Erhebungsmethode. Die Anwendung in der Praxis kann nur gerechtfertigt werden, wenn standardisierte Beobachtungsprotokolle entwickelt werden, die den Aufwand reduzieren. Außerdem sollte zukünftige Forschung diagnostische Routinen entwickelt, die Informationen von verschiedenen diagnostischen Quellen integrieren und nutzbar machen.

References

Abikoff, H., Gittelman, R., & Klein, D. F. (1980). Classroom observation code for hyperactive children: A replication of validity. *Journal of Consulting and Clinical Psychology*, *48*(5), 555–565. https://doi.org/10.1037//0022-006X.48.5.555

Achenbach, T. M. (2011). Commentary: Definitely more than measurement error: but how should we understand and deal with informant discrepancies? *Journal of Clinical Child and Adolescent Psychology*, *40*(1), 80–86. https://doi.org/10.1080/15374416.2011.533416

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. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *49*(3), 251–75. https://doi.org/10.1111/j.1469-7610.2007.01867.x

Alderson, R. M., Rapport, M. D., & Kofler, M. J. (2007). Attention-deficit/hyperactivity disorder and behavioral inhibition: A meta-analytic review of the stop-signal paradigm. *Journal of Abnormal Child Psychology*, *35*(5), 745–758. https://doi.org/10.1007/s10802-007-9131-6

Allport, G., & Odbert, H. S. (1936). *Trait-names A psycho-lexical study*. Princeton, N. J. and Albany, N. Y.: Psychological Review Company.

Alvarez, J., & Freides, D. (2004). Research on attention deficit hyperactivity disorder using the covert orienting paradigm of posner. *Developmental Neuropsychology*, *26*(2), 627–645. https://doi.org/10.1207/s15326942dn2602

Anastasi, A. (1983). traits, states and situations: a comprehensive view. In H. Wainer & S. Messick (Eds.), *Principles of modern psychological measurement* (pp. 345–356). London: Lawrence Erlbaum Associates, Inc.

Antrop, I. (2002). Stimulation seeking and hyperactive behavior in children with ADHD: a re-analysis. *Perceptual and Motor Skills*, *95*(5), 71. https://doi.org/10.2466/PMS.95.5.71-90

Antrop, I., Stock, P., Verté, S., Wiersema, J. R., Baeyens, D., & Roeyers, H. (2006). ADHD and delay aversion: The influence of non-temporal stimulation on choice for delayed rewards. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *47*(11), 1152–1158. https://doi.org/10.1111/j.1469-7610.2006.01619.x

APA. (2003). *Diagnostisches und Statistisches Manual Psychischer Störungen - Textrevision (DSM-IV-TR)[Diagnostic and Statistical Manual of Mental Disorders – Textrevision(DSM-IV-TR)]*. (H. Saß, H. Wittchen, M. Zaudig, & I. Houben, Eds.). Göttingen: Hogrefe.

APA. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

APA. (2015). *Diagnostisches und Statistisches Manual Psychischer Störungen DSM-5® [Diagnostic Statistical Manual of Mental Disorders DSM-5]*. (P. Falkai, H.

Wittchen, M. Döpfner, W. Gaebel, W. Maier, & H. Saß, Eds.). Göttingen: Hogrefe.

Asarnow, J. R., Tompson, M., Woo, S., & Cantwell, D. P. (2001). Is expressed emotion a specific risk factor for depression or a nonspecific correlate of psychopathology? *Journal of Abnormal Child Psychology*, *29*(6), 573–583. https://doi.org/10.1023/A:1012237411007

Ayduk, O., Mendoza-Denton, R., Mischel, W., Downey, G., Peake, P. K., & Rodriguez, M. (2000). Regulating the interpersonal self: strategic self-regulation for coping with rejection sensitivity. *Journal of Personality and Social Psychology*, *79*(5), 776–792. https://doi.org/10.1037/0022-3514.79.5.776

Bagwell, C. L., Molina, B. S. G., Pelham, W. E., & Hoza, B. (2001). Attention-deficit hyperactivity disorder and problems in peer relations: predictions from childhood to adolescence. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*(11), 1285–1292. https://doi.org/10.1097/00004583-200111000-00008

Balázs, J., & Keresztény, Á. (2014). Subthreshold attention deficit hyperactivity in children and adolescents: a systematic review. *European Child & Adolescent Psychiatry*, *23*(6), 393–408. https://doi.org/10.1007/s00787-013-0514-7

Banaschewski, T., Roessner, V., Uebel, H., & Rothenberger, A. (2004). Neurobiologie der Aufmerksamkeitsdefizit- /hyperaktivitätsstörung (ADHS) [Neurobiology of Attention Deficit Hyperactivity Disorder (ADHD)]. *Kindheit und Entwicklung*, *13*(3), 137–147. https://doi.org/10.1026/0942-5403.13.3.137

Barkley, R. (1991). The ecological validity of laboratory and analog assessment methods of adhd symptoms. *Journal of Abnormal Child Psychology*, *19*(2), 149–178.

Barkley, R. (1997a). *ADHD and the nature of self-control*. New York: Guilford Press.

Barkley, R. (1997b). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin*, *121*(1), 65–94. https://doi.org/10.1037/0033-2909.121.1.65

Barry, T. D., Lyman, R. D., & Klinger, L. G. (2002). Academic underachievement and attention-deficit / hyperactivity disorder: the negative impact of symptom severity on school performance. *Journal of School Psychology*, *40*(3), 259–283. https://doi.org/10.1016/S0022-4405(02)00100-0

Bastiaansen, J. A., Meffert, H., Hein, S., Huizinga, P., Ketelaars, C., Pijnenborg, M., … de Bildt, A. (2011). Diagnosing autism spectrum disorders in adults: the use of autism diagnostic observation schedule (ADOS) module 4. *Journal of Autism and Developmental Disorders*, *41*(9), 1256–1266. https://doi.org/10.1007/s10803-010-1157-x

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-report and finger movements: whatever happened to actual behavior? *Perspectives on Psychological Science*, *2*(4), 396–403. https://doi.org/10.1111/j.1745-6916.2007.00051.x

Bertram, H., Naumann, S., Harstick-Koll, S., Kuschel, A., Heinrichs, K., Brix, G., & Döpfner, M. (2008). *Revised Family Observation Schedule (FOS-RIII; Sanders,*

M.R., Waugh, L., Tully, L. & Hynes, K., 1996) - German version.

