(New) Approaches and Challenges in Measuring Self-Regulation

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

M.Sc. Lilly Buhr

aus Uelzen

Tübingen

2023

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

Tag der mündlichen Qualifikation:	24.04.2023
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatterin:	Prof. Dr. Caterina Gawrilow
2. Berichterstatter:	Prof. Dr. Tobias Renner

Danksagungen

Ich bin unglaublich dankbar für das wunderbare Netz aus großartigen Menschen, die mich in den letzten Jahren begleitet haben und mir das Gefühl gegeben haben, nie allein zu sein!

Vielen Dank Prof. Dr. Caterina Gawrilow für die Ermöglichung dieser besonderen Zeit, für die Freiheit eigene Wege zu gehen und für die Ermutigung auch Herausforderungen zu meistern.

Danke an meine Kolleg*innen (oder doch eher Freund*innen?!), an Dr. Ulrike Schwarz, Anne Eppinger, Tomasz Moschko, und Julia Steigleder für die gegenseitige Selbstregulation, für das Nutzen von Synergieeffekten, für FuFs und für das gemeinsame Feiern von Erfolgen.

Danke den ehemaligen und aktuellen Mitarbeitenden des Arbeitsbereichs Schulpsychologie, sowie den Mitgliedern der Forschungsnetzwerke LEAD und IDeA für Kooperationen, Unterstützung und Austausch.

Danke an die Co-Autor*innen der drei in der Dissertation verwendeten Manuskripte, Prof. Dr. Florian Schmiedek, Dr. Jan Kühnhausen und Finja Greiner für die produktive und hilfreiche Zusammenarbeit.

Danke an Prof. Dr. Tobias Renner für die kurzfristige Bereitschaft, die Begutachtung meiner Arbeit zu übernehmen.

Danke an alle Studierenden, mit denen ich über die Jahre durch Lehre, Abschlussarbeiten, Praktika und Hiwi-Zeiten zusammenarbeiten durfte und die mich immer wieder daran erinnert haben, wie schön es sein kann die Freude an seiner Arbeit weitergeben zu dürfen.

Danke an alle meine Freundinnen und Freunde, egal wie weit entfernt sie sind, fürs Zuhören und Ablenken, fürs Mitfiebern und Mitfeiern.

Zu guter Letzt ein riesiges Danke an das Fundament, auf das ich alles bauen kann, meine gesamte Familie, insbesondere an meine Eltern und meine große kleine Schwester, für das wahnsinnige Vertrauen, das ihr in mein Können setzt und die Unterstützung, wann immer ich sie brauche.

Content

List of Tables

List of Figures

Abbreviations

ADHD – Attention-Deficit/Hyperactivity Disorder

CAARS – Conners Adult ADHD Rating Scales

DSM – Diagnostic and Statistical Manual of Mental Disorders

HASE – Homburger ADHS Skalen für Erwachsene

ICC – Intraclass Correlation Coefficient

ICD – International Statistical Classification of Diseases and Related Health Problems

ROC-analysis – Receiver Operating Characteristic Analysis

SCS – Self-control scale

SWAN – Strengths and Weaknesses of ADHD and Normal Behavior

WHO – World Health Organization

Abstract

Research and clinical practice have long identified self-regulation, the ability to work towards a specific goal and ignoring internal as well as external distractions, as an important precursor for positive life outcomes, like vocational success, favorable health behavior, and better social relationships. However, the measurement methods to assess individuals' self-regulation vary widely. Within this dissertation, I am examining different approaches in the measurement of self-regulation: depicting interindividual differences and intraindividual fluctuations via ambulatory assessment, investigating associations with internal and external factors, in this case sleep and executive functions, using the dimensional perspective to assess strengths and weaknesses, and comparing self- and observer reports to understand differences between assessments. In the first manuscript we investigated 70 schoolchildren aged 10-12 years on their self-regulation ability and their night sleep as well as their daytime sleepiness via ambulatory assessment (i.e., repeated measurement on technical devices) in their daily life. The second manuscript is based on the same sample and examines more closely the temporal fluctuations of self-regulation and working-memory performance (i.e., executive functioning), as well as their associations on the inter- and the intraindividual level. The third manuscript investigates self-regulation in 142 adults, using a dimensional questionnaire and comparing self- versus observer reports by a significant other. The results of the presented studies show that all measurement approaches have individual advantages and challenges. Measuring temporal fluctuations of self-regulation helps to differentiate inter- and intraindividual associations with internal and external factors like sleep and executive functions. However, it poses a high burden on the participants and requires considerable resources. Dimensional measurement permits the investigation of strengths and weaknesses within an individual and allows for new research questions but needs more groundwork in the development of valid and reliable scales. Including self- and observer reports into the assessment of self-regulation provides additional information about the target behavior but further research is needed concerning the factors which might influence the differences in reports. In conclusion, the discussed measurement approaches suggest a better understanding of self-regulation

abilities in humans. However, more precise theory and additional research is needed to sufficiently understand how individuals self-regulate their cognition and behavior throughout their life.

German Abstract

(Neue) Ansätze zur Messung von Selbstregulation

Menschen benötigen die Fähigkeit zur Selbstregulation, um ihre Ziele zu erreichen. Selbstregulation steht in Verbindung mit schulischem und beruflichem Erfolg, vorteilhaftem Gesundheitsverhalten, und positiven sozialen Interaktionen. Trotz dieser Erkenntnisse über die Bedeutung der Selbstregulation fehlt bis heute ein Konsens über einschlägige Methoden zur Messung der Selbstregulationsfähigkeit von Individuen. In der vorliegenden Dissertation werden neue und vielversprechende Ansätze zur Messung von Selbstregulation untersucht und diskutiert: Die Untersuchung von zeitlichen Schwankungen in der Selbstregulation durch die Unterscheidung zwischen inter- und intraindividuellen Unterschieden, der Zusammenhang von Selbstregulation mit internalen und externalen Faktoren, Messung von Selbstregulation als dimensionales Konstrukt, und der Vergleich von Selbst- und Fremdbericht über die Selbstregulationsfähigkeit. Das erste Manuskript untersucht die inter- und intraindividuellen Verbindungen von nächtlichem Schlaf sowie Schläfrigkeit am Tag und Selbstregulation im Alltag von 70 Schulkindern im Alter von 10 bis 12 Jahren mithilfe von ambulantem Assessment (d.h., wiederholten Messungen auf Smartphones). Im zweiten Manuskript wird mithilfe der gleichen Stichprobe die zeitliche Fluktuation von Selbstregulation genauer untersucht sowie der Zusammenhang mit dem Arbeitsgedächtnis als exekutive Funktion. Manuskript drei vergleicht anhand einer dimensionalen Selbstregulationsskala den Selbst- mit dem Fremdbericht von 142 Erwachsenen und einer ihnen nahestehenden Person. Die Ergebnisse der Studien zeigen, dass alle vorgestellten Ansätze zur Messung von Selbstregulation Chancen und Herausforderungen aufweisen. Studienprotokolle, welche ein ambulantes Assessment implementieren, ermöglichen die Unterscheidung auf der inter- sowie der intraindividuellen Ebene und erlauben außerdem die Untersuchung von Zusammenhängen mit internalen und externalen Einflussfaktoren, zum Beispiel Schlaf oder Exekutiven Funktionen auf beiden Ebenen. Sie stellen jedoch eine hohe Belastung für die Teilnehmenden dar und erfordern erhebliche Ressourcen. Mithilfe einer dimensionalen Skala können sowohl Stärken als auch Schwächen in der

Selbstregulationsfähigkeit gemessen werden und diese Skalen können somit neue Fragestellungen in der Wissenschaft ermöglichen, es mangelt allerdings noch an ausreichend aussagekräftigen Studien zu Validität und Reliabilität der Skalen. Der Vergleich von Selbst- und Fremdberichten macht es möglich einzuschätzen, inwiefern die untersuchte Person selbst und nahestehende Personen in ihrem Urteil übereinstimmen und welche zusätzlichen Informationen die jeweiligen Beobachter*innen in den Erhebungsprozess einbringen können, auch wenn Ursachen für diese Unterschiede in zukünftigen Studien genauer erforscht werden sollten. Zusammengefasst kann man sagen, dass die vorgestellten Ansätze einen deutlichen Mehrwert in der Messung von Selbstregulation darstellen können. Nichtsdestotrotz benötigen wir Anpassung der Theorien an die neuen Erkenntnisse, sowie zusätzliche empirische Forschung, um ein umfassenderes Verständnis über die Selbstregulation in Individuen zu ermöglichen.

List of Publications

The dissertation is based on the following three manuscripts:

a) Accepted Manuscripts

Buhr, L., Moschko, T., Eppinger Ruiz de Zarate, A., Schwarz, U., Kühnhausen, J., & Gawrilow, C. (2022). The Association of Self-Reported ADHD Symptoms and Sleep in Daily Life of a General Population Sample of School Children: An Inter- and Intraindividual Perspective. *Brain Sciences*, *12*(4), 440. https://doi.org/10.3390/brainsci12040440

b) Submitted Manuscripts

Buhr, L., Greiner, F., Schwarz, U., & Gawrilow, C. (2023). Dimensional Self- versus Observer Report of Inattention, Hyperactivity/Impulsivity, and Emotion Regulation in a General Adult Sample. *Submitted to: Journal of Attention Disorders*.

Buhr, L., Schmiedek, F., & Gawrilow, C. (2023). Temporal Fluctuations and Associations of Self-Regulation Skills and Working-Memory Performance in the Daily Life of Schoolchildren. *Submitted to: Journal of Educational Psychology*.

1. Introduction - (New) Approaches in Measuring Self-Regulation

How do children stay focused in class? How do adults maintain a conversation without interrupting the conversational partner with constantly new thoughts? How can we stay seated when it is socially appropriate without constantly fidgeting and moving around? Psychologists have identified self-regulation as an important answer to all these questions and as an essential ability for individuals. Human beings depend on their individual ability to self-regulate their cognition and behavior in all areas of life as for instance in school and at the workplace, in social interactions, and when working towards a specific goal. High self-regulation is associated with a number of positive life-outcomes, like educational success, better social relationships and healthy behaviors across the life span (Barkley, 2002; Caye et al., 2016; Kuriyan et al., 2013; Loe & Feldman, 2007; Sciberras et al., 2009). For example, research could show that self-regulation of attention at the age of 4 significantly predicted the odds of successfully finishing college at the age of 25 (McClelland et al., 2013). Furthermore, a meta-analysis found that children who showed higher self-regulation when they were 8 years old had higher academic achievement, less internalizing and externalizing problems, and less substance abuse disorders when they were 13 years old (Robson et al., 2020). On the other hand, problems with self-regulation as for instance manifested in people with self-regulation disorders such as Attention-Deficit/Hyperactivity Disorder (ADHD) have shown correlations with negative life outcomes like school problems, substance abuse, and delinquent behavior (Caye et al., 2016).

In recent years, the relevance of self-regulation abilities in human lives has led researchers to investigate self-regulation even further. Despite the ample amount of research, experts in the field still do not agree on one golden standard to measure self-regulation reliably in various setting requiring the self-regulation of cognition and behavior (Inzlicht et al., 2021). For example, it has been shown that the worldwide prevalence rate of ADHD depends heavily on the measurement technique which is used to identify individuals with and without the disorder (Polanczyk et al., 2014). New technologies as well as new foci in the research community have emerged and created exciting new possibilities for the exploration of the measurement of psychological constructs (e.g., Koch et al., 2021; Swanson et al., 2012). Within my dissertation, I will inspect some promising (new) measurement techniques and core questions on the measurement of self-regulation across the spectrum of high self-regulation on the one side, as well as self-regulation difficulties (i.e., ADHD symptoms) on the other side.

2. Theoretical Background and Previous Research

2.1. Self-Regulation

As described earlier, self-regulation is an important ability for positive life outcomes (Robson et al., 2020). However, to appropriately inform research and clinical practice about the measurement of self-regulation in individuals, we have to define what exactly is meant by the term.

2.1.1. Definition of Self-Regulation

Self-regulation is a diffuse construct which has been defined differently by many researchers (Inzlicht et al., 2021; Nigg, 2017). Within this dissertation, by following Carver and Scheier (2011), I define self-regulation as the ability of an individual to orchestrate their cognition and behavior in order to work towards a specific goal, thereby ignoring distractions and suppressing impulses. Some researchers interpret the term as being synonymous with self-control, which, according to Carver and Scheier (2011) only describes a part of the complex internal adjustments that take place in the self-regulation of an individual. In their definition, self-control describes the process of working against an impulse to reach another goal, while self-regulation is a more complex procedure involving many internal processes monitoring and correcting one’s cognition and behavior to be goal-directed. Therefore, in the following I will treat self-control as a component of self-regulation.

2.1.2. ADHD as Disorder of Self-Regulation

People with a diagnosis of Attention-Deficit/Hyperactivity Disorder (ADHD) are challenged by difficulties in keeping up attention, exceptional high activity and an adversity to suppress their impulses (American Psychiatric Association, 2013). In other words, they have problems in self-regulating their cognition and behavior to lead to a specific goal. A diagnosis of ADHD, according to the guidelines of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5; American Psychiatric Association, 2013) or the *International Statistical Classification of Diseases and Related Health Problems* (ICD-11; World Health Organization [WHO], 2018), is associated with many negative life outcomes. People with an ADHD diagnosis show more vocational problems (i.e., lower educational success, lower income, higher unemployment), adverse health related behavior (i.e., manifested in higher obesity rates or more substance abuse disorders), and worse social relationships (i.e., fewer friendships, lower marital satisfaction) (Caye et al., 2016; Erskine et al., 2016). The etiology of ADHD is widely discussed in the literature and current consensus agrees on a combination of genetical, biological, social, and environmental factors (Kooij et al., 2019). Many psychological theories have identified a deficiency in self-regulation as one of the core underlying processes in the development of ADHD (e.g., Drechsler et al., 2020; Johnson et al., 2009).

In a very early theory, Barkley (1997) described ADHD as a disorder stemming from a deficit in four neuropsychological functions, namely (1) working memory, (2) self-regulation of affect, motivation and arousal, (3) internalization of speech, and (4) reconstitution through behavioral analysis and synthesis. All of these four functions are, according to Barkley, based on the inhibition of behavior. Later, Barkley defines ADHD as a “disorder of self-regulation” (Barkley & Murphy, 2011, p. 559). To integrate this view of ADHD as a self-regulation disorder with disparate theories of the disorder as a motivational style, and explain the heterogeneity of symptoms in individuals with ADHD, Sonuga-Barke (2002) developed the dual-pathway model. In this model, different symptoms stem from deficits on one of the two pathways, the self-regulation or inhibition of cognition and behavior, and the ability to delay rewards. Current models about the etiology of ADHD incorporate a combination of biological, psychological, and social precursors to the development of the core symptoms inattention and hyperactivity/impulsivity (e.g., Döpfner et al., 2020; Nigg et al., 2020). These models integrate the lack self-regulation as an important aspect of cognition and emotion in individuals with ADHD (Nigg et al., 2020). Supporting the hypothesis that self-regulation deficits are an essential underlying component of ADHD, developmental studies have found that early self-regulation deficits in children predict an ADHD diagnosis later in life (Kostyrka-Allchorne et al., 2020).

Due to this theoretical and empirical evidence, I will in the following refer to people with ADHD symptoms or diagnosis as people with low self-regulation ability interchangeably. Whenever possible, I will, however, distinguish whether specific clinical samples or clinical diagnostic scales were used in the described research.

2.2. Measurement of Self-Regulation

Conventional measures of self-regulation comprise questionnaires (e.g., Conners ADHD Rating Scales [CAARS]; Christiansen et al., 2013; self-control scale [SCS]; Rauch et al., 2014; and Homburger ADHS Skalen für Erwachsene [HASE]; Rösler et al., 2008) and laboratory experiments (e.g., delay aversion tasks; Marx et al., 2021). However, researchers have always seen the limitations of these measurement methods and permanently tried to improve their ways to draw conclusions about the internal mechanisms and expression of self-regulation in humans. In the following chapter, I will discuss a fragment of these limitations in more classical measurement methods as well as the efforts that have been made to improve the way of measuring self-regulation.

2.2.1. Inter- versus Intraindividual Differences in Self-Regulation

Conventionally, self-regulation is measured once in a person to inform research and clinical practice. However, the ability to self-regulate cognition and behavior does not only differ between persons, but has been found to also fluctuate within individuals over time and in different contexts (Leonard et al., 2021; Ludwig et al., 2016; Schmid et al., 2016). A person might have moments of high self-regulation and moments with pronounced inattention and hyperactivity/impulsivity. Separating these two levels, namely the interindividual (between-person) level and the intraindividual (within-person) level has theoretical, methodological, and practical implications.

From a theoretical view, it should be distinguished whether self-regulation or ADHD symptoms are defined as trait variables, which might develop during childhood but then stay stable during adulthood and can only be influenced over long timeframes, or whether we interpret self-regulation as a state variable, which can fluctuate highly within individuals. Although these distinctions have to be made carefully, they can also be combined within one overarching theory of trait-like self-regulation which includes long-term developmental changes as well as short-time fluctuations (Nesselroade, 1991). Theories about antecedents, correlations and consequences of self-regulation and ADHD symptoms should therefore carefully define which of the levels they target with their definition.

From a methodological perspective, a rising use of measurement technologies like ambulatory assessment, where data is assessed repeatedly in the daily life of participants via technical devices like smartphones or tablets, provides the necessary information to disentangle inter- from intraindividual effects in self-regulation and ADHD symptoms (Ebner-Priemer et al., 2009; Koch et al., 2021). Statistically, procedures like multilevel modeling approaches help to detect systematic associations between self-regulation and internal or external factors on both levels.

Finally, the differentiation between the inter- and intraindividual level gives rise to practical issues. Associations between two variables might have completely different effects on the two levels, in strength as well as in direction (Molenaar & Campbell, 2009; Wang & Maxwell, 2015). For example, on the interindividual level, children who are more physically active might show higher self-regulation skills but intraindividually, in moments when a child has high physical activation rates, self-regulation might be lower. These two levels should be considered when discussing antecedents, correlates, and consequences of self-regulation and ADHD symptoms in scientific research and clinical practice, for example when planning and evaluating interventions.

2.2.2. Correlates and Antecedents of Self-Regulation

An important advantage of measuring self-regulation intraindividually in the daily life of participants is the possibility to explore the context in which self-regulation is higher or lower. Correlations between the fluctuations of self-regulation and internal or external factors might convey important information about psychological mechanisms through which self-regulation is established.

When taking the temporal order into account, antecedents of self-regulation might provide empirical evidence for theories of causal mechanisms regarding the development of high or low self-regulation. For example, the knowledge about antecedents of self-regulation is especially important in the development of interventions: only if we know what internal and external factors trigger high self-regulation, we can actively try to promote these circumstances (Purdie et al., 2002). For example, if high sleep quality precedes high self-regulation, we might subsequently test whether an intervention on sleep quality also improves self-regulation (Gruber et al., 2012). Additionally, if theory hypothesizes a relationship between self-regulation and executive functions, we should find an association between their indicators on the inter- and/or the intraindividual level (Molenaar & Campbell, 2009). In the following, I am going to explain the hypothesized associations of self-regulation with sleep and executive functions more thoroughly.

2.2.2.1. Sleep.

Humans need sleep for their cognitive and biological functioning (Chokroverty, 2017). Sleep is a fascinating state, where at the same time an individual has highly diminished consciousness but, nevertheless, research can detect very high brain activity during sleeping states. Consequently, sleep is not only considered a resting state for the body and brain, but it is also assumed to be important for the restoration of body and brain tissue and for the consolidation of memory (Chokroverty, 2017). Researching sleep is challenging, because individuals might lack the introspection to reliably inform about their internal states during sleep, but there is evidence that during sleep the prerequisites for emotion regulation and cognitive functioning are built (Vriend et al., 2012). According to the state regulation model, sleep is a necessary foundation for children to be able to regulate their arousal and activation (Van der Meere, 2005). The lack of sleep might therefore lead to the inability to self-regulate cognition and behavior, represented in high inattention and hyperactivity/impulsivity.

Scientific research has tried to find evidence for the relationship between sleep and ADHD symptoms through two modes of investigation: correlational interindividual comparisons, and experimental studies for the assessment of intraindividual effects. The first mode includes scientific research, where symptoms of ADHD and sleep is assessed once (Becker, 2020). Studies of this category found for example that individuals with a diagnosis of ADHD have higher onset latency of sleep (i.e., the time it takes to fall asleep) and lower efficiency of sleep than individuals without an ADHD diagnosis (De Crescenzo et al., 2014). Additionally, they found that adolescents with ADHD reported insufficient sleep and even fell asleep in class more often than adolescents without ADHD (Becker et al., 2019). However, these studies are only correlational, and they do not inform about the causal relationship within the association, which might point into different directions. The state regulation theory explained above assumes that bad sleep leads to lower self-regulation and therefore more ADHD symptoms, but it is also possible that the symptoms of ADHD cause problems in sleeping (e.g., hyperactivity hinders the necessary relaxation of the body) or ADHD related medication could interfere with normal sleeping patterns (Becker, 2020; Gregory et al., 2017; Gregory & Sadeh, 2012; Hvolby, 2015; Van Der Heijden et al., 2005).

To investigate such causal relationships between ADHD symptoms and sleep more closely, experimental studies were conducted with either short term interventions (one night/few nights of restricted or prolonged sleep) or long-term behavioral trainings for parents. When 7-11-year-old children were restricted in their sleep by one hour, they showed significantly lower self-regulatory behavior the following day. Conversely, after a night of more than usual sleep, they could self-regulate their cognition and behavior better (Gruber et al., 2012). Another study found that parent behavior training instructing parents about sleep hygiene procedures and assisting them in the implementation of evening routines reduced ADHD symptoms half a year later (Hiscock et al., 2015). Although it can be seen above that the association of self-regulation and sleep is already extensively examined on the inter- as well as on the intraindividual level, these studies have investigated only one of the two levels. Additionally, the intraindividual differences have only been investigated in experiments where sleep restriction has been implemented artificially. No study has yet investigated how the natural fluctuations of sleep might influence self-regulation abilities the following day. This, however, seems very important because these fluctuations have a high impact on daily functioning and therefore, for example, on school success and cognitive functioning (Dewald et al., 2010; Könen et al., 2015).

Next to the quality and quantity of night sleep, the feeling of sleepiness during the day might also influence self-regulation ability in individuals. When children feel tired, their behaviour sometimes includes symptoms of inattention and even hyperactivity/impulsivity (Owens et al., 2012). This might equally be explained by the state regulation theory, where the internal evaluation mechanism detects a state of underarousal due to sleepiness and consequently enhances activity, which might lead to symptoms of inattention and hyperactivity/impulsivity (Van der Meere, 2005). This daytime sleepiness is highly related to school performance, even higher than sleep quality and sleep duration during the night, and should thus be investigated more thoroughly (Dewald et al., 2010).

In manuscript one, we therefore investigated the association of two different sleep indicators (night sleep and daytime sleepiness) and self-reported inattention and hyperactivity/impulsivity on the inter- and intraindividual level in the daily lives of 70 German schoolchildren for three bursts of 18 days.

2.2.2.2. Executive Functions and Working-Memory Performance.

When individuals have to carry out a complex task, they need elaborate cognitive functions. Those are generally summarized under the umbrella term of executive functions, with the three underlying constructs shifting, inhibition, and updating (Miyake & Friedman, 2012). Mental set-shifting refers to the ability to switch between tasks or vary between mental sets. When inhibiting, a person suppresses an impulsive response to reach some previously defined goal. The concept of information updating describes the ability to store and at the same time alter knowledge within one's working memory. All these constructs were shown to be clearly separated in a confirmative factor analysis, but were still linked through a unifying underlying executive functions factor (Miyake et al., 2000).

A relationship of self-regulation and ADHD symptoms has been hypothesized with executive functions. This relationship is supposed to be bidirectional, with the development of executive functions building on self-regulation abilities, but on the other hand self-regulation being dependent on executive function abilities (Blair & Ursache, 2011). Core regulatory mechanisms like self-regulation are hypothesized to develop through core processes in executive functions (Bailey & Jones, 2019; Schmidt et al., 2022). Building on this theoretical background, empirical research on the association of self-regulation with executive functions has discovered that individuals with an ADHD diagnosis performed worse in executive function tasks, compared to healthy controls (Alderson et al., 2013; Campez et al., 2020; Willcutt et al., 2005). Nevertheless, the effect sizes of the group differences were only medium and some executive function tasks did not show any association to ADHD symptoms. One of the tasks which had been repeatedly used in empirical studies to investigate about executive functions examines spatial working memory updating performance. Working-memory updating is for example needed when a person is doing mental math. The person has to keep all necessary information in mind and at the same time update it, to come to the correct result. Research has shown that individuals with better self-regulation skills also seem to show better working-memory performance (Willcutt et al., 2005). In the following, I will focus on scientific studies investigating working-memory performance as a measure of executive function and its association with self-regulation.

Most previous research on the association of self-regulation and working-memory performance has been conducted on the interindividual level. Thus, the studies hypothesized that children who depict higher self-regulation on the trait level would also demonstrate higher working-memory performance. Theory, however, assumes that working memory might support self-regulation more on the intraindividual level, where moments of high working memory would be linked with moments of high self-regulation within a person (Hofmann et al., 2012). Hofmann and colleagues (2012) interpret self-regulation as goal directed behavior and explain the underlying mechanisms in the support from working memory in representing the goal more actively, concealing interference with the goal, suppression of hindering thoughts that do not lead towards the goal and controlling unwanted emotions. Another theory assumes that self-regulation and working memory both rely on the active control of attention and due to this dependency on the same resource, the intraindividual processes are associated (Ilkowska & Engle, 2010b).

Until now, to my knowledge only one study has investigated the intraindividual relationship between self-regulation and working-memory performance. An ambulatory assessment study examining 9-11-year-old students daily over a period of 4 weeks has found that self-regulation and working-memory performance independently are associated with different academic success measures, but not associated with each other on the inter- or the intraindividual level (Blume et al., 2022). However, their measure of self-regulation lacked reliability since they investigated it with only one item.

To further understand the association of self-regulation and working-memory performance on the inter- as well as on the intraindividual level, we investigated data from an ambulatory assessment study of 70 German schoolchildren who indicated their self-regulation ability and conducted a working memory updating task three times a day for three bursts of 18 days in manuscript two.