Biederman, J. (2000). Age-dependent decline of symptoms of attention deficit hyperactivity disorder: impact of remission definition and symptom type. *American Journal of Psychiatry*, *157*(5), 816–818. https://doi.org/10.1176/appi.ajp.157.5.816

Biederman, J., Faraone, S., Milberger, S., Curtis, S., Chen, L., Marrs, A., … Spencer, T. (1996). Predictors of persistence and remission of ADHD into adolescence: results from a four-year prospective follow-up study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *35*(3), 343–351. https://doi.org/10.1097/00004583-199603000-00016

Biederman, J., Petty, C., Clarke, A., Lomedico, A., & Faraone, S. (2011). Predictors of persistent ADHD: An 11-year follow-up study. *Journal of Psychiatric Research*, *45*(2), 150–155. https://doi.org/10.1016/j.jpsychires.2010.06.009

Breuer, D., Rettig, K., & Döpfner, M. (2009). Die Erfassung von Aufmerksamkeits- und Verhaltensproblemen im Unterricht mit dem Fragebogen zur Verhaltensbeurteilung im Unterricht (FVU) [The asssessment of attention problems and behavioral problems at school with the Fragebogen zur Verhaltensbeurteilun. *Diagnostica*, *55*(1), 11–19. https://doi.org/10.1026/0012-1924.55.1.11

Brown, G. W., Monck, E. M., Carstairs, G. M., & Wing, J. K. (1962). Influence of family life on the course of schizophrenic illness. *British Journal of Preventive and Social Medicine*, *16*(2), 55.-68. https://doi.org/10.1192/bjp.121.3.241

Bünder, P. (2017). MarteMeo - Netzwerk Deutschland. Retrieved April 18, 2017, from http://www.martemeo-deutschland.de/index.html

Bundesärztekammer. (2005). *Stellungnahme zur „Aufmerksamkeitsdefizit- / Hyperaktivitätsstörung (ADHS)“ - Langfassung - [statement on attention deficit/hyperactivity disorder (ADHD) - longversion]*. Retrieved from http://www.bundesaerztekammer.de/fileadmin/user_upload/downloads/ADHSLang.pdf

Bunte, T. L., Schoemaker, K., Hessen, D. J., Van Der Heijden, P. G. M., & Matthys, W. (2013). Clinical usefulness of the Kiddie-Disruptive Behavior Disorder Schedule in the diagnosis of DBD and ADHD in preschool children. *Journal of Abnormal Child Psychology*, *41*(5), 681–690. https://doi.org/10.1007/s10802-013-9732-1

Burns, G. L., Servera, M., del Mar Bernad, M., Carrillo, J. M., & Geiser, C. (2014). Ratings of ADHD symptoms and academic impairment by mothers, fathers, teachers, and aides: Construct validity within and across settings as well as occasions. *Psychological Assessment*, *26*(4), 1247–1258. https://doi.org/10.1037/pas0000008

Butzlaff, R. L., & Hooley, J. M. (1998). Expressed emotion and psychiatric relapse: a meta-analysis. *Archives of General Psychiatry*, *55*(6), 547–552. https://doi.org/10.1001/archpsyc.55.6.547

Campbell, S. B., & Von Stauffenberg, C. (2009). Delay and inhibition as early predictors of ADHD symptoms in third grade. *Journal of Abnormal Child Psychology*, *37*(1), 1–15. https://doi.org/10.1007/s10802-008-9270-4

Cartwright, K. L., Bitsakou, P., Daley, D., Gramzow, R. H., Psychogiou, L., Simonoff, E., … Sonuga-Barke, E. J. S. (2011). Disentangling child and family influences on maternal expressed emotion toward children with attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *50*(10), 1042–1053. https://doi.org/10.1016/j.jaac.2011.07.006

Christiansen, H., Oades, R. D., Psychogiou, L., Hauffa, B. P., & Sonuga-Barke, E. J. (2010). Does the cortisol response to stress mediate the link between expressed emotion and oppositional behavior in Attention-Deficit/Hyperactivity-Disorder (ADHD)? *Behavioral and Brain Functions*, *6*(1), 45. https://doi.org/10.1186/1744-9081-6-45

Coghill, D., & Sonuga-Barke, E. J. S. (2012). Annual research review: Categories versus dimensions in the classification and conceptualisation of child and adolescent mental disorders - Implications of recent empirical study. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *53*(5), 469–489. https://doi.org/10.1111/j.1469-7610.2011.02511.x

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: Lawrence Erlbaum Associates, Inc.

Colombo, J. (2001). The development of visual attention in infancy. *Annual Review of Psychology*, *52*(1), 337–367. https://doi.org/10.1146/annurev.psych.52.1.337

Conners, C. K. (2011). *Conners 3rd edition: Manual*. Toronto: Multi-Health Systems.

Cortese, S., Ferrin, M., Brandeis, D., Buitelaar, J., Daley, D., Dittmann, R. W., … Sonuga-Barke, E. J. S. (2015). Cognitive training for attention-deficit/hyperactivity disorder: Meta-analysis of clinical and neuropsychological outcomes from randomized controlled trials. *Journal of the American Academy of Child and Adolescent Psychiatry*, *54*(3), 164–174. https://doi.org/10.1016/j.jaac.2014.12.010

Cortese, S., Ferrin, M., Brandeis, D., Holtmann, M., Aggensteiner, P., Daley, D., … Zuddas, A. (2016). Neurofeedback for attention-deficit/hyperactivity disorder: meta-analysis of clinical and neuropsychological outcomes from randomized controlled trials. *Journal of the American Academy of Child & Adolescent Psychiatry*, *55*(6), 444–455. https://doi.org/10.1016/j.jaac.2016.03.007

Crosbie, J., Pérusse, D., Barr, C. L., & Schachar, R. J. (2008). Validating psychiatric endophenotypes: Inhibitory control and attention deficit hyperactivity disorder. *Neuroscience and Biobehavioral Reviews*, *32*(1), 40–55. https://doi.org/10.1016/j.neubiorev.2007.05.002

Cunningham, C. E., & Siegel, L. S. (1987). Peer interactions of normal and attention-deficit-disordered boys during free-play, cooperative task, and simulated classroom situations. *Journal of Abnormal Child Psychology*, *15*(2), 247–268. https://doi.org/10.1007/BF00916353

Daley, D., & Birchwood, J. (2010). ADHD and academic performance: Why does ADHD impact on academic performance and what can be done to support ADHD children in the classroom? *Child: Care, Health and Development*, *36*(4), 455–464. https://doi.org/10.1111/j.1365-2214.2009.01046.x

Daley, D., Sonuga-Barke, E. J. S., & Thompson, M. (2003). Assessing expressed emotion in mothers of preschool AD/HD children: Psychometric properties of a

modified speech sample. *British Journal of Clinical Psychology*, *42*(1), 53–67. https://doi.org/10.1348/014466503762842011

Daley, D., Van Der Oord, S., Ferrin, M., Danckaerts, M., Doepfner, M., Cortese, S., & Sonuga-Barke, E. J. S. (2014). Behavioral interventions in attention-deficit/hyperactivity disorder: A meta-analysis of randomized controlled trials across multiple outcome domains. *Journal of the American Academy of Child and Adolescent Psychiatry*, *53*(8), 835–847. https://doi.org/10.1016/j.jaac.2014.05.013

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. *Journal of Clinical Child & Adolescent Psychology*, *40*(1), 1–9. https://doi.org/10.1080/15374416.2011.533405

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D. E., & Rabinowitz, J. (2014). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, *141*(2), 858–900. https://doi.org/10.1037/a0038498

Deault, L. C. (2010). A systematic review of parenting in relation to the development of comorbidities and functional impairments in children with attention-deficit/ hyperactivity disorder (ADHD). *Child Psychiatry and Human Development*, *41*(2), 168–192. https://doi.org/10.1007/s10578-009-0159-4

DGKJP. (2007). Leitlinien zur Diagnostik und Therapie von psychischen Störungen im Säuglings-, Kindes- und Jugendalter [Guidelines for diagnosis and therapy of psychiatric illnesses in infants, children and adolescents]. In Dt.Ges.f. Kinder- und Jugendpsychiatrie und Psychotherapie [German Association for children and adolescents psychiatry and psychotherapy] (Ed.) (3. überarb, pp. 239–254). Deutscher Ärzte Verlag.