2.2.3. Dimensional versus Categorical Perspective

Originally, psychologists defined (psychological) disorders as being categorical in nature. For example, patients and research participants are either classified as *with ADHD* or *without ADHD*. For some time, however, theories have been developed that identify behaviors and perceptions related to certain psychological disorders as extreme expressions on a scale ranging from very low to very high presentations. In other words, in the dimensional view, "disorder and normality differ only in degree, but not kind" of symptoms (Coghill & Sonuga-Barke, 2012, p. 469). With regard to ADHD, research supports this theoretical view, since studies repeatedly find empirical evidence for a dimensional structure of the symptoms inattention and hyperactivity/impulsivity (Bitto et al., 2017; Frazier et al., 2007; Haslam et al., 2006; Larsson et al., 2012; Marcus & Barry, 2011). This implies the notion that all humans depict an individual capacity for self-regulating their cognition and behavior. People with a diagnosis of ADHD can be ranked at the lower end of the scale.

Despite this evidence for a dimensional structure of ADHD symptoms in theory and research, until recently almost all scales assessing ADHD symptoms in research and clinical practice were categorical in nature (e.g., CAARS; Christiansen et al., 2013; HASE; Rösler et al., 2008). This circumstance can be explained by several reasons. First, cutoff scores are the most important output of ADHD scales for clinical practice (Coghill & Sonuga-Barke, 2012). Clinicians usually need to decide whether a patient is eligible for a diagnosis and therefore for clinical support. A precise cutoff score within a questionnaire helps to make this decision objectively, reliably and with high validity. The diagnosis can then also be the source of emotional relief for patients themselves, as well as parents and teachers (Young et al., 2008). In addition, researchers use categorical means to compare groups of people with and without ADHD diagnoses (Bitto et al., 2017).

Dimensional scales, on the other hand, have some advantages over the conventional categorical scales in measuring self-regulation. By integrating both extremes of a characteristic, they can capture strengths in self-regulation as well as weaknesses of an individual. Additionally, they usually capture higher variance between participants and can therefore identify at-risk patients before they might be identified by cutoff scores for a disorder (Frazier et al., 2007). This higher variance also makes it possible to depict small changes of symptoms within participants, for example because of a normal developmental trajectory or because of treatment effects that do not change the category from *with ADHD* to *without ADHD* yet (Marcus & Barry, 2011). Dimensional scales do not only show advantages in clinical practice but also for research purposes. While categorical scales depict a skewed distribution in the general population, with most participants scoring in the *no symptoms* area, dimensional scales should result in a normality distribution, which is a necessary preliminary assumption for many statistical tests (Polderman et al., 2007). Additionally, these dimensional scales allow for not only comparing clinical with healthy samples, but are also able to depict variance between healthy participants in population-based studies (Arnett et al., 2013).

Given all these benefits of dimensional scales, as well as the empirical evidence for a normal distribution of self-regulation in the general population, it is surprising that only one dimensional scale for capturing ADHD symptoms has been developed so far, the *Strengths and Weaknesses of ADHD and Normal behavior* (SWAN) scales (Swanson et al., 2012). The SWAN scales contain 18 neutrally formulated statements based on the ADHD symptom list in the DSM-5 (American Psychiatric Association, 2013). The self-regulation of cognition and behavior of a child can then be rated by parents or teachers on a 7-point Likert scale from *far below average* to *far above average* in comparison to other children at the same age. The SWAN scales have been translated into several languages, for example Spanish (Lakes et al., 2012), French (Robaey et al., 2007), Chinese (Chan et al., 2014), and German (Schulz-Zhecheva et al., 2017). Since ADHD can also be present in adults (Caye et al., 2016), we developed a German self-report version for adults by adapting the original questionnaire to the living conditions of adults (Blume et al., 2020). All scientific studies examining the different versions of the SWAN scales confirm their high validity and reliability (e.g., Blume et al., 2020; Lakes et al., 2012; Schulz-Zhecheva et al., 2017). As further evidence for the dimensional view on ADHD symptoms, the data collected through the scales proofs to be normally distributed in general population samples (e.g., Arnett et al., 2013; Polderman et al., 2007).

Next to the symptoms of inattention and hyperactivity/impulsivity, people with ADHD often also experience difficulties in emotion regulation (Retz et al., 2012). Although emotion regulation is currently not described as a core symptom in the DSM-5 (American Psychiatric Association, 2013), some categorical ADHD questionnaires examine these difficulties nevertheless in the assessment process to receive a comprehensive picture of the individual (e.g., CAARS; Christiansen et al., 2013). The mechanism underlying the association of inattention and hyperactivity/impulsivity with emotion regulation has, however, not been exhaustively studied yet (Shaw et al., 2014). Dimensional emotion regulation scales might help to better understand the interrelation of ADHD symptoms and emotion regulation in the general population as well as in clinical samples.

Taken together, we conclude the overwhelming evidence and advantages for the use of dimensional scales to measure ADHD symptoms in research and clinical practice. Although categorical scales can be helpful in specific circumstances, much valuable information about smaller differences is lost by categorizing individuals in few groups. Nevertheless, additional potential measurement practices have yet to be investigated. Until now, the SWAN scales have been used for parent- or teacher report of children and self-report for adults. I argue that we need to better understand the output of the scales by testing them, for example, in an observer-report for adults.

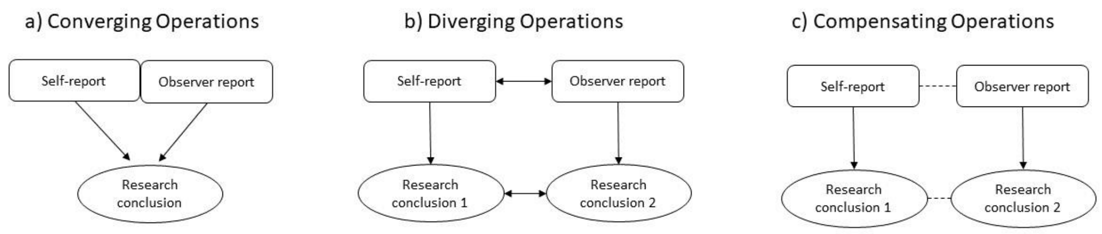
2.2.4. Self-report versus Observer Report

The integration of several perspectives for the diagnostic process of ADHD in adults was noted as an important aspect in a recent consensus statement (Kooij et al., 2019). Possible raters are the target person themselves, a clinician, a close significant other (i.e., a partner, friend, or parent), a teacher, or a co-worker. When both the target person and a significant other agree on the severity of the symptoms, this provides meaningful indication for an ADHD diagnosis. However, it is repeatedly found that self- and observer-reports about the self-regulation capacity of an individual might differ (e.g., Kooij et al., 2008; Van Voorhees et al., 2011; Zucker et al., 2002). Until now, there has not been a consensus about the question which perspective should be integrated in in research questions or the diagnostic process. Similarly, when multiple reports are collected, there is no consensus on how to integrate these different perspectives.

In their Operations Triad Model, De Los Reyes and colleagues (2013) try to formulate guidelines on the integration of different perspectives and distinguish three types of information: converging, diverging, and compensating (Figure 1). When a multiple informant process results in converging information, the informers agree in their ratings and a conclusion can easily be made. In the case that information from observers differs, there are two possibilities to interpret this difference. On the one side, we can see information as diverging, that is, the difference can be explained by empirically measurable characteristics of the observers and/or the context. This means that we can use both ratings as providing important individual information to our conclusion. On the other side, when we doubt the credibility of one of the observers, we can use the information from the second observer as compensating information. In this case, only the information of the second observer should be used to reach a conclusion.

Figure 1

Operations Triad Model (adapted from De Los Reyes, 2013)



The distinction between diverging and compensating information has been further specified by Alexander and colleagues (2017). According to them, differences in information can be accounted to context and insight. Context describes the fact that individual behavior might not always be the same in varying settings, which might lead to diverging observations by different observers. For example, a good friend seeing the target person usually in active sports courses might not be able to detect hyperactivity like a colleague, who experiences the same person in a quiet office. Both observations incorporate important information that can be used in research or clinical processes. A different proceeding is necessary when it comes to a lack of insight for one of the observers. A lack of insight might apply when an external observer is asked to judge about internal processes of a target person or when a target person themselves does (for example due to a psychological disorder) not have the ability to recognize important cues. For example, there has been empirical evidence that individuals with ADHD might lack the insight in their own symptoms (Smith et al., 2000).

Previous research has compared information on ADHD symptoms by different observers, for example comparing adolescents with their parents, self-reports with clinicians, and self-reports with significant others. They usually find a moderate but significant correlation between self- and observer-report. Additionally, it seems that the self-report is usually more negative (i.e., reporting more/stronger ADHD symptoms) than the observer report, indicating that target persons seem to have more insight in their inattention and hyperactivity/impulsivity than significant others (e.g., Kooij et al., 2008; Van Voorhees et al., 2011; Zucker et al., 2002). Some research has also tried to examine what internal and external factors might account for differences in observer reports (i.e., gender, age, relationship of target person and observer). However, these individual influences are not yet understood to the full extent.

The above-described research gives some interesting insight into the different information that individual observers provide. However, all these studies have been conducted with help of categorical ADHD scales, comparing decisions on group level (e.g., indicating a diagnosis or not) and investigating exclusively the weaknesses in self-regulation. Additional studies using dimensional scales allow for very different research questions. Initial studies have been conducted, comparing parent and teacher report of ADHD symptoms in children with help of the SWAN scales (Gooch et al., 2017; Jungersen & Lonigan, 2021). For example, measurement invariance analyses found that parents and teachers interpreted the items on the SWAN scales differently, which might explain the divergence in the ratings (Jungersen & Lonigan, 2021). Another explanation for differences in parent and teacher ratings are child characteristics, where evidence is found that teachers rate children with low language ability higher on inattention than parents (Gooch et al., 2017).

In conclusion, dimensional scales might help to better understand the differences in ratings by individual observers. However, to my understanding there has yet not been any research on observer differences on dimensional scales of ADHD symptoms with an adult sample. In manuscript three, we therefore compared self- and observer ratings assessed with the dimensional SWAN-DE scales in 142 adult dyads.

3. Aim of the Dissertation

Humans rely on their individual ability in self-regulation to successfully reach their goals. To foster self-regulation, we have to understand the underlying mechanisms and internal structure of self-regulation within individuals. Innovative and promising approaches to the measurement of self-regulation have been developed recently, for example ambulatory assessment for investigations in daily life and dimensional scales to examine strengths and weaknesses. At the same time, established research questions might be reassessed with new measurement methods, for example whose observation to favor when measuring self-regulation with questionnaires. These new developments might all have specific advantages. For example, when assessing fluctuations in daily life of individuals, we can better define precursors for a specific state, and it might be easier to develop interventions.

The aim of my dissertation is to explore new ways to measure and depict self-regulation across the lifespan. Thereby, I am going to present research examining self-regulation in children and in adults. The co-authors and I have been working with data of participants from general population samples instead of clinical samples, to depict a wider range of self-regulation capacities in our data. With this, I hope to be able to make more overarching conclusions that might apply for a majority of humans.

Although I think that the presented approaches are pointing towards a promising direction to measure self-regulation more reliably, I am aware that much more research is needed to make universally valid conclusions for research and clinical practice. The presented approaches are just a fragment of research questions that might be examined in the future.

In my dissertation, I am going to present three manuscripts that I have been working on as a first author. The first manuscript concerns the inter- and intraindividual differences of self-regulation in the daily life of German schoolchildren as measured with help of ambulatory assessment and its association with night sleep quality and daytime sleepiness. In my second manuscript, I examine the timescales of the fluctuations of self-regulation and working memory performance and their associations on the inter- and intraindividual level with data from the same study. In the third manuscript, I applied a dimensional self-regulation scale and investigate the concordance and differences of self- and observer report about self-regulation abilities in adults. Altogether, these three manuscripts embrace the overarching research question of how to best measure self-regulation. To answer this research question, I would like to propose the following four Hypotheses:

1. We should measure self-regulation on the inter- as well as on the intraindividual level to learn more about the temporal fluctuations of self-regulation performance
2. We should measure self-regulation on the inter- as well as on the intraindividual level to learn more about potential correlates and antecedents of high and low self-regulation
3. We should measure self-regulation dimensionally to depict the strengths and weaknesses of individuals
4. We should measure self-regulation with help of different observers to gain better understanding of context and insight

The following chapters are written as separately readable manuscripts. This results in overlapping contents to this introduction and between the empirical chapters.

4. Empirical Manuscripts

4.1. Manuscript 1: The Association of Self-Reported ADHD Symptoms and Sleep in Daily Life of a General Population Sample of School Children: An Inter- and Intraindividual Perspective

Published as:

Buhr, L., Moschko, T., Eppinger Ruiz de Zarate, A., Schwarz, U., Kühnhausen, J., & Gawrilow, C. (2022). The Association of Self-Reported ADHD Symptoms and Sleep in Daily Life of a General Population Sample of School Children: An Inter- and Intraindividual Perspective. *Brain Sciences*, *12*(4), 440. https://doi.org/10.3390/brainsci12040440

Abstract

Sleep and Attention-Deficit/Hyperactivity Disorder (ADHD) have repeatedly been found to be associated with each other. However, the ecological validity of daily life studies to examine the effect of sleep on ADHD symptoms is rarely made use of. In an ambulatory assessment study with measurement burst design, consisting of three bursts (each 6 months apart) of 18 days each, 70 German schoolchildren aged 10-12 years reported on their sleep quality each morning and on their subjective ADHD symptom levels as well as their sleepiness three times a day. It was hypothesized that nightly sleep quality is negatively associated with ADHD symptoms on the inter- as well as the intraindividual level. Thus, we expected children who sleep better to report higher attention and self-regulation. Additionally, sleepiness during the day was hypothesized to be positively associated with ADHD symptoms on both levels, meaning that when children are sleepier, they experience more ADHD symptoms. No association of sleep quality and ADHD symptoms between or within participants was found in multilevel analyses; also, no connection was found between ADHD symptoms and daytime sleepiness on the interindividual level. Unexpectedly, a negative association was found on the intraindividual level for ADHD symptoms and daytime sleepiness, indicating that in moments when children are sleepier during the day, they experience less ADHD symptoms. Explorative analyses showed differential links of nightly sleep quality and daytime sleepiness, with the core symptoms of inattention and hyperactivity/impulsivity, respectively. Therefore, future analyses should take the factor structure of ADHD symptoms into account.

Keywords: Attention-Deficit/Hyperactivity Disorder; Sleep; Ambulatory Assessment; Multilevel Analysis

Introduction

People with attention deficit/hyperactivity disorder (ADHD) have self-regulation difficulties and frequently experience symptoms of inattention, hyperactivity and impulsivity (American Psychiatric Association, 2013; Sonuga-Barke, 2002). These problems concern around 3,4% of children worldwide (Polanczyk et al., 2015). With an ADHD diagnosis in childhood the probability of negative life outcomes concerning health, vocational, and social areas increases (Caye et al., 2016). High self-regulation on the other hand is associated positively with academic achievement, healthy behaviors and interpersonal relationships (Robson et al., 2020). Therefore, it seems necessary to understand antecedents and correlates of ADHD, to tailor practices for therapy or prevention and enhance self-regulation. Today, theory assumes that ADHD is caused by a combination of biological, psychological, and social factors (Sonuga-Barke, 2002). However, the question how daily life circumstances affect individuals' ability for attention and regulation of behaviour has only scarcely been researched. The study at hand therefore investigates how ADHD symptoms in a general population sample of German schoolchildren are associated with sleep. Thereby, variance cannot only be found with help of clinical samples but also in a general population sample, since people differ in their ability to self-regulate and most people experience at least some ADHD symptoms from time to time.

Dimensionality of ADHD symptoms

Current theories define self-regulation as dimensional in nature. According to this dimensional view, every person lies on a continuum between two extreme poles of high ADHD symptoms on the one side and high self-regulation in behaviour on the other side (Coghill & Sonuga-Barke, 2012; Larsson et al., 2012). Taking that into account, children should not be categorised into those with an ADHD diagnosis and those without the disorder, but differ on the extent of their capability to self-regulate attention and behaviour (Swanson et al., 2012). Supporting that view, research has found that ADHD symptoms in the general population depict a normal distribution ranging from high attention and self-regulation of behaviour to extreme inattention, hyperactivity and impulsivity as well for children as for adults (Blume et al., 2020; Schulz-Zhecheva et al., 2017). To depict the whole continuum of attention and self-regulation of behaviour, research should therefore consider differences in and correlates of ADHD symptomatology in a general population sample instead of applying group comparisons.

Fluctuations in ADHD symptomatology

Recent findings suggest that differences in self-regulation do not only exist between individuals (interindividual; between-person) but ADHD symptoms also fluctuate within individuals (intraindividual; within-person) (Schmid et al., 2016). Ambulatory assessment studies are the golden standard to capture these moments of high and low symptomatology (Koch et al., 2021). Thereby, participants indicate their current experiences repeatedly, for example several times per day on a digital device, like smartphones or tablets (Trull & Ebner-Priemer, 2019). Self-regulation capacities and ADHD symptoms fluctuate highly in the daily lives of children with and without ADHD diagnosis over days and weeks (Ludwig et al., 2016; Schwarz & Gawrilow, 2019). These fluctuations as well as their preceding and following events need to be investigated more thoroughly to better understand the disorder. Indicating which events and experiences lead to better self-regulation of cognition (i.e., attention) and behaviour (i.e., impulsivity) could help to improve the daily lives of people with high levels of ADHD symptoms (Leonard et al., 2021). One phenomenon which has been shown to be associated to cognitive and behavioural measures like executive functions and therefore might also be related to ADHD symptoms is sleep (Hvolby, 2015).

Importance of sleep

Sleep is an important factor for cognitive and psychological functioning in daily life (Gregory & Sadeh, 2012). Sleep is defined as a state with highly diminished consciousness and responsiveness, while brain activity can still be high (Chokroverty, 2017). It is assumed that this brain activity is crucial for memory construction as well as restoration of body and brain tissue. Lack of sleep might thus impair emotion regulation and cognitive functions (Vriend et al., 2012). Therefore, it has often been hypothesized that sleep might also impact the capability to self-regulate ones' behaviour and thereby influence symptoms of inattention, hyperactivity, and impulsivity (e.g. Becker, 2020).

The relation between sleep and ADHD symptomatology

According to the state regulation model, children with sleep loss might not have the energy to adequately regulate their arousal and activation (Van der Meere, 2005). A few studies have implemented sleep restriction and extension experiments, where children followed a strict sleep schedule including significantly less or more sleep than their average sleeping hours. When seven- to eleven-year-old children slept one hour shorter than usual, their teachers described them as less emotionally stable and more hyperactive/impulsive. In the opposite condition, when children slept one hour longer than normally, they were rated as more alert and showed more emotional stability (Gruber et al., 2012). In another study implementing a similar intervention of sleep restriction, this intervention functioned as a moderator of response inhibition and self-regulation in preschool children. With normal sleep schedules, children who showed higher response inhibition applied more self-regulation strategies while playing with an unsolvable puzzle. After the sleep restriction, no association between response inhibition and self-regulation strategies was found (Schumacher et al., 2017). Given this empirical evidence, consequently longer or better sleep should have positive consequences on childrens's self-regulation. When parents of five to twelve year old children with an ADHD diagnosis received a behavioral sleep intervention, which consisted of psychoeducation concerning sleep hygiene practices and standardised behavioural strategies, ADHD symptom levels of the children six months later showed a significantly greater decrease than those of a control group (Hiscock et al., 2015). These findings indicate that there might be an effect of sleep on the ability for attention and self-regulation of behaviour on the between-person level. Thus, children sleeping more and better than others might experience less ADHD symptoms. However, while restricting or extending individuals' sleep under laboratory conditions mirrors typically occurring, intraindividual fluctuations of sleep quality in daily life, to our knowledge no study has explicitly investigated the intraindividual associations between sleep and ADHD symptoms in daily life up to now. Intraindividual fluctuations describe the changes which happen within an individual, for example a child might sleep very good in one night and experience bad sleep in the next. This has to be distinguished from interindividual differences, the between-person difference, where one child has in general better sleep than the other. Both, inter- and intraindividual differences should be considered when investigating the effect of sleep on ADHD symptoms.

Besides the quality of night sleep, which might influence the regulation of attention and behaviour, there also might exist an effect of the current personal experience of tiredness during a specific moment of the day. Although this might seem paradoxical, children who are feeling sleepy could be more instead of less active than usual, as it indeed has been described by many parents (Owens et al., 2012). The feeling of sleepiness might therefore lead to more hyperactive and impulsive symptoms. This observation can also be explained with support of the state regulation model: the evaluation mechanism of the individual might register a state of underarousal due to sleepiness, and therefore react with an enhanced hyperactivity/impulsivity (Van der Meere, 2005). The state of tiredness might also interfere with attention, since children do not possess the energy to regulate their cognition and behaviour adequately. Thus, it is important to examine the daily life of individuals to disentangle how natural fluctuations in sleep quality and tiredness during the day interact with the fluctuations of ADHD symptoms.

Measurement of sleep

The overarching construct of sleep seems to be composed of several different sleep indicators like sleep duration, sleep efficacy, or sleep quality. These indicators in turn might be measured by calculating the hours of total sleep time, the number of awakenings, the time needed to fall asleep (sleep onset latency), and the subjective feeling of being rested in the morning (Van Der Heijden et al., 2005). All of these indicators might thereby be related to other aspects of human functioning. Past research has for example found groups of children with and without ADHD diagnosis to differ in sleep onset latency (the time needed to fall asleep) and sleep efficiency but not in the number of awakenings during the night or the actual hours of being asleep (De Crescenzo et al., 2014). Sleep onset latency has been found to be related to night awakenings, deeper sleep, subjective sleep quality and longer sleep (Gaina et al., 2005). This would make sleep onset latency an economical and short indicator of sleep quality in general.

Another aspect of sleep is the feeling of being tired or sleepy during the day. According to a meta-analysis, this daytime sleepiness has shown higher correlations to school performance than sleep quality and sleep duration (Dewald et al., 2010). This indicates that feelings of sleepiness might be partly independent from the actual sleeping time but still have an impact on self-regulation (Owens et al., 2012). Consequently, researchers should be aware of these different parameters when deciding for an index to measure sleep.

The current study

Considering the above-described research, we were interested in examining how self-reported nightly sleep quality, sleepiness over the day, and ADHD symptoms interact with each other on a between- as well as a within-person level in the daily life of German schoolchildren. ADHD symptomatology was defined on a dimensional level, therefore a general population sample was gathered to depict as much variance in the construct as possible. To account for fluctuations in the measured constructs, ambulatory assessment was used. Both constructs were examined through self-report. Sleep quality was defined by a combination of sleep onset latency and subjective sleep quality. Daytime sleepiness was assessed through indication of the activation level. In the current study, 10–12-year-old children were asked to report on their sleep, sleepiness, and ADHD symptoms on 18 consecutive days. These assessment periods were repeated three times, each time half a year apart, resulting in a maximum of 54 days of assessment. Such an ambulatory assessment not only decreases memory bias, but also ensures a high ecological validity (Koch et al., 2021), and allows to determine both, interindividual differences between the children, as well as intraindividual fluctuations over time.

Building on the state regulation theory as well as on previous findings about the relationship of sleep and ADHD symptoms, we expected the following effects: we predicted a negative relationship between self-rated night sleep quality and self-rated ADHD symptoms on (1) the between-person level across all assessments, and (2) on the within-person level (relation between prior night sleep quality and following day ADHD symptoms). ADHD symptoms should be higher for children who on average sleep worse than other children (interindividual difference), and be higher after a night of worse sleep than a child usually has (intraindividual fluctuation). Further, we expected a positive relationship between self-rated daytime sleepiness and self-rated ADHD symptoms (3) on a between-person level across all assessments, as well as (4) on a within-person level. Children who are sleepier in general are supposed to experience more ADHD symptoms (interindividual difference), and in moments when a child is more tired than usual it is expected to indicate more symptoms (intraindividual fluctuation).

Method

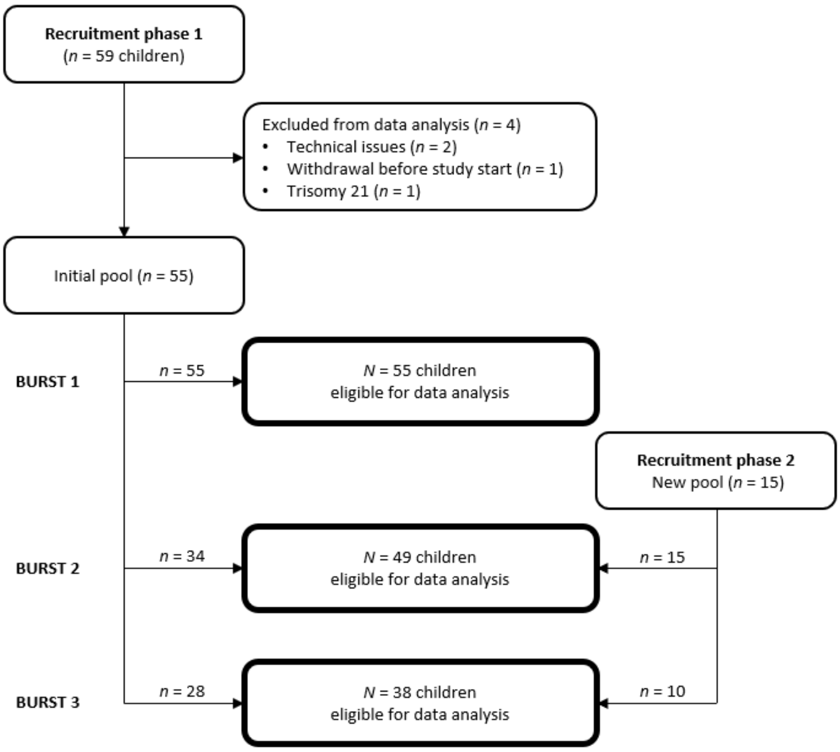
Data was collected within the research project „Adaptive dynamics of cognitive and behavioral variability in children with symptoms of attention deficit /hyperactivity disorder (AttentionGO!)", an intensive longitudinal study which was conducted at the Department of School Psychology at the University of Tübingen in cooperation with the Goethe University, Frankfurt. The project was funded by the German Research Foundation (project number GA 1277/9-1) and approved by the ethics committee of the German Society for Psychology (DGPs, CG 102018_amd_112013). The Ministry of Culture, Youth, and Sport in Baden-Württemberg, Germany, approved recruitment in schools (file number 31-6499.20/1087). The present study refers to three measurement bursts (each lasting 18 days), which took place between autumn 2017 and autumn 2018.