DGKJP. (2016). *ADHS bei Kindern, Jugendlichen und Erwachsenen [ADHD in children, adolescents and adults]*. (T. Banaschewski, Ed.). Mannheim: Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften [Working unit of scientific medical associations]. Retrieved from http://www.awmf.org/leitlinien/detail/anmeldung/1/ll/028-045.html

DGKJP, & DGPPPN. (2016). *Leitlinien zur Diagnostik und Therapie von psychischen Störungen im Säuglings-, Kindes- und Jugendalter [Guidelines for diagnosis and therapy of psychiatric illnesses in infants, children and adolescents]*. Retrieved from http://www.awmf.org/leitlinien/detail/anmeldung/1/ll/028-045.html

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. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *53*(5), 558–574. https://doi.org/10.1111/j.1469-7610.2012.02537.x

Döpfner, M., & Görtz-Dorten, A. (2016). *Diagnostik-System für psychische Störungen nach ICD-10 und DSM-5 für Kinder- und Jugendliche (DISYPS-III) [Diagnostic system for psychiatric disorders according to ICD-10 and DSM-5 for children and adolescents]*. Bern: Huber.

Downer, J. T., Booren, L. M., Lima, O. K., Luckner, A. E., & Pianta, R. C. (2010). The

Individualized Classroom Assessment Scoring System (inCLASS): Preliminary reliability and validity of a system for observing preschoolers' competence in classroom interactions. *Early Childhood Research Quarterly*, *25*(1), 1–16. https://doi.org/10.1016/j.ecresq.2009.08.004

Doyle, A. E., Biederman, J., Seidman, L. J., Weber, W., & Faraone, S. V. (2000). Diagnostic efficiency of neuropsychological test scores for discriminating boys with and without attention deficit-hyperactivity disorder. *Journal of Consulting and Clinical Psychology*, *68*(3), 477–488. https://doi.org/10.1037//0022-006X.68.3.477

Düker, H., & Lienert, G. A. (2001). *Konzentrations-Leistungs-Test: KLT-R [Concentration-Performance-Test: KLT-R]*. Göttingen: Hogrefe.

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. *School Psychology Quarterly*, *29*(4), 409–421. https://doi.org/10.1037/spq0000095

Eid, M., Gollwitzer, M., & Schmitt, M. (2015). *Statistik und Forschungsmethoden : Lehrbuch [Statistics and research methodology: a textbook]*. Weinheim: Beltz.

Eigsti, I.-M., Zayas, V., Mischel, W., Shoda, Y., Ayduk, O., Dadlani, M. B., … Casey, B. J. (2006). Predicting cognitive control from preschool to late adolescence and young adulthood. *Psychological Science*, *17*(6), 478–484. https://doi.org/10.1111/j.1467-9280.2006.01732.x

Ervin, R. a, DuPaul, G. J., Kern, L., & Friman, P. C. (1998). Classroom-based functional and adjunctive assessments: proactive approaches to intervention selection for adolescents with attention deficit hyperactivity disorder. *Journal of Applied Behavior Analysis*, *31*(1), 65–78. https://doi.org/10.1901/jaba.1998.31-65

Faraone, S. V., Asherson, P., Banaschewski, T., Biederman, J., Buitelaar, J. K., Ramos-Quiroga, J. A., … Franke, B. (2015). Attention-deficit/hyperactivity disorder. *Nature Reviews Disease Primers*, *1*, 15020. https://doi.org/10.1038/nrdp.2015.20

Faraone, S. V., Perlis, R. H., Doyle, A. E., Smoller, J. W., Goralnick, J. J., Holmgren, M. A., & Sklar, P. (2005). Molecular genetics of attention-deficit/hyperactivity disorder. *Biological Psychiatry*, *57*(11), 1313–1323. https://doi.org/10.1016/j.biopsych.2004.11.024

Faraone, S. V, & Biederman, J. (1998). Neurobiology of attention-deficit hyperactivity disorder. *Biological Psychiatry*, *44*(10), 951–958. https://doi.org/10.1016/S0006-3223(98)00240-6

Frazier, T. W., Youngstrom, E. A., Glutting, J. J., & Watkins, M. W. (2007). ADHD and achievement: meta-analysis of the child, adolescent, and adult literatures and a concomitant study with college students. *Journal of Learning Disabilities*, *40*(1), 49–65. https://doi.org/10.1177/00222194070400010401

Furr, M., & Funder, D. C. (2007). Behavioral observation. In R. Robins, C. Fraley, & R. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 273–291). New York: Guilford Press.

Gamer, M., Lemon, J., Fellows, I., & Sing, P. (2012). irr: Various coefficients of

interrater reliability and agreement (Version 0.84) [software]. Retrieved from http://cran.r-project.org/package=irr

Gaub, M., & Carlson, C. L. (1997). Behavioral characteristics of DSM-IV ADHD subtypes in a school-based population. *Journal of Abnormal Child Psychology*, *25*(2), 103–111. https://doi.org/10.1023/A:1025775311259

Gawrilow, C., Gollwitzer, P. M., & Oettingen, G. (2011). If-then plans benefit delay of gratification performance in children with and without ADHD. *Cognitive Therapy and Research*, *35*(5), 442–455. https://doi.org/10.1007/s10608-010-9309-z

Glutting, J. J., Youngstrom, E. A., & Watkins, M. W. (1996). Situational specificity and generality of test behaviors for samples of normal and referred children. *School Psycology Review*, *25*(1), 94–107.