Participants

Participants were recruited in seven schools in southern Germany (n secondary school = 5, n community school = 2). The sample consisted of a total of 70 pupils in Grade 5 (55,71 % female). The age range of the children at the beginning of the study period was 10 to 12 years (M = 10;9 years, SD = 5.7 months). Eight of the participating children had a diagnosis of ADHD, all of them were receiving medical treatment. Exclusion criteria consisted of psychological health (no other diagnosed psychological disorder than ADHD). Figure 2 shows the recruitment process and retention of the participants throughout the study period. Parents and children learned about the possibility to participate in the project through presentations at their respective schools and registered via school. Participation was voluntary and only possible with the written consent of the children and their parents. Participants could end their participation in the study at any time without giving reasons. As compensation for their participation, each family received a voucher worth 40€ for an excursion of their choice (e.g., swimming pool, zoo).

Figure 2

Recruitment process and retention of the participants



In order to prevent any conclusions regarding personal data, all collected data was pseudonymised and stored in a password-protected manner on internal servers of the University of Tübingen. The participating persons were informed about the type of data storage, the handing over of data on request and the deletion of data in accordance with the Basic Data Protection Regulation (DSGVO).

Procedure

For the length of each 18-day survey period, the participating children were given a smartphone (Motorola MotoG4plus ©). Children were trained in school to use the smartphones and fill out the daily questionnaires. It was made sure that all items and instructions were understood by the participants. Each survey period started on a Wednesday and we used a time-contingent sampling method. Smartphones rang three times a day (i.e., in the morning directly after getting up, in the afternoon after school, and in the evening before going to bed) within specific time ranges which were adapted according to individual schedules of the participants. Assessment times could vary on weekends to better fit into the lives of the children. After the signal, children had up to 30 minutes to participate, otherwise the occasion was indicated as missing. Children were asked to give information about their current ADHD symptoms as well as their current feeling of sleepiness on all three assessment moments per day and to indicate their sleep quality during the prior night in the morning measurement directly after getting up.

The study protocol of the ambulatory assessment phase was similar for all three bursts, which were administered approximately half a year apart. However, within the project an intervention to enhance self-regulation was conducted before Burst 2. Children were assigned to one of two groups, with the experimental group receiving the full intervention and the control group receiving a reduced intervention. Both groups showed slight improvement in their self-regulation with no significant difference between the groups (Schwarz & Gawrilow, 2019).

Measures

The ambulatory assessment design poses specific challenges to the scales which are used within scientific studies. First, their wording has to be in such a way that repetitive assessment actually captures fluctuations in the concepts. Therefore, in the current projects all instructions included the phrase “Since the last time I filled in the form…”. Second, participant burden is already very high due to a long study period. Consequently, scales have to be as short as possible to minimize disruption of the daily life of participants and keep compliance rates as high as possible. To account for these concerns, all scales used in the present study were tested in a pilot study. Only items which proved to depict substantial variance were included in the study. Additionally, we tried to apply broad questionnaires, to assess as many research questions as possible without excessive extension of each assessment occasion. In the following, we will describe the adapted items which are relevant for our research questions.

ADHD symptoms

Four items of the children self-report version of the Conners C3-AI Scales (Lidzba et al., 2013) on attention and behaviour were modified for daily recording ("Since the last time I filled in the form I talked too much."; "Since the last time I filled in the form I forgot what I was supposed to do."; "Since the last time I filled in the form I had too much energy to sit still."; "Since the last time I filled in the form I could hardly concentrate."). The self-report scales are suited for children of eight to 18 years of age (Lidzba et al., 2013). The children indicated on a Likert scale how much the statements applied to them since the last assessment (1 = not at all to 6 = exactly). High values therefore expressed high ADHD symptom levels. To obtain an ADHD score for each measurement time point, we calculated averages across all four items for each moment the child answered at least three of the four items. We computed multilevel reliability estimates using generalizability theory analyses (Shrout & Lane, 2012) to determine the reliability of these scores to capture individual differences (between-person reliability R_{KF}; .98-.99)[1], as well as day-to-day fluctuations in symptom levels (within-person reliability R_C; .61-.69). Multilevel reliability estimates for only afternoon assessments was .93-.97 (between-person), and .59-.70 (within-person). Additionally, we checked for validity by comparing our modified version of the Conners scales for the ambulatory assessment with the standardized measures (without modification) of the Conners ADHD index score that the children filled out in school before each burst. Mean ADHD scores from daily assessment within each study burst are associated weakly but significantly with the child's ADHD index score, assessed at the

[1] As initial sighting of the data indicated that there might be substantial differences in children's ADHD symptom levels between bursts we computed the reliability estimates for the ADHD scale separately for each burst

beginning of each measurement burst, respectively. More specifically, there was a weak correlation for Burst 1, $r(53) = 0.47$, $p < .001$, for Burst 2, $r(41) = 0.31$, $p = .040$, and for Burst 3, $r(34) = 0.42$, $p = .011$, indicating that higher mean ADHD scores from daily assessment within each study burst were associated with higher ADHD index scores. Therefore, we concluded that the modified ADHD scales were valid to measure the construct we intended.

Sleep

The items for self-report of subjective sleep quality were adapted from the study by Könen and colleagues (2015). Children rated their sleep quality of the previous night on a Likert scale from one (*poor*) to six (*good*). The time taken to fall asleep was also recorded on a Likert scale from one (*long*) to six (*not long*). Thus, a high value of the duration of falling asleep indicated that children fell asleep quickly. To calculate a sleep quality score, the average of both items was computed, with higher scores indicating better night sleep quality. Between-person reliability for this score was .97, and within-person reliability was .54.

Daytime Sleepiness

To indicate their current affect, children filled out a slightly modified and shortened version of the Multidimensional Mood Questionnaire (MDMQ; Steyer et al., 1997). Eight items were answered on individual 6-point Likert scales. The daytime sleepiness was calculated by averaging the following two items: (1.) “At the moment I feel tired (1) or well rested (6)” and (2.) “At the moment I feel sleepy (1) - awake (6)”. Between-person reliability for this score was .99, and within-person reliability was .81.

Statistical Analysis

All analyses for the current research question were preregistered (Doi: 10.17605/OSF.IO/T9XEA, https://archive.org/details/osf-registrations-t9xea-v1). The data was processed and analysed with help of the programme R (R Core Team, 2020, version 4.1.1.), using the nlme package (version 3.1.-153) to conduct multilevel regression analyses. To analyze between- and within-person associations between children’s night sleep quality and ADHD symptom levels the following day (Hypothesis 1 and 2), we used a multilevel model including a random intercept and random slopes for time and within-person fluctuations in sleep quality (e.g. Bolger & Laurenceau, 2013). Due to the specific assessment design, we used several time variables. Data is nested within bursts, which were administered each half a year apart. To account for this nested structure and possible trends in missing data, variables were included to account for the 18 days within each burst and the respective differences in results of Burst 2 and Burst 3 compared to Burst 1. We expected missing data to be higher on weekends and additionally assumed differences in sleep quality between weekends and weekdays. Therefore, we included weekend as a control variable into the models. For this specific analysis, we paired night sleep quality ratings assessed in the morning and rating of ADHD symptom levels assessed the following afternoon. To avoid biased results just due to extreme individual reports of either night sleep quality, or ADHD symptom levels on certain days, we considered data points that lie three standard deviations above or below a participant's individual mean across time as outliers and excluded them from all data analyses. To differentiate the effects of within-person fluctuations from trait-like individual differences in sleep quality, we split the raw scores into two components: a between-person component indicating individual i's trait-like tendency for better/worse sleep than other individuals[2], and a within-person component indicating individual i's tendency on day t to have slept better/worse than usual. To facilitate the interpretation of results and comparison of within- and between-person effects, we divided the predictor (within-person fluctuations and between-person differences in sleep quality) by the between-person standard deviation across the study period to identify small, moderate, and large effect sizes in standard deviation units (Cohen, 1992). Based on previous findings, we included gender, age and ADHD medication as control variables in the model without specific hypotheses. Equation 1 describes the full model tested:

---

[2] this between-person component was calculated by subtracting the sample's grand mean from each person mean (a participants average across all study days). The grand means for all variables of interest (Table 1) were obtained by calculating the average of all person means.

Equation 1:

$$ADHD_{it} = (\gamma_{00} + u_{i0}) + (\gamma_{01} + u_{i1})\, Time_{it} + \gamma_{02}\, SleepB_i + (\gamma_{03} + u_{i2})\, SleepW_{it} + \gamma_{04}\, Weekend_{it} + \gamma_{10}\, Burst2_{it} + \gamma_{11}\, Burst2_{it}{*}Time_{it} + \gamma_{12}\, Burst2_{it}{*}SleepB_i + \gamma_{13} Burst2_{it}{*}SleepW_{it} + \gamma_{14}\, Burst2_{it}{*}Weekend_{it} + \gamma_{20}\, Burst3_{it} + \gamma_{21}\, Burst3_{it}{*}Time_{ti} + \gamma_{22}\, Burst3_{it}{*}SleepB_i + \gamma_{23}\, Burst3_{it}{*}SleepW_{it} + \gamma_{24}\, Burst3_{it}{*}Weekend_{it} + \gamma_{30}\, Gender_i + \gamma_{31}\, Age_i + \gamma_{32}\, Medication_i + \varepsilon_{it}$$

Using this equation, we tested whether the following fixed effects differ from 0:

(a) an intercept, γ_{00}, representing the average level of ADHD symptoms on study day 1 during Burst 1;

(b) an average linear time trend, γ_{01}, indicating the change in ADHD symptom levels over the 18 study time days during Burst 1, centered on Day 1;

(c) the between-person effect of sleep quality during Burst 1, centered at the sample's grand mean in sleep quality across all three bursts, γ_{02}, indicating the difference in ADHD symptom levels for participants with better sleep quality of one unit (i.e., one between-person standard deviation in sleep quality), compared to the typical participant's sleep quality;

(d) the within-person effect of sleep quality during Burst 1, centered at the participant's personal mean in sleep quality across all three bursts, γ_{03}, indicating the change in ADHD symptom levels on days following night with better sleep of one unit (i.e., one between-person standard deviation in sleep quality) than the participant's usual level in sleep quality;

(e) the weekend effect, γ_{04}, indicating the mean difference in ADHD symptom levels on weekend days (i.e., Saturday and Sunday; coded 1), and school days (i.e., Monday to Friday; coded 0);

(f) the difference in the mean level of ADHD symptoms on study day 1 in Burst 2 (coded 1) compared to Burst 1 (coded 0), γ_{10};

(g) the difference in the average linear time trend in Burst 2 (coded 1) compared to Burst 1 (coded 0), γ_{11};

(h) the difference in the between-person effect of sleep quality in Burst 2 compared to Burst 1, γ_{12};

(i) the difference in the within-person effect of sleep quality in Burst 2 compared to Burst 1, γ_{13};

(j) the difference in the weekend effect in Burst 2 compared to Burst 1, γ_{14};

(k) the difference in the mean level of ADHD symptom on study day 1 in Burst 3 (coded 1) compared to Burst 1 (coded 0), γ_{20};

(l) the difference in the average linear time trend in Burst 3 (coded 1) compared to Burst 1 (coded 0), γ_{21};

(m) the difference in the between-person effect of sleep quality in Burst 3 compared to Burst 1, γ_{22};

(n) the difference in the within-person effect of sleep quality in Burst 3 compared to Burst 1, γ_{23};

(o) the difference in the weekend effect in Burst 3 compared to Burst 1, γ_{24};

(p) the effect of children's gender, γ_{30}, indicating the mean difference in ADHD symptom levels between boys (coded 1), and girls (coded 0);

(q) the effect of children's age, γ_{31}, indicating the difference in ADHD symptom levels for older participants of one unit (month);

(r) the effect of ADHD medication, γ_{32}, indicating the mean difference in ADHD symptom levels between children receiving ADHD medication (coded 1), and children not receiving ADHD medication (coded 0).

The model in Equation 1 also tested whether the following between- and within-person random effects differ from 0:

(s) u_{0i} captures how much a particular participant deviates from the average intercept (i.e., random intercept);

(t) u_{1i} captures how much a particular participant deviates from the average time slope (i.e., random time slope);

(u) u_{2i} captures how much a particular participant deviates from the average within-person effect (i.e., random sleepiness slope);

(v) ε_{it} indicates how much a particular participant's ADHD symptom levels on a gives study time point deviates from the value predicted by their person-specific regression line (i.e., residual error).

We allowed for a maximal random effects structure with covariances of all random effects. To account for the intensive longitudinal data structure, we modelled time dependence of the residuals with a first-order autoregressive structure (AR1; e.g. Bolger & Laurenceau, 2013). Model analyses were conducted with restricted maximum likelihood estimation and a probability level of $p < .05$ to indicate significance of effects based on t-values of each model coefficient.

Likewise, we tested the between- and within-person associations between children's daytime sleepiness and their ADHD symptom levels (Hypothesis 3 and 4), using children's sleepiness and ADHD symptom ratings collected three times a day – that is, on up to 54 study time points per burst – with between-person differences and within-person fluctuations in daytime sleepiness rather than sleep quality as predicting variable within the regression model.

Results

Descriptive results

The number of possible observations was calculated by multiplying 18 days with 3 bursts for all 55 children that were recruited in November 2017 and with 2 bursts for the 15 children that started with the study in April 2018. This procedure ensured that those participants who were newly recruited for the second burst, were not inflating the dropout rate. Thus, there were up to 3,510 observations of night sleep quality possible. As daytime sleepiness and ADHD symptom levels were assessed three times a day, this results in up to 10,530 observations, respectively. Data on night sleep quality was collected 2051 times, resulting in a participation rate of 58%, while data on daytime sleepiness was collected 5799 times (55%).