Godwin, K. E., Almeda, M. V., Seltman, H., Kai, S., Skerbetz, M. D., Baker, R. S., & Fisher, A. V. (2016). Off-task behavior in elementary school children. *Learning and Instruction*, *44*, 128–143. https://doi.org/10.1016/j.learninstruc.2016.04.003

Gomez, R., Burns, G. L., Walsh, J. a, & de Moura, M. A. (2003). A multitrait-multisource confirmatory factor analytic approach to the construct validity of ADHD rating scales. *Psychological Assessment*, *15*(1), 3–16. https://doi.org/10.1037/1040-3590.15.1.3

Gordon, M. (2006). Symptoms versus impairment: the case for respecting DSM-IV's criterion D. *Journal of Attention Disorders*, *9*(3), 465–475. https://doi.org/10.1177/1087054705283881

Gordon, M., DiNiro, D., Mettelman, B. B., & Tallmadge, J. (1989). Observations of test behavior, quantitative scores, and teacher ratings. *Journal of Psychoeducational Assessment*, *7*(2), 141–147. https://doi.org/10.1177/073428298900700204

Graf Schimmelmann, B., Meng, H., Koch, E., Karle, M., Preuss, U., & Schulte-Markwort, M. (2003). Expressed Emotion im Verlauf kinder- und jugendpsychiatrischer Störungen [Expressed Emotion and the outcome of child and adolescent psychiatric disorders]. *Fortschritte der Neurologie und Psychiatrie*, *71*(10), 517–526. https://doi.org/10.1055/s-2003-42870

Hallgreen, K. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(2), 23–34. https://doi.org/10.1016/j.pestbp.2011.02.012

Hawellek, C. (2012). *Entwicklungsperspektiven öffnen : Grundlagen beobachtungsbegleiteter Beratung nach der Marte-Meo-Methode [Open developmental perspectives: Basics of the observational counseling method marte meo]*. Göttingen: Vandenhoeck & Ruprecht.

Hawellek, C. (2014). Einladung zum Perspektivwechsel. Die Möglichkeiten der Marte-Meo-Methode in Beratung und Psychotherapie [Inviting a change of perspective - the potential of the marte meo method in counseling and psychotherapy]. *Familiendynamik*, *1*(38–49).

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, *1*(1), 77–89. https://doi.org/10.1080/19312450709336664

Honour, H. (2016). What is Krippendor ’s Alpha? Retrieved April 18, 2017, from https://github.com/foolswood/krippendorffs_alpha/blob/master/krippendorff.pdf

IDEA Center Indivdual Development and Adaptive Education. (2011). GIDeCA Gene by Environment Interactions on Decision Making in Children with different ADHD Symptoms. Retrieved April 18, 2017, from http://www.idea-frankfurt.eu/en/research/theme/individual-development/gideca?set_language=en

Imeraj, L., Antrop, I., Roeyers, H., Deboutte, D., Deschepper, E., Bal, S., & Sonuga-Barke, E. (2016). The impact of idle time in the classroom: differential effects on children with ADHD. *Journal of Attention Disorders*, *20*(1), 71–81. https://doi.org/10.1177/1087054713478464

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., … Wang, P. (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*, *167*(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

Jacob, R. G., O’Leary, K. D., & Rosenblad, C. (1978). Formal and informal classroom settings: Effects on hyperactivity. *Journal of Abnormal Child Psychology*, *6*(1), 47–59. https://doi.org/10.1007/BF00915781

Johnson, K. A., Wiersema, J. R., & Kuntsi, J. (2009). What would Karl Popper say? Are current psychological theories of ADHD falsifiable? *Behavioral and Brain Functions*, *5*(1), 15. https://doi.org/10.1186/1744-9081-5-15

Johnston, C., & Mash, E. J. (2001). Families of children with attention-deficit / hyperactivity disorder : review and recommendations for future research. *Clinical Child and Family Psychology Review*, *4*(3), 183–207. https://doi.org/10.1023/A:1017592030434

Kofler, M. J., Rapport, M. D., & Matt Alderson, R. (2008). Quantifying ADHD classroom inattentiveness, its moderators, and variability: A meta-analytic review. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *49*(1), 59–69. https://doi.org/10.1111/j.1469-7610.2007.01809.x

Krippendorff, K. (1970). Estimating the reliability, systematic error and rondom error of interval data. *Educational and Psychological Measurement*, *30*(1), 61–70. https://doi.org/10.1177/001316447003000105

Kuntsi, J., Stevenson, J., Oosterlaan, J., & Sonuga-Barke, E. J. S. (2001). Test-retest reliability of a new delay aversion task and executive function measures. *British Journal of Developmental Psychology*, *19*(3), 339–348. https://doi.org/10.1348/026151001166137

Lauth, G. W., Heubeck, B. G., & Mackowiak, K. (2006). Observation of children with attention-deficit hyperactivity (ADHD) problems in three natural classroom contexts. *British Journal of Educational Psychology*, *76*(2), 385–404. https://doi.org/10.1348/000709905X43797

Leeb, B., Hahlweg, K., Goldstein, M. J., Feinstein, E., Mueller, U., Dose, M., & Magana-Amato, A. (1991). Cross-national reliability, concurrent validity, and stability of a brief method for assessing expressed emotion. *Psychiatry Research*, *39*(1), 25–31. https://doi.org/10.1016/0165-1781(91)90005-A

Lidzba, K., Christiansen, H., & Drechsler, R. (2013). *Conners Skalen zu Aufmerksamkeit und Verhalten–3. Deutschsprachige Adaptation der Conners 3 Edition [Conners scales on attention and behavior-3. German adaption of the Conners 3 edition]*. Bern: Huber.

Lienert, G. A., & Raatz, U. (1994). *Testaufbau und Testanalyse [Testconstruction and Testanalysis]*. Weinheim: Psychologie Verlags Union.

Loe, I. M., & Feldman, H. M. (2007). Academic and educational outcomes of children with ADHD. *Journal of Pediatric Psychology*, *32*(6), 643–654. https://doi.org/10.1093/jpepsy/jsl054

Lord, C., Risi, S., Lambrecht, L., Cook, E. H. J., Leventhal, B. L., DiLavore, P. C., … Rutter, M. (2000). The Autism Diagnostic Schedule – Generic: A standard measures of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, *30*(3), 205–223. https://doi.org/10.1023/A:1005592401947

Magana, A. B., Jenkins, J., & Miklowitz, J. (1986). A brief method for assessing Expressed Emotion in relatives of psychiatric patients. *Journal of Psychiatric Research*, (17), 203–212. https://doi.org/http://dx.doi.org/10.1016/0165-1781(86)90049-1

Maguin, E., Nochajski, T. H., De Wit, D. J., & Safyer, A. (2015). Examining the validity of the adapted alabama parenting questionnaire-parent global report version. *Psychological Assessment*, *28*(5), 613–625. https://doi.org/10.1037/pas0000214

Mahone, E. M., & Pritchard, A. E. (2013). Neuropsychological assessment of ADHD in preschoolers. In J. Ghuman & H. Ghuman (Eds.), *ADHD in Preschool Children* (pp. 42–65). Oxford University Press. https://doi.org/10.1093/med/9780199948925.003.0003

Manfra, L., Davis, K. D., Ducenne, L., & Winsler, A. (2014). Preschoolers' motor and verbal self-control strategies during a resistance-to-temptation task. *The Journal of Genetic Psychology*, *175*(4), 332–345. https://doi.org/10.1080/00221325.2014.917067

Mangold International GmbH. (2017). Lösungen für die Beobachtungsforschung und Datenanalyse [Soluations for observational research and data analysis]. Retrieved April 18, 2017, from https://www.mangold-international.com/de/impressum

McConaughy, S. H., Harder, V. S., Antshel, K. M., Gordon, M., Eiraldi, R., & Dumenci, L. (2010). Incremental validity of test session and classroom observations in a multimethod assessment of attention deficit/hyperactivity disorder. *Journal of Clinical Child & Adolescent Psychology*, *39*(5), 650–666. https://doi.org/10.1080/15374416.2010.501287

Mehl, M. R., & Conner, T. S. (Eds.). (2013). *Handbook of research methods for studying daily life*. Guilford Publications.