In total, data on ADHD symptom levels was collected 5733 times (54%), with 1669 observations collected in the afternoon (48% of all possible observations in the afternoon). Before further data analyses, 36 night sleep quality, 21 daytime sleepiness and 110 ADHD symptom level observations were excluded due to being defined as outliers. When only considering ADHD symptom level in the afternoon, 132 observations had to be excluded. Within each study burst, missing values were more likely to occur on weekends compared to school days by up to 68% (Burst 1: *OR* = 1.38, 95% CI [1.14, 1.67], Burst 2: *OR* = 1.68, 95% CI [1.40, 2.01], and Burst 3: *OR* = 1.43, 95% CI [1.17, 1.76]). Moreover, with each day within a study burst the likelihood for missing values increased by up to 6% compared to the previous day (Burst 1: *OR* = 1.06, 95% CI [1.04, 1.08], Burst 2: *OR* = 1.06, 95% CI [1.04, 1.07], and Burst 3: *OR* = 1.04, 95% CI [1.02, 1.06]).

Table 1

Descriptive statistics for children's ADHD symptom levels, night sleep quality, and daytime sleepiness across all 54 study days.

	Between-person			Within-person			
	M	*SD*	range	M_{ISD}	*SD*	range	*ICC*
ADHD (afternoon)[1]	1.62	0.60	1.00-3.64	0.59	0.41	0.00-1.92	0.44
ADHD (overall)[2]	1.52	0.50	1.00-2.91	0.56	0.37	0.00-1.79	0.43
Night sleep quality	4.58	0.82	2.23-6.00	1.11	0.57	0.00-2.37	0.43
Daytime sleepiness	2.68	0.98	1.00-4.42	1.48	0.55	0.00-2.28	0.34

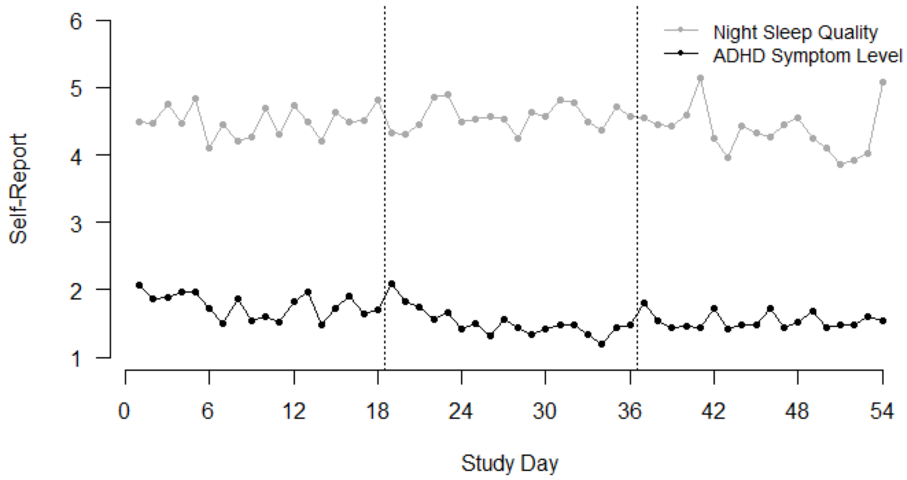
Note. M_{ISD} = mean intra-individual standard deviation, ICC = intraclass correlation coefficient, theoretical range for all variables: 1-6, with higher values indicating higher ADHD symptom levels, better night sleep quality, and higher daytime sleepiness, respectively.
[1] ADHD symptom level reports collected only at afternoon time points.
[2] ADHD symptom level reports collected at all three time points during a day.

Mean self-report of ADHD symptom levels in the sample was relatively low (*M* = 1.52, *SD* = 0.50; only afternoon: *M* = 1.62, *SD* = 0.60). Children reported medium to high sleep quality (*M* = 4.58, *SD* = 0.82). Most variance was indicated for daytime sleepiness (*M* = 2.68, *SD* = 0.98). Table 1 lists descriptive statistics of all constructs utilized for testing of hypotheses. The intraclass correlation coefficients (ICC) indicate how much of the total variance can be explained by variance on the interindividual level. Thus, around 44% of the variance in ADHD symptom levels can be explained by interindividual differences. Consequently, around 56% of the variance is composed of intraindividual fluctuations and measurement error.

Fluctuations of the variables can be inspected more thoroughly in Figure 3 for night sleep and Figure 4 for daytime sleepiness. For sleepiness, the graph indicates higher values in the mornings as the evenings, as would be expected in normal circadian rhythms.

Figure 3

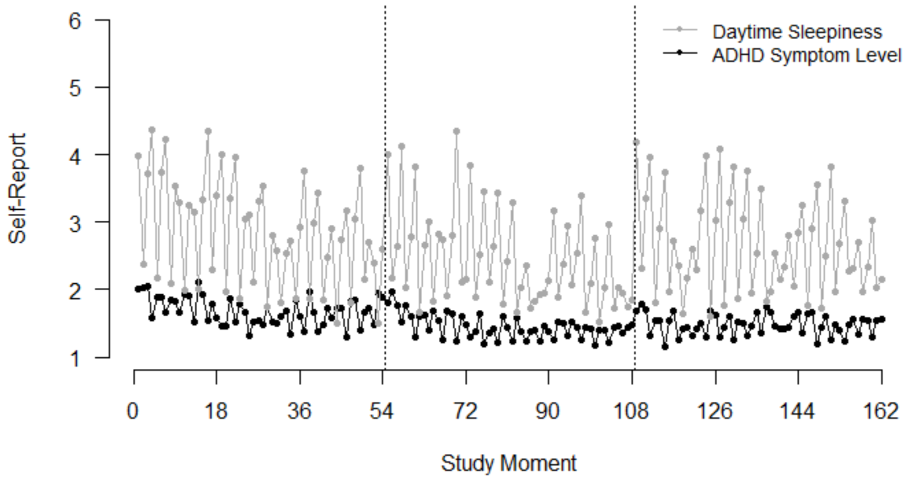
Time course of self-reported night sleep quality and ADHD symptom levels (in the afternoon) across all 54 study days



Note. The dashed lines indicate breaks between each burst.

Figure 4

Time course of self-reported daytime sleepiness and ADHD symptom levels across all 162 study moments, with three moments per day



Note. The dashed lines indicate breaks between each burst.

Multilevel Analyses

Association of night sleep quality and ADHD symptoms

To assess the association of sleep quality during the preceding night and ADHD symptoms during the school day, multilevel models were conducted. As can be seen in

Table 2, we found no significant associations between night sleep quality and ADHD symptom levels during the initial study burst, neither on the interindividual nor on the intraindividual level. This did not change during the subsequent bursts, except for a significant increase in the between-person association of night sleep quality and ADHD symptom level in Burst 2 compared to Burst 1 (γ_{12} = 0.18 (SE = 0.08), *p* = 0.03), which almost annulated a non-significant trend for this between-person association in Burst 1 (γ_{02} = -0.16 (SE = 0.09), *p* = 0.07). However, we found that ADHD symptoms decreased significantly during Burst 1 (γ_{01} = -0.27 (SE = 0.13), *p* = 0.04), with no significant changes in within-burst decrease rates across bursts. Also, children reported significantly lower ADHD symptom level at the beginning of

Burst 3 compared to Burst 1 (γ_{01} = -0.25 (SE = 0.11), p = 0.02). The random intercept showed to be significant in the analyses, indicating significant differences in children's initial ADHD symptom levels.

Table 2

Multilevel model to test the between- and within-person association between children's night sleep quality and ADHD symptom levels the following day.

Fixed Effects		Estimate		*SE*	*p*
Burst 1					
Intercept: initial level	γ_{00}	2.07		1.60	.20
Time slope[a]	γ_{01}	-0.27	*	0.13	.04
Night sleep quality, between-person differences	γ_{02}	-0.16		0.09	.07
Night sleep quality, within-person fluctuations	γ_{03}	-0.02		0.02	.40
Weekend effect	γ_{04}	-0.02		0.06	.75
Change at Burst 2, compared to Burst 1					
Change in level	γ_{10}	-0.03		0.10	.74
Change in time slope	γ_{11}	-0.21		0.18	.24
Change in effect of night sleep quality (between-person)	γ_{12}	0.18	*	0.08	.03
Change in effect of night sleep quality (within-person)	γ_{13}	-0.002		0.04	.95
Change in weekend effect	γ_{14}	0.02		0.09	.85
Change at Burst 3, compared to Burst 1					
Change in level	γ_{20}	-0.25	*	0.11	.02
Change in time slope	γ_{21}	0.02		0.19	.92
Change in effect of night sleep quality (between-person)	γ_{22}	0.12		0.09	.17
Change in effect of night sleep quality (within-person)	γ_{23}	0.02		0.04	.61
Change in weekend effect	γ_{24}	-0.07		0.09	.44
Control variables					
Gender	γ_{30}	0.08		0.14	.56
Age[b]	γ_{31}	-0.003		0.01	.80
ADHD medication[c]	γ_{32}	0.29		0.23	.22
Random Effects & Covariances		Estimate			*p*[d]
Level 2 (between-person)					
Intercept: initial level	$SD(u_{0i})$	0.62	***		<.001
Time slope	$SD(u_{1i})$	0.49			.93
Sleep quality within-person fluctuations	$SD(u_{2i})$	0.06			.93
Intercept and time	$r(u_{0i}, u_{1i})$	-0.51	**		.003
Intercept and sleep quality fluctuations	$r(u_{0i}, u_{2i})$	-0.78	*		.03
Time and sleep quality fluctuations	$r(u_{1i}, u_{2i})$	0.78			.99
Level 1 (within-person)					
Residual	$SD(\varepsilon_{it})$	0.66			
Autocorrelation	ρ	0.34	***		<0.001

Note. N = 70 children, n = 1450 considered observations, *p < 0.05, **p < 0.01, ***p < 0.001

[a] Time is coded 0 = study day 1 within a measurement burst, 1 = study day 18 within a measurement burst, with equal intervals for the intervening study moments
[b] We were not able to collect data on one child's age. To avoid it falling out from data analysis, we set its age to the sample mean age
[c] We were not able to collect data on one child's medication status. To avoid it falling out from data analysis, we assumed it was not receiving ADHD medication
[d] The respective p-values for the random effect estimates were obtained by testing in pairs a model that includes the parameter in question against a model missing just this parameter via likelihood ratio tests.

Association of daytime sleepiness and ADHD symptoms

Similarly, to test for the association of self-reported daytime sleepiness and ADHD symptoms at the same time, we likewise calculated a multilevel model (Table 3). No effect of daytime sleepiness on ADHD symptoms can be seen on the interindividual level in the initial study burst. Again, we found a significant change in the between-person association of daytime sleepiness and ADHD symptom levels in Burst 2 compared to Burst 1 ($\gamma 12 = -0.12$ (SE = 0.06), $p = 0.04$), which almost annulated a non-significant trend for this between-person association in Burst 1 ($\gamma 02 = -0.16$ (SE = 0.10), $p = 0.11$). However, the data indicated a negative within-person association of daytime sleepiness and ADHD symptoms in the initial study burst ($\gamma 02 = -0.04$ (SE = 0.02), $p = 0.01$), with no significant changes in the subsequent bursts. This result suggests that in moments when participants felt more tired during the day they indicated less ADHD symptoms, therefore contradicting our hypothesis. Also, we again found significant decreases in ADHD symptom levels throughout each study burst, as well as overall decreased symptom levels at the beginning of Burst 2, and Burst 3, compared to Burst 1, respectively. Regarding our control variables, we found increased ADHD symptom levels in children receiving ADHD medication, which overlaps with an ADHD diagnosis in our sample.

Significant effects were found for the random intercept and both random slopes, implying substantial variance in children's initial ADHD symptom levels, variance in symptom level change within bursts, and variance in the size of the within-person association of ADHD symptom levels and daytime sleepiness.

Table 3

Multilevel model to test the between- and within-person association between children's daytime sleepiness and concurrent ADHD symptom levels.