Merkt, J., Siniatchkin, M., & Petermann, F. (2016). Neuropsychological measures in the diagnosis of ADHD in preschool: can developmental research inform diagnostic practice? *Journal of Attention Disorders*. https://doi.org/10.1177/1087054716629741

Merrell, C., & Tymms, P. B. (2001). Inattention, hyperactivity and impulsiveness: their impact on academic achievement and progress. *The British Journal of Educational Psychology, 71*, 43–56. https://doi.org/10.1348/000709901158389

Merten, E. C., Cwik, J. C., Margraf, J., & Schneider, S. (2017). Overdiagnosis of mental disorders in children and adolescents (in developed countries). *Child and Adolescent Psychiatry and Mental Health*, *11*(1), 5. https://doi.org/10.1186/s13034-016-0140-5

Metcalfe, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: dynamics of willpower. *Psychological Review*, *106*(1), 3–19. https://doi.org/10.1037/0033-295X.106.1.3

Millstein, R. B., Wilens, T. E., Biederman, J., & Spencer, T. J. (1997). Presenting ADHD symptoms and subtypes in clinically referred adults with ADHD. *Journal of Attention Disorders*, *2*(3), 159–166. https://doi.org/10.1177/108705479700200302

Mischel, W. (1996). From good intentions to willpower. In P. Gollwitzer & J. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 197–218). New York, NY: Guilford Press.

Mischel, W., & Baker, N. (1975). Cognitive appraisals and transformations in delay behavior. *Journal of Personality and Social Psychology*, *31*(2), 254–261. https://doi.org/10.1037/h0076272

Mischel, W., Ebbesen, E., & Zeiss, A. (1972). Cognitive and attentional mechanisms in delay of gratification. *Journal of Personality and Social Psychology*, *21*(2), 204–218. https://doi.org/10.1037/h0032198

Mischel, W., & Moore, B. (1973). Effects of attention to symbolically presented rewards on self-control. *Journal of Personality and Social Psychology*, *28*(2), 172–179.

Mitsis, E. M., McKay, K. E., Schulz, K. P., Newcorn, J. H., & Halperin, J. M. (2000). Parent–teacher concordance for DSM-IV attention-deficit/hyperactivity disorder in a clinic-referred sample. *Journal of the American Academy of Child & Adolescent Psychiatry*, *39*(3), 308–313. https://doi.org/10.1097/00004583-200003000-00012

Mulder, M. J., Bos, D., Weusten, J. M. H., van Belle, J., van Dijk, S. C., Simen, P., … Durston, S. (2010). Basic impairments in regulating the speed-accuracy tradeoff predict symptoms of attention-deficit/hyperactivity disorder. *Biological Psychiatry*, *68*(12), 1114–1119. https://doi.org/10.1016/j.biopsych.2010.07.031

Müller, A., Candrian, G., & Kropotov, J. (2011). *ADHS-Neurodiagnostik in der Praxis [ADHD Neurodiagnosis in practice]*. Berlin-Heidelberg: Springer.

Musser, E. D., Karalunas, S. L., Dieckmann, N., Peris, T. S., & Nigg, J. T. (2016). Attention-deficit/hyperactivity disorder developmental trajectories related to parental expressed emotion. *Journal of Abnormal Psychology*, *125*(2), 182–195. https://doi.org/10.1037/abn0000097

Neubauer, A., Gawrilow, C., & Hasselhorn, M. (2011). Belohnungsaufschub: Ein Ansatz zur Frühprognose volitionaler Kompetenzen [Delay of gratification: an approach for early prognosis of volitional competences]. In M. Hasselhorn & W.

Schneider (Eds.), *Frühprognose schulischer Kompetenzen [early prognosis of academic competences]* (Hogrefe, pp. 203–220). Göttingen.

Neuenschwander, R., & Blair, C. (2017). Zooming in on children's behavior during delay of gratification: Disentangling impulsigenic and volitional processes underlying self-regulation. *Journal of Experimental Child Psychology*, *154*, 46–63. https://doi.org/10.1016/j.jecp.2016.09.007

Norén Selinus, E., Molero, Y., Lichtenstein, P., Anckarsäter, H., Lundström, S., Bottai, M., & Hellner Gumpert, C. (2016). Subthreshold and threshold attention deficit hyperactivity disorder symptoms in childhood: psychosocial outcomes in adolescence in boys and girls. *Acta Psychiatrica Scandinavica*, *134*(6), 533–545. https://doi.org/10.1111/acps.12655

Norman, L. J., Carlisi, C., Lukito, S., Hart, H., Mataix-Cols, D., Radua, J., & Rubia, K. (2016). Structural and functional brain abnormalities in attention-deficit/hyperactivity disorder and obsessive-compulsive disorder. *JAMA Psychiatry*, *73*(8), 815. https://doi.org/10.1001/jamapsychiatry.2016.0700

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity*, *41*(5), 673–690. https://doi.org/10.1007/s11135-006-9018-6

Patros, C. H. G., Alderson, R. M., Kasper, L. J., Tarle, S. J., Lea, S. E., & Hudec, K. L. (2016). Choice-impulsivity in children and adolescents with attention-deficit/hyperactivity disorder (ADHD): A meta-analytic review. *Clinical Psychology Review*, *43*, 162–174. https://doi.org/10.1016/j.cpr.2015.11.001

Pauli-Pott, U., & Becker, K. (2011). Neuropsychological basic deficits in preschoolers at risk for ADHD: A meta-analysis. *Clinical Psychology Review*, *31*(4), 626–637. https://doi.org/10.1016/j.cpr.2011.02.005

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: our words, our selves. *Annual Review of Psychology*, *54*(1), 547–577. https://doi.org/10.1146/annurev.psych.54.101601.145041

Perez, E., Turner, M., Fisher, A., Lockwood, J., & Daley, D. (2014). Linguistic analysis of the Preschool Five Minute Speech Sample: What the parents of Preschool children with early signs of ADHD say and how they say it? *PLoS ONE*, *9*(9). https://doi.org/10.1371/journal.pone.0106231

Peris, T., & Baker, B. (2000). Applications of the expressed emotion construct to young children with externalizing behavior: stability and prediction over time. *Journal of Child Psychology and Psychiatry*, *41*(4), 457–462. https://doi.org/http://dx.doi.org/10.1111/1469-7610.00630

Peris, T., & Hinshaw, S. (2003). Family dynamics and preadolescent girls with ADHD: The relationship between expressed emotion, ADHD symptomatology, and comorbid disruptive behavior. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *44*(8), 1177–1190. https://doi.org/10.1111/1469-7610.00199

Peris, T., & Miklowitz, D. (2015). Parental expressed emotion and youth psychopathology: new directions for an old construct. *Child Psychiatry and Human Development*, *46*(6), 863–873. https://doi.org/10.1007/s10578-014-0526-7

Peterson, B. S. (2015). Editorial: Research Domain Criteria (RDoC): A new psychiatric nosology whose time has not yet come. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *56*(7), 719–722. https://doi.org/10.1111/jcpp.12439

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2006). *Classroom assessment scoring system: Manual k-3 version*. Charlotesville, VA: Center for Advanced Study of Teaching and Learning, University of Virginia.