Fixed Effects		Estimate		*SE*	*p*
Burst 1					
Intercept: initial level	γ_{00}	1.65		1.13	.14
Time slope[a]	γ_{01}	-0.23	**	0.08	.004
Daytime sleepiness, between-person differences	γ_{02}	0.16		0.10	.11
Daytime sleepiness, within-person fluctuations	γ_{03}	-0.04	*	0.02	.01
Weekend effect	γ_{04}	0.05		0.03	.14
Change at Burst 2, compared to Burst 1					
Change in level	γ_{10}	-0.16	**	0.06	.006
Change in time slope	γ_{11}	-0.06		0.10	.57
Change in effect of daytime sleepiness (between-person)	γ_{12}	-0.12	*	0.06	.04
Change in effect of daytime sleepiness (within-person)	γ_{13}	0.01		0.02	.50
Change in weekend effect	γ_{14}	-0.09		0.05	.10
Change at Burst 3, compared to Burst 1					
Change in level	γ_{20}	-0.28	***	0.06	<.001
Change in time slope	γ_{21}	0.06		0.10	.57
Change in effect of daytime sleepiness (between-person)	γ_{22}	-0.06		0.07	.39
Change in effect of daytime sleepiness (within-person)	γ_{23}	0.03		0.02	.18
Change in weekend effect	γ_{24}	-0.09		0.06	.10
Control variables					
Gender	γ_{30}	-0.002		0.10	.98
Age[b]	γ_{31}	0.0003		0.009	.97
ADHD medication[c]	γ_{32}	0.55	**	0.17	.001
Random Effects & Covariances		Estimate			p^d
Level 2 (between-person)					
Intercept: initial level	$SD(u_{0i})$	0.50	***		<.001
Time slope	$SD(u_{1i})$	0.40	***		<.001
Sleepiness within-person fluctuations	$SD(u_{2i})$	0.07	***		<.001
Intercept and time	$r(u_{0i}, u_{1i})$	-0.30			.05
Intercept and sleepiness fluctuations	$r(u_{0i}, u_{2i})$	-0.73	***		<.001
Time and sleepiness fluctuations	$r(u_{1i}, u_{2i})$	-0.04			.54
Level 1 (within-person)					
Residual	$SD(\varepsilon_{it})$	0.64			
Autocorrelation	ρ	0.27	***		<.001

Note. N = 70 children, n = 5559 considered observations, *p < 0.05, **p < 0.01, ***p < 0.001

[a] Time is coded 0 = study day 1 within a measurement burst, 1 = study day 18 within a measurement burst, with equal intervals for the intervening study moments

[b] We were not able to collect data on one child's age. To avoid it falling out from data analysis, we set its age to the sample mean age

[c] We were not able to collect data on one child's medication status. To avoid it falling out from data analysis, we assume it was not receiving ADHD medication
[d] The respective p-values for the random effect estimates were obtained by testing in pairs a model that includes the parameter in question against a model missing just this parameter via likelihood ratio tests.

Explorative post-hoc analysis

After conducting all planned and pre-registered analyses, we decided to investigate children's ADHD symptom levels separately for symptoms of inattention and symptoms of hyperactivity-impulsivity. As we will argue in the discussion section, these core dimensions of ADHD might show a unique and discriminative link to markers of sleep quality in everyday life. To this end, we conducted two separate scores for children's levels of inattention ("Since the last time I filled in the form I forgot what I was supposed to do."; "Since the last time I filled in the form I could hardly concentrate."), and levels of hyperactive-impulsive behavior ("Since the last time I filled in the form I talked too much."; "Since the last time I filled in the form I had too much energy to sit still."), and reran our analysis in correspondence to the procedure described above, however, with separate multilevel models for levels of inattention, and hyperactive-impulsive behavior. The complete results of these post-hoc analyses are added to the appendix (Table 4 and Table 5). In summary, regarding our initial inter- and intraindividual hypotheses, we found the following results with respect to the association of night sleep quality and inattention: (a) a negative between-person association between children's night sleep quality and levels of inattention the following day in Burst 1 – that is, children who sleep better than others report to have lower levels of inattention – but the size of this association decreased significantly in Burst 2 compared to Burst 1, and in Burst 3 compared to Burst 1; and (b) a negative within-person association between children's night sleep quality and levels of inattention the following day across all measurement bursts – that is, after sleeping better than usual, children report to have lower levels of inattention the following day. No inter- or intraindividual associations were found between night sleep quality and hyperactivity-impulsivity. Regarding the relationship between daytime sleepiness and inattention (c) a positive between-person association between children's daytime sleepiness and levels of inattention in Burst 1 – that is, children with higher levels of daytime sleepiness than others report higher levels of inattention – but the size of this association decreased significantly in Burst 2 compared to Burst 1, and in Burst 3 compared to Burst 1. There was no within-person association evident between daytime sleepiness and levels of inattention. For the core symptom of hyperactivity-impulsivity, we found (d) a negative within-person association to daytime sleepiness across all measurement bursts – that is, children report lower levels of hyperactive-impulsive behavior in moments of higher daytime sleepiness than usual. There was no between-person association evident between daytime sleepiness and levels of hyperactive-impulsive behavior.

Discussion

In the current study, we investigated the relationship of self-reported sleep variables and ADHD symptoms on a between- as well as a within-person level in German schoolchildren. With an intensive longitudinal study, applying a measurement burst design with ambulatory assessment, daily fluctuations in the constructs of sleep quality, daytime sleepiness and ADHD symptom levels were assessed. With this measurement approach we expanded the current literature on sleep and ADHD. In contrast to earlier studies, which compared groups of children with and without ADHD diagnosis, we defined ADHD symptoms on a dimensional level by using a general population sample, in line with current dimensional theories for the classification of psychological disorders (Coghill & Sonuga-Barke, 2012). Fluctuations in all constructs could be investigated further due to the repeated measurement. Finally, ecological validity was enhanced in comparison to laboratory studies by implementing ambulatory assessment in the daily life of participants with the help of smartphones.

Multilevel analyses did not confirm a relationship between sleep quality during the night and ADHD symptoms on the subsequent day on the inter- or the intraindividual level. Accordingly, we must reject our first two hypotheses, since we had expected a negative effect of sleep quality on ADHD symptoms within and between children. Although this is to our knowledge the first study to investigate the relationship of sleep and ADHD symptoms in (school) children's daily lives, theory and previous research would have hinted to such a connection. State regulation theory implies that with worse sleep (inter- and intraindividually), children have less capacity for self-regulation and therefore show more ADHD symptoms (Van der Meere, 2005). Support for this has been found in earlier studies. For example, when their sleep was restricted for several nights within an experimental study, children experienced significantly more problems with alertness and emotional regulation (Gruber et al., 2012). In contrast to such an experimental study, where sleep restriction was externally exerted, children in our study seemed to have a relatively good night sleep and indicated an overall good sleep quality. With this lack of variance, it might have been difficult to find an effect of sleep quality on ADHD symptoms even if it was present. Additionally, earlier research findings are hinting to an effect in form of an inverted U shape of sleep on cognitive functioning, with too much sleep provoking a negative effect. After nights when children sleep either much less or much more than on average, they perform worse in a working memory task than after nights with their usual sleep length (Könen et al., 2015).

Similarly, we did not find an interindividual effect of subjective daytime sleepiness on ADHD symptoms. However, on the within-person level, we found a negative association between daytime sleepiness and ADHD symptoms. Our data implies that in moments when children feel more tired and sleepy during the day, they report less ADHD symptoms than when they feel more activated and awake. This finding is contrary to our hypothesized effect. It seems that children who are well rested also experience more energy to feel restless. One possible explication for the effect could lie in the timely structure of the study. Sleepiness and ADHD symptoms were measured three times a day. As Figure 3 shows, children indicated high sleepiness in the morning and evening but low sleepiness in the mid-day measurement. However, the time when self-regulation is most needed and therefore ADHD symptoms might be most easily detected is the time that children spend in school (Schwarz & Gawrilow, 2019). Therefore, it might be assumed that children were not actually able to inform about their ADHD symptoms at specific times of the day, since self-regulatory processes were not needed that strongly. In future studies, it might be interesting to examine the interaction of sleepiness and ADHD symptoms during the school day.

Especially interesting is the found effect of the three bursts on ADHD symptom levels. The findings indicate that in Burst 3 ADHD symptoms at the beginning of the burst are significantly lower than in Burst 1. In the second model, this effect can also be found for Burst 2, where children start lower than in Burst 1. The difference between the models results from the fact that in the first model only ADHD symptoms on the second moment at the middle of the day are considered. The second model uses all three indications of ADHD symptom levels on each day. Several possible explanations can be found for this effect of burst on ADHD symptom levels. The most obvious explanation might be an aging effect. Within the course of normal development, children get more attentive and learn to better self-regulate their behaviour. Therefore, symptoms of inattention and hyperactivity/impulsivity decrease with age (Döpfner et al., 2015; Faraone et al., 2006). Another possible cause for the effect might lie in the format of the study protocol. The study included an intervention after Burst 1, aiming at promoting self-regulatory behaviour. To this end, children were allocated to two different intervention groups (mental contrasting with implementation intentions vs. mental contrasting), with both groups showing similar improvement in self-regulation following the interventions (Schwarz & Gawrilow, 2019). As there were no differential intervention effects, we would not expect that the implementation of the intervention confounds the relationship between sleep quality and ADHD symptoms investigated in the current study. However, to reassure that results are not influenced by this experiment, we integrated the intervention as a control variable in a post-hoc analysis. None of the effects changed due to this additional variable as can be seen in Appendix B.

Another result we found is the significant decrease of ADHD symptoms within each of the bursts. In general, children reported significantly less symptoms at the end of the burst than in the beginning. This could be explained by an initial-elevation bias (Shrout et al., 2018). Independently of the topic of ambulatory assessment studies, self-reports are often higher in the first measurement timepoints and get more stable after a while. A support for this assumption in our data can be found in Figure 3 where higher values at the beginning of each burst are graphically depicted. It might be helpful to further investigate this effect and possibly conduct future analyses without the first few measurement timepoints.

In our original models, we integrated all three core symptoms of ADHD into one common factor to enhance the reliability of the scale. However, previous research has found that different ADHD subtypes might be associated with different sleeping patterns (Hvolby, 2015; Paavonen et al., 2009). For our study this might imply that children feel less hyperactivity/impulsivity symptoms when they are tired but are at the same time more inattentive. Therefore, in an explorative post-hoc analysis, which was not preregistered, we individually examined the two ADHD symptom factors inattention and hyperactivity/impulsivity separately in models with nightly sleep quality and daytime sleepiness. We found a significant negative effect of night sleep quality on inattention on the inter- as well as on the intraindividual level. Thus, children who slept better on average indicated less inattentive symptoms in general and after a night when they slept better, children indicated less inattentive symptoms. Interestingly, the interindividual effect decreases in Burst 2 and Burst 3 respective to Burst 1. Furthermore, we found a positive interindividual effect between sleepiness and inattention; children who report in general to be more tired also report to experience more inattention. However, also this effect seems to be smaller in Burst 2 and 3 than in Burst 1. We did not find an intraindividual effect between sleepiness and inattention. For hyperactivity/impulsivity as the dependent variable, we solely found a significant negative intraindividual effect of sleepiness. Thus, in moments when children were sleepier, they indicated less hyperactive/impulsive symptoms. Other than that, no effect of sleep on hyperactive/impulsive symptoms was found.

These exploratory analyses incorporate some interesting insights into our data and the implications should be investigated more thoroughly in future research. To summarize, it seems that our hypotheses apply better for the inattentive factor of ADHD while hyperactivity/impulsivity seem not to be related to sleep and sleepiness as measured in our study, or even in the opposite direction than expected. These findings might reflect a general effect where sleep quality and sleepiness only affect attention. It might however also be an effect of the specific age group. People tend to grow calmer with age and are better able to self-regulate their behaviour with age. This effect has often been shown in ADHD research, where adults report less hyperactive/impulsive symptoms than children but still significant impairments in their attention (Faraone et al., 2006). Thus, children in our sample might already have outgrown the tendency to show more hyperactive/impulsive symptoms when they have slept badly or feel sleepier.

Limitations

Despite the numerous advantages the current study adds to the existing research literature, the study design also might incorporate specific drawbacks and potential for improvement. In general, ambulatory assessment has great potential to capture daily fluctuations in ADHD symptoms and sleep of children. However, there is still a lack of adequately tested scales to use within this specific research design (Shrout & Lane, 2012). We tried to account for this by slightly modifying the scales and selecting only specific items which proved to show substantial variances within a pilot study. Nevertheless, future research might show that different scales are better suited to depict the fluctuations of ADHD symptoms and sleep in the daily life of schoolchildren.

Another drawback of ambulatory assessment always is the high participant burden which is put on the participants. Answering the same questions three times a day for 18 days in three different bursts is very exhausting, especially for children. Although we shortened the scales as much as possible, occasions with missing answers increased with time within each burst and many children dropped out of the study between the bursts. We tried to control for these dropouts and missing data by including burst, day within burst and weekend into the models. With compliance rates of 48-58% we received enough data to model inter- and intraindividual differences of the children. Still, future research should try to prevent this dropout effect by reducing burden and enhancing commitment of the participants.

Most obvious seems to be the question whether self-reports of sleep are a valid instrument to measure actual sleep quality in children. For measuring sleep, subjective and objective measurements all show their own advantages and drawbacks (Hvolby, 2015). The utilization of polysomnography in sleep laboratories leads to well documented physical and neurological data but lacks ecological validity. Actigraphs can easily be worn at home in the participants' natural environment but have their drawbacks in only measuring movement and therefore being fault-prone in indicating sleep. In sleep research with children, parent report is often used to gather information about quantity and quality of sleep, however, as children grow older and get more independent, parents' might lose insight into their actual sleeping behaviour. Research with adults and adolescents is often relying on self-report measures of sleep. It has been found that children report more problems falling asleep and retaining sleep than their parents indicate (Fricke-Oerkermann et al., 2007). Therefore, we were interested to see how children would self-report their quality of sleep in their daily life and how this data is related to other measures like self-reported ADHD measures (Könen et al., 2015).

Critics might object that participants of this age might lack the relevant introspection and humans in general might not be able to give valid reports of their sleep, given that the key feature of sleep is the lack of consciousness (Chokroverty, 2017). This limitation of the study should be integrated into future studies, which might use combinations of self-report with more objective measures like polysomnography or actigraphy (Hvolby, 2015). In the current study, actigraphs were only administered throughout the day to minimize participant burden, therefore we had to rely on self-report of sleep quality during the night. This question of the amount of introspection for self-report in children of this age group might also apply the assessment of ADHD symptoms. Here as well, future research should compare these self-reports with more objective measures or parent- and teacher-rated scales to examine the validity of the children's responses. However, since we found very high between-person reliability and high within-person reliability in our analyses, we figured the self-report scales to be adequate for the assessment of ADHD symptoms in the daily life of children.