Polanczyk, G. (2007). The worldwide prevalence of ADHD: A systematic review and metaregression analysis. *American Journal of Psychiatry*, *164*(6), 942. https://doi.org/10.1176/appi.ajp.164.6.942

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. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *56*(3), 345–365. https://doi.org/10.1111/jcpp.12381

Psychogiou, L., Daley, D. M., Thompson, M. J., & Sonuga-Barke, E. J. S. (2007). Mothers' expressed emotion toward their school-aged sons: Associations with child and maternal symptoms of psychopathology. *European Child and Adolescent Psychiatry*, *16*(7), 458–464. https://doi.org/10.1007/s00787-007-0619-y

R Core Team and contributors worldwide. (n.d.). The R Stats Package 3.4.0.

R Development Core Team. (2012). R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing.

Rapport, M. D., Kofler, M. J., Alderson, R. M., Timko, T. M., & Dupaul, G. J. (2009). Variability of attention processes in ADHD: observations from the classroom. *Journal of Attention Disorders*, *12*(6), 563–573. https://doi.org/10.1177/1087054708322990

Reinelt, T., Wirth, A., Rauch, W., & Gawrilow, C. (2014). Duration discrimination predicts delay of gratification in children with and without ADHD. *Procedia - Social and Behavioral Sciences*, *126*, 220–221. https://doi.org/10.1016/j.sbspro.2014.02.383

Rejchtman, G. (2011). *Ubongo–Das Kartenspiel [Ubongo – card game]*. Stuttgart: KOSMOS Verlag.

Rescorla, L. A., Ginzburg, S., Achenbach, T. M., Ivanova, M. Y., Almqvist, F., Begovac, I., … Verhulst, F. C. (2013). Cross-informant agreement between parent-reported and adolescent self-reported problems in 25 societies. *Journal of Clinical Child & Adolescent Psychology*, *42*(2), 262–273. https://doi.org/10.1080/15374416.2012.717870

Reynolds, B., & Schiffbauer, R. (2005). Delay of Gratification and Delay Discounting: A unifying feedback model of delay-related impulsive behavior. *The Psychological Record*, *55*, 439–460.

Richards, J. S., Vásquez, A. A., Rommelse, N. N. J., Oosterlaan, J., Hoekstra, P. J., Franke, B., … Buitelaar, J. K. (2014). A follow-up study of maternal expressed emotion toward children with attention-deficit/hyperactivity disorder (ADHD): Relation with severity and persistence of adhd and comorbidity. *Journal of the*

American Academy of Child and Adolescent Psychiatry, *53*(3), 311–320. https://doi.org/10.1016/j.jaac.2013.11.011

Roberts, M. A. (1990). A behavioral observation method for differentiating hyperactive and agressive boys. *Journal of Abnormal Child Psychology*, *18*(2), 131–142. https://doi.org/10.1007/BF00910726

Rodriguez, M. L., Mischel, W., & Shoda, Y. (1989). Cognitive person variables in the delay of gratification of older children at risk. *Journal of Personality and Social Psychology*, *57*(2), 358–367. https://doi.org/10.1037/0022-3514.57.2.358

Rutter, M., & Brown, G. W. (1966). The reliability and validity of measures of family life and relationships in families containig a psychiatric patient. *Social Psychiatry*, *1*(1), 38–53. https://doi.org/10.1007/BF00583828

Sander, A. (1983). Towards a model of stress and human performance. *Acta Psychologica*, *53*, 61–97. https://doi.org/doi.org/10.1016/0001-6918(83)90016-1

Schloß, S., Schramm, M., Christiansen, H., Scholz, K.-K., Schuh, L. C., Döpfner, M., … Pauli-Pott, U. (2015). Expressed Emotion, Mutter-Kind-Beziehung und ADHS-Symptome im Vorschulalter [Expressed emotion, mother-child relationship, and ADHD symptoms in preschool – a study on the validity of the German Preschool Five Minute Speech Sample]. *Zeitschrift Für Kinder- und Jugendpsychiatrie und Psychotherapie*, *43*(6), 425–431. https://doi.org/10.1024/1422-4917/a000384

Schmid, J., Stadler, G., Dirk, J., Fiege, C., & Gawrilow, C. (2016). ADHD symptoms in adolescents everyday life: fluctuations and symptom structure within and between individuals. *Journal of Attention Disorders*. https://doi.org/10.1177/1087054716629214

Schneider, S., Unnewehr, S., & Margraf, J. (2009). *Kinder-DIPS. Diagnostisches Interview bei psychischen Störungen im Kindes- und Jugendalter [Diagnostic interview for psychiatric disorders in children and adolescents]*. Berlin: Springer.

Scholtens, S., Rydell, A. M., & Yang-Wallentin, F. (2013). ADHD symptoms, academic achievement, self-perception of academic competence and future orientation: A longitudinal study. *Scandinavian Journal of Psychology*, *54*(3), 205–212. https://doi.org/10.1111/sjop.12042

Schuh, L. (2015). Expressed Emotion in Familien von Kindern mit Störung des Sozialverhaltens [Expressed Emotion in families of children with conduct disorder]. *Dissertation, Medizinische Fakultät Der Universität Zu Köln [Dissertation, Medical Faculty of the University of Cologne]*. https://doi.org/10.4126/FRL01-006399218.Gedruckt

Schwarz, N., Knäuper, B., Oyserman, D., & Stich, C. (2009). The psychology of asking questions. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 18–34). New York, NY: Psychology Press.

Seongho, K. (2015). ppcor Partial and Semi-Partial (Part) Correlation Version 1.1. Retrieved from https://cran.r-project.org/web/packages/ppcor/ppcor.pdf

Shoda, Y., Mischel, W., & Peake, P. K. (1990). Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions. *Developmental Psychology*, *26*(6), 978–986.

https://doi.org/10.1037/0012-1649.26.6.978

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–8. https://doi.org/10.1037/0033-2909.86.2.420

Sjöwall, D., Roth, L., Lindqvist, S., & Thorell, L. B. (2013). Multiple deficits in ADHD: Executive dysfunction, delay aversion, reaction time variability, and emotional deficits. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *54*(6), 619–627. https://doi.org/10.1111/jcpp.12006

Skinner, B. F. (1966). What is the experimental analysis of behavior? *Journal of the Experimental Analysis of Behavior*, *9*(3), 213–218. https://doi.org/10.1901/jeab.1966.9-213

Solanto, M. V, Abikoff, H., Sonuga-Barke, E., Schachar, R., Logan, G. D., Wigal, T., … Turkel, E. (2001). The ecological validity of delay aversion and response inhibition as measures of impulsivity in AD/HD: a supplement to the NIMH multimodal treatment study of AD/HD. *Journal of Abnormal Child Psychology*, *29*(3), 215–228. https://doi.org/10.1023/A:1010329714819

Sonuga-Barke, E. J. . (2002). Psychological heterogeneity in AD/HD—a dual pathway model of behaviour and cognition. *Behavioural Brain Research*, *130*(1–2), 29–36. https://doi.org/10.1016/S0166-4328(01)00432-6

Sonuga-Barke, E. J. S. (2003). The dual pathway model of AD/HD: An elaboration of neuro-developmental characteristics. *Neuroscience and Biobehavioral Reviews*, *27*(7), 593–604. https://doi.org/10.1016/j.neubiorev.2003.08.005

Sonuga-Barke, E. J. S., Bitsakou, P., & Thompson, M. (2010). Beyond the dual pathway model: evidence for the dissociation of timing, inhibitory, and delay-related impairments in attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*(4), 345–355. https://doi.org/10.1016/j.jaac.2009.12.018

Sonuga-Barke, E. J. S., Brandeis, D., Cortese, S., Daley, D., Ferrin, M., Holtmann, M., … Sergeant, J. (2013). Nonpharmacological interventions for ADHD: systematic review and meta-analyses of randomized controlled trials of dietary and psychological treatments. *American Journal of Psychiatry*, *170*(3), 275–289. https://doi.org/10.1176/appi.ajp.2012.12070991

Steiner, N. J., Sidhu, T., Rene, K., Tomasetti, K., Frenette, E., & Brennan, R. T. (2013). Development and testing of a direct observation code training protocol for elementary aged students with attention deficit/hyperactivity disorder. *Educational Assessment, Evaluation and Accountability*, *25*(4), 281–302. https://doi.org/10.1007/s11092-013-9166-x

Steyer, R., Ferring, D., & Schmitt, M. J. (1992). States and traits in psychological assessment. *European Journal of Psychological Assessment*, *8*(2), 79–98.

Stubbe, D., Zahner, G., Goldstein, M., & Leckman, J. (1993). Diagnostic specificity of a brief measure of Expressed Emotion: a community study of children. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *34*(2), 139–54. https://doi.org/10.1111/j.1469-7610.1993.tb00976.x

Sturge-Apple, M. L., Suor, J. H., Davies, P. T., Cicchetti, D., Skibo, M. A., & Rogosch, F. A. (2016). Vagal tone and children's delay of gratification. *Psychological Science*, *27*(6), 885–893. https://doi.org/10.1177/0956797616640269

Sudman, S., Bradburn, N., & Schwarz, N. (1996). Answering a survey question: Cognitive and Communicative Processes. In *Asking about answers: The application of cognitive processes to survey methodology* (pp. 55–79). San Francisco: Jossey-Bass.

Tamm, L., Narad, M. E., Antonini, T. N., O'Brien, K. M., Hawk, L. W., & Epstein, J. N. (2012). Reaction time variability in ADHD: a review. *Neurotherapeutics*, *9*(3), 500–508. https://doi.org/10.1007/s13311-012-0138-5

Tannock, R. (2013). Rethinking ADHD and LD in DSM-5: Proposed changes in diagnostic criteria. *Journal of Learning Disabilities*, *46*(1), 5–25. https://doi.org/10.1177/0022219412464341

Theule, J., Wiener, J., Tannock, R., & Jenkins, J. M. (2013). Parenting stress in families of children with ADHD: a meta-analysis. *Journal of Emotional and Behavioral Disorders*, *21*(1), 3–17. https://doi.org/10.1177/1063426610387433

Thomas, R., Sanders, S., Doust, J., Beller, E., & Glasziou, P. (2015). Prevalence of attention-deficit/hyperactivity disorder: A systematic review and meta-analysis. *Pediatrics*, *135*(4), 994–1001. https://doi.org/10.1542/peds.2014-3482

Thome, J., Ehlis, A.-C., Fallgatter, A. J., Krauel, K., Lange, K. W., Riederer, P., … Gerlach, M. (2012). Biomarkers for attention-deficit/hyperactivity disorder (ADHD). A consensus report of the WFSBP task force on biological markers and the World Federation of ADHD. *The World Journal of Biological Psychiatry*, *13*(5), 379–400. https://doi.org/10.3109/15622975.2012.690535

Thorell, L. B., Sjöwall, D., Diamatopoulou, S., Rydell, A.-M., & Bohlin, G. (2016). Emotional functioning, ADHD symptoms, and peer problems: A longitudinal investigation of children age 6-9.5 years. *Infant and Child Development*, 1–13. https://doi.org/10.1002/icd.2008

Tosto, M. G., Momi, S. K., Asherson, P., & Malki, K. (2015). A systematic review of attention deficit hyperactivity disorder (ADHD) and mathematical ability: current findings and future implications. *BMC Medicine*, *13*(1), 204. https://doi.org/10.1186/s12916-015-0414-4

Tymms, P., & Merrell, C. (2011). ADHD and academic attainment: Is there an advantage in impulsivity? *Learning and Individual Differences*, *21*(6), 753–758. https://doi.org/10.1016/j.lindif.2011.07.014

Van der Meere, J. (2002). The role of attention. In S. Sandberg (Ed.), *Monographs on Child and Adolescent Psychiatry: Hyperactivity Disorders.* (2nd ed.). Cambridge: Cambridge University Press.

Van der Meere, J. (2005). State regulation and attention deficit hyperactivity disorder. In *Attention Deficit Hyperactivity Disorder* (pp. 413–433). Totowa, NJ: Humana Press. https://doi.org/10.1385/1-59259-891-9:413

Van Humbeeck, G., Van Audenhove, C., De Hert, M., Pieters, G., & Storms, G. (2002). Expressed emotion: A review of assessment instruments. *Clinical Psychology*

Review, *22*(3), 321–341. https://doi.org/10.1016/S0272-7358(01)00098-8

Vaughn, B. E., Kopp, C. B., Krakow, J. B., Johnson, K., & Schwartz, S. S. (1986). Process analyses of the behavior of very young children in delay tasks. *Developmental Psychology*, *22*(6), 752–759. https://doi.org/10.1037/0012-1649.22.6.752

Volpe, R. J., DiPerna, J. C., Hintze, J. M., & Shapiro, E. S. (2005). Observing students in classroom settings: A review of seven coding schemes. *School Psychology Review*, *34*(4), 454–474.

Wåhlstedt, C. (2009). Neuropsychological deficits in relation to symptoms of ADHD: independent contributions and interactions. *Child Neuropsychology : A Journal on Normal and Abnormal Development in Childhood and Adolescence*, *15*(3), 262–79. https://doi.org/10.1080/09297040802524198

Wakschlag, L., Briggs-Gowan, M. J., Hill, C., Danis, B., Leventhal, B. L., Keenan, K., … Carter, A. S. (2008). Observational assessment of preschool disruptive behavior, part II: validity of the disruptive behavior diagnostic observation schedule (DB-DOS). *Journal of the American Academy of Child & Adolescent Psychiatry*, *47*(6), 632–641. https://doi.org/10.1097/CHI.0b013e31816c5c10

Wakschlag, L., Hill, C., Carter, A. S., Danis, B., Egger, H. L., Keenan, K., … Briggs-Gowan, M. J. (2008). Observational assessment of preschool disruptive behavior, part I: reliability of the disruptive behavior diagnostic observation schedule (DB-DOS). *Journal of the American Academy of Child & Adolescent Psychiatry*, *47*(6), 622–631. https://doi.org/10.1097/CHI.0b013e31816c5bdb

Watson, J. B. (1994). Psychology as the behaviorist views it. *Psychological Review*, *101*(2), 248–253. https://doi.org/10.1037/0033-295X.101.2.248

Weiß, R. H. (2006). *Grundintelligenztest Skala 2 Revision (CFT 20-R) [Culture Fair Intelligence Test (CFT 20-R)]*. Göttingen: Hogrefe.

Whalen, C. K., Henker, B., Ishikawa, S. S., Floro, J. N., Emmerson, N. A., Johnston, J. A., & Swindle, R. (2009). ADHD and anger contexts: Electronic diary mood reports from mothers and children. *Journal of Pediatric Psychology*, *34*(9), 940–953. https://doi.org/10.1093/jpepsy/jsn138

WHO. (1993). *The ICD-10 classification of mental and behavioural disorders: Diagnostic criteria for research*. Geneva: Author.

WHO. (2011). *Internationale Klassifikation psychischer Störungen : ICD-10 Kapitel V(F). Klinisch-diagnostische Leitlinien.[International classification of mental disorders. ICD-10, chapter V (F): Clinical descriptions and diagnostic guidelines]*. (H. Dilling, W. Mombour, M. H. Schmidt, E. Schulte-Markwort, & H. Remschmidt, Eds.) (8th ed.). Bern.

Willcutt, E. G. (2012). The prevalence of DSM-IV attention-deficit/hyperactivity disorder: A meta-analytic review. *Neurotherapeutics*, *9*(3), 490–499. https://doi.org/10.1007/s13311-012-0135-8

Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., & Pennington, B. F. (2005). Validity of the executive function theory of attention-deficit/hyperactivity disorder: a meta-analytic review. *Biological Psychiatry*, *57*, 1336–1346.

https://doi.org/10.1016/j.biopsych.2005.02.006

Wirtz, M. A., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen [Inter-rater agreement and inter-rater reliablity: methods for assessmet and* . Göttingen: Hogrefe.

Zhou, K., Dempfle, A., Arcos-Burgos, M., Bakker, S. C., Banaschewski, T., Biederman, J., … Asherson, P. (2008). Meta-analysis of genome-wide linkage scans of attention deficit hyperactivity disorder. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, *147*(8), 1392–1398. https://doi.org/10.1002/ajmg.b.30878

Zobel, A., & Maier, W. (2004). Endophänotypen - ein neues Konzept zur biologischen Charakterisierung psychischer Störungen [Endophenotypes - a new concept for biological characterization of psychological disorders]. *Der Nervenarzt*, *75*(3), 205–214. https://doi.org/10.1007/s00115-003-1642-y

Eigenleistungserklärung

Alle Texte dieser Dissertation wurden von Frau Keri Hartman in Bezug auf Sprache korrigiert.

Studie 1 und 3

Die in Studie 1 und 3 verwendeten Daten stammen aus einer Studie und derselben Datenerhebung.

Finanzierungsantrag bei der Graduiertenschule LEAD

- Verfassen des Antrags: Leona Hellwig, in Rücksprache mit Caterina Gawrilow, Johanna Schmid, Ute Dürrwächter und Tobias Renner
- Revision des Antrags: Leona Hellwig, in Rücksprache mit Caterina Gawrilow, Johanna Schmid, Ute Dürrwächter und Tobias Renner

Studienplanung

- Auswahl und Entwicklung der Erhebungsinstrumente: Leona Hellwig in Rücksprache mit Caterina Gawrilow und Johanna Schmid
- Entwicklung und Pilotierung des Studienaufbaus und Auswahl der technischen Geräte: Leona Hellwig in Rücksprache mit Caterina Gawrilow und Johanna Schmid

Studiengenehmigungen

- Verfassen, Revision und Einreichung des Ethikantrags: Leona Hellwig in Rücksprache mit Caterina Gawrilow und Johanna Schmid
- Verfassen und Einreichung des Antrags zur Genehmigung der Rekrutierung an Schulen beim Regierungspräsidium Tübingen: Leona Hellwig in Rücksprache mit Caterina Gawrilow und Johanna Schmid

Rekrutierung und Datenerhebung

Leona Hellwig mit Unterstützung von wiss. Hilfskräften

Entwicklung und Training der Ratingschemata

- Studie 1: Leona Hellwig in Absprache mit Caterina Gawrilow und Johanna Schmid und Abschlussarbeitskandidatinnen
- Studie 3: Leona Hellwig in Absprache mit Caterina Gawrilow und Johanna Schmid und wiss. Hilfskräften

Aufbereitung von Audio- und Videodaten

Leona Hellwig mit Unterstützung von Abschlussarbeitskandidatinnen und wiss. Hilfskräften

Rating

- Studie 1: Abschlussarbeitskandidatinnen
- Studie 3: Abschlussarbeitskandidatinnen und wiss. Hilfskräfte

Datenanalyse

Leona Hellwig in Rücksprache mit Caterina Gawrilow und Johanna Schmid

Verfassen der Manuskripte

- Anfertigen eines Entwurfs: Leona Hellwig
- Revisionen: Caterina Gawrilow und Johanna Schmid
- Einarbeitung der Revisionen: Leona Hellwig

Studie 2

Die Daten der Studie 2 stammen aus dem Projekt GIDeCA (IDEA Center Indivdual Development and Adaptive Education, 2011). Die ProjektleiterInnen Caterina Gawrilow, Wolfgang Rauch, und Christine Freitag waren für Finanzierung, Studienplanung, Studiengenehmigungen, sowie Rekrutierung und Datenerhebung verantwortlich, ohne Beteiligung von Leona Hellwig.

Konzeption und Auswahl der Fragestellung

Caterina Gawrilow, Jan Kühnhausen und Leona Hellwig

Aufbereitung von Audio- und Videodaten

Leona Hellwig mit Unterstützung von Abschlussarbeitskandidatinnen

Entwicklung und Pilotierung des Ratingschemas

Leona Hellwig in Rücksprache mit Jan Kühnhausen

Ratertraining

Leona Hellwig mit Jan Kühnhausen

Rating

Wiss. Hilfskräfte und Abschlussarbeitskandidatinnen

Datenanalyse

Leona Hellwig in Absprache mit Jan Kühnhausen und Caterina Gawrilow

Verfassen des Manuskripts

- Anfertigen eines Entwurfs: Leona Hellwig
- Revisionen: Jan Kühnhausen, Tilman Reinelt, Andrea Wirth, Christine Freitag, Wolfgang Rauch, Christina Schwenck, Caterina Gawrilow
- Einarbeitung der Revisionen: Leona Hellwig