Furthermore, we decided to measure sleep quality by combining self-reported sleep onset latency and subjective sleep quality. These constructs have shown to be related to other psychological factors in earlier studies (e.g. Könen et al., 2015) and depicted most variance in a pilot study. Given the already high participant burden in the study, questionnaires had to be as concise as possible. Other constructs indicating sleep should be investigated further. For example, the total hours of sleep could be examined (Paavonen et al., 2009; Pesonen et al., 2010). Nevertheless, also this construct has its drawbacks, since sleep needs might differ between children. Also, number of awakenings during the night could be a good indicator of sleep. This construct however might be difficult to measure in self-report since people often do not remember their awakenings the next day.

Implications and Future Research

Although the current study did not confirm the hypotheses, it might bring new ideas and questions to the research area. A very positive finding is the fact that children in our general population sample indicated overall relatively high sleep quality and low ADHD symptoms throughout their daily lives. Comparing different methods to evaluate sleep quality might help to define which measurement might be related to other physiological and psychological constructs. The finding that ADHD symptoms seem to decrease over time, both between and within the bursts should be further examined. The first might be investigated in future research by further examining developmental changes throughout the lifespan. For the second, the initial elevation bias should be integrated more into the planning and evaluations of ambulatory assessment studies.

Since most items that were used in this project were originally developed for one-time assessment, an important goal for future research should be the development of well investigated questionnaires that can be used for daily measures, especially in self-report with children. These scales should prove to be valid, reliable, economic, minimally disruptive, not reactive, able to capture fluctuations in the daily experiences of participants, and ideally show accordance with objective measures (Shrout & Lane, 2012). Ambulatory assessment studies which can access such resources have the potential to capture important aspects of cognitive and behavioural functioning in humans.

Especially interesting for future research might be the results from our exploratory analyses. As we found that most of our hypotheses would have been confirmed, had we only considered the inattention factor of ADHD symptoms, this discrimination of the core symptoms in research should be pursued further.

Conclusion

In the current study we examined the association of sleep quality, daytime sleepiness, and ADHD symptoms in the daily life of German schoolchildren on an inter- and an intraindividual level. A significant negative intraindividual effect was found for daytime sleepiness on ADHD symptoms within participants, contrary to the hypotheses. Explorative analyses found significant effects of sleep and sleepiness on inattention on the inter- and the intraindividual level in the expected directions: Children who sleep better on average report less inattention; On days when children report better sleep, they indicate less inattention; And children who are sleepier on average during the day report more inattention. For hyperactivity/impulsivity we found an opposite effect to our expectations: in moments when children indicate to be sleepier during the day, they report less hyperactive/impulsive symptoms. We conclude that future research should preserve the advantages concerning ecological validity which the ambulatory assessment entails and possibly integrate it with the benefits that more objective measurements like actigraphy might add. Studies examining the precursors, correlations and effects of ADHD symptoms should split the construct in the two factors of attention and hyperactivity/impulsivity.

Author Contributions: Conceptualization, Caterina Gawrilow; Data curation, Ulrike Schwarz and Jan Kühnhausen; Formal analysis, Tomasz Moschko and Anne Eppinger Ruiz de Zarate; Funding acquisition, Caterina Gawrilow; Investigation, Ulrike Schwarz; Methodology, Lilly Buhr, Tomasz Moschko, Anne Eppinger Ruiz de Zarate, Ulrike Schwarz, Jan Kühnhausen and Caterina Gawrilow; Project administration, Ulrike Schwarz; Resources, Caterina Gawrilow; Supervision, Caterina Gawrilow; Visualization, Lilly Buhr and Tomasz Moschko; Writing – original draft, Lilly Buhr, Tomasz Moschko, Anne Eppinger Ruiz de Zarate and Ulrike Schwarz; Writing – review & editing, Jan Kühnhausen and Caterina Gawrilow.

All authors have read and agreed to the published version of the manuscript.

Funding: The project was funded by the German Research Foundation (project number GA 1277/9-1). We acknowledge support by Open Access Publishing Fund of University of Tübingen and the LEAD Graduate School & Research Network.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the ethics committee of the German Society for Psychology (DGPs, CG 102018_amd_112013). The Ministry of Culture, Youth, and Sport in Baden-Württemberg, Germany, approved recruitment in schools (file number 31-6499.20/1087).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study and at least one legal guardian (parent).

Data Availability Statement: The authors work together with the Leibniz Institute for Psychology Information (ZPID; http://www.zpid.de/index.php?lang=EN) to archive and document the data collected in the project leading to this study. The data gathered in this project will be made available for the scientific community two years after the final project report has been written.

Acknowledgments: We thank all children, their parents and teachers for participation in this study. Additionally, we thank Merle Reuter and our student assistants for their help with data collection and preparation.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table 4

Multilevel model to test the between- and within-person association between children's night sleep quality and ADHD symptom levels, separately for symptoms of hyperactivity-impulsivity and inattention, the following day.

		Hyperactivity-impulsivity (n = 1436)				Inattention (n = 1438)			
Fixed Effects		Estimate		*SE*	*p*	Estimate		*SE*	*p*
Burst 1									
I Intercept: initial level	γ_{00}	3.12		2.29	.17	1.55		1.24	.21
Time slope[a]	γ_{01}	-0.18		0.15	.24	-0.28	*	0.12	.02
Night sleep quality, between-person differences	γ_{02}	-0.15		0.12	.25	-0.17	*	0.07	.02
Night sleep quality, within-person fluctuations	γ_{03}	0.03		0.03	.35	-0.06	*	0.02	.01
Weekend effect	γ_{04}	-0.10		0.08	.21	0.02		0.06	.78
Change at Burst 2, compared to Burst 1									
Change in level	γ_{10}	-0.007		0.14	.96	-0.05		0.10	.63
Change in time slope	γ_{11}	-0.48	*	0.23	.04	-0.006		0.17	.97
Change in effect of night sleep quality (between-person)	γ_{12}	0.14		0.10	.20	0.23	**	0.07	.001
Change in effect of night sleep quality (within-person)	γ_{13}	-0.06		0.05	.23	0.05		0.04	.15
Change in weekend effect	γ_{14}	0.13		0.12	.31	-0.03		0.09	.74
Change at Burst 3, compared to Burst 1									
Change in level	γ_{20}	-0.32	*	0.15	.03	-0.14		0.11	.20
Change in time slope	γ_{21}	-0.28		0.25	.25	0.20		0.18	.26
Change in effect of night sleep quality (between-person)	γ_{22}	0.06		0.12	.59	0.20	**	0.08	.009
Change in effect of night sleep quality (within-person)	γ_{23}	-0.05		0.06	.40	0.07		0.04	.07
Change in weekend effect	γ_{24}	0.01		0.13	.94	-0.09		0.10	.37
Control variables									
Gender	γ_{30}	0.20		0.20	.32	0.02		0.11	.82
Age[b]	γ_{31}	-0.009		0.02	.61	-0.0003		0.01	.97
ADHD medication[c]	γ_{32}	0.27		0.34	.42	0.33		0.18	.07

Random Effects & Covariances		Estimate		p	Estimate		p[d]
Level 2 (between-person)							
Intercept: initial level	$SD(u_{0i})$	0.77	***	<.001	0.57	***	<.001
Time slope	$SD(u_{1i})$	0.004		.77	0.46		.37
Sleep quality within-person fluctuations	$SD(u_{2i})$	0.06		.77	0.04		.96
Intercept and time	$r(u_{0i}, u_{1i})$	0.006		.95	-0.80	***	<.001
Intercept and sleep quality fluctuations	$r(u_{0i}, u_{2i})$	-0.44		.18	-0.89	*	.05
Time and sleep quality fluctuations	$r(u_{1i}, u_{2i})$	-0.02		.84	-0.81		.93
Level 1 (within-person)							
Residual	$SD(\varepsilon_{it})$	0.90			0.65		
Autocorrelation	ρ	0.33	***	<.001	0.27	***	<.001

Note. *p < 0.05, **p < 0.01, ***p < 0.001

[a] Time is coded 0 = study day 1 within a measurement burst, 1 = study day 18 within a measurement burst, with equal intervals for the intervening study occasions

[b] We were not able to collect data on one child's age. To avoid it falling out from data analysis, we set its age to the sample mean age

[c] We were not able to collect data on one child's medication status. To avoid it falling out from data analysis, we assumed it was not receiving ADHD medication

[d] The respective p-values for the random effect estimates were obtained by testing in pairs a model that includes the parameter in question against a model missing just this parameter via likelihood ratio tests.

Table 5

Multilevel model to test the between- and within-person association between children's daytime sleepiness and concurrent ADHD symptom levels, separately for symptoms of hyperactivity-impulsivity and inattention.

		Hyperactivity-impulsivity (n = 5541)				Inattention (n = 5525)			
Fixed Effects		Estimate		*SE*	*p*	Estimate		*SE*	*p*
Burst 1									
I Intercept: initial level	γ_{00}	2.63	*	1.32	.05	1.19		1.03	.24
Time slope[a]	γ_{01}	-0.19		0.10	.05	-0.14		0.08	.09
Daytime sleepiness, between-person differences	γ_{02}	0.04		0.12	.71	0.27	**	0.09	.002
Daytime sleepiness, within-person fluctuations	γ_{03}	-0.07	***	0.02	<.001	-0.01		0.02	.48
Weekend effect	γ_{04}	0.09	*	0.05	.03	0.03		0.03	.45
Change at Burst 2, compared to Burst 1									
Change in level	γ_{10}	-0.19	*	0.07	.01	-0.08		0.06	.16
Change in time slope	γ_{11}	-0.31	*	0.12	.01	0.18		0.11	.10
Change in effect of daytime sleepiness (between-person)	γ_{12}	-0.006		0.08	.94	-0.23	***	0.06	<.001
Change in effect of daytime sleepiness (within-person)	γ_{13}	0.01		0.03	.66	0.02		0.02	.39
Change in weekend effect	γ_{14}	-0.12		0.07	.09	-0.09		0.05	.08
Change at Burst 3, compared to Burst 1									
Change in level	γ_{20}	-0.38	***	0.08	<.001	-0.12		0.06	.06
Change in time slope	γ_{21}	-0.15		0.13	.25	0.27	*	0.12	.02
Change in effect of daytime sleepiness (between-person)	γ_{22}	0.08		0.08	.35	-0.15	*	0.06	.01
Change in effect of daytime sleepiness (within-person)	γ_{23}	0.02		0.03	.44	0.04		0.02	.06
Change in weekend effect	γ_{24}	-0.17	*	0.07	.03	-0.01		0.06	.81
Control variables									
Gender	γ_{30}	0.06		0.12	.64	-0.02		0.09	.84
Age[b]	γ_{31}	-0.006		0.01	.55	0.002		0.008	.83
ADHD medication[c]	γ_{32}	0.69	***	0.19	<.001	0.46	**	0.15	.003

Random Effects & Covariances		Estimate		p^b	Estimate		p^d
Level 2 (between-person)							
Intercept: initial level	$SD(u_{0i})$	0.68	***	<.001	0.46	***	<.001
Time slope	$SD(u_{1i})$	0.45	***	<.001	0.46	***	<.001
Sleepiness within-person fluctuations	$SD(u_{2i})$	0.11	***	<.001	0.08	***	<.001
Intercept and time	$r(u_{0i}, u_{1i})$	-0.17		.19	-0.61	***	<.001
Intercept and sleepiness fluctuations	$r(u_{0i}, u_{2i})$	-0.93	***	<.001	-0.29		.17
Time and sleepiness fluctuations	$r(u_{1i}, u_{2i})$	0.02		.84	0.10		.73
Level 1 (within-person)							
Residual	$SD(\varepsilon_{it})$	0.85			0.62		
Autocorrelation	ρ	0.24	***	<.001	0.27	***	<.001

Note. *p < 0.05, **p < 0.01, ***p < 0.001.
[a] Time is coded 0 = study day 1 within a measurement burst, 1 = study day 18 within a measurement burst, with equal intervals for the intervening study occasions
[b] We were not able to collect data on one child's age. To avoid it falling out from data analysis, we set its age to the sample mean age
[c] We were not able to collect data on one child's medication status. To avoid it falling out from data analysis, we assumed it was not receiving ADHD medication
[d] The respective p-values for the random effect estimates were obtained by testing in pairs a model that includes the parameter in question against a model missing just this parameter via likelihood ratio tests.

Appendix B

Responding to a reviewer's comment, we repeated multilevel modelling on both outcome variables (Night sleep quality, and daytime sleepiness), respectively, including an effect of the intervention which children received before Burst 2, γ_{33}. This indicated the mean difference in ADHD symptom levels between children receiving a full self-regulation intervention (coded 1) and children receiving a reduced self-regulation intervention (coded 0) from Burst 2 onwards. As children who did not participate in Burst 2 (*n* = 21) were not allocated to any of these intervention conditions, the models were computed with a reduced sample size of *n* = 49 children. In general, the main results of our analyses remained unchanged after controlling for the intervention effect (see Table 6 and Table 7)

Table 6

Multilevel model to test the between- and within-person association between children's night sleep quality and ADHD symptom levels the following day.

Fixed Effects		Estimate		*SE*	*p*
Burst 1					
I Intercept: initial level	γ_{00}	1.96		1.96	.32
Time slope[a]	γ_{01}	-0.36	*	0.16	.03
Night sleep quality, between-person differences	γ_{02}	-0.24		0.12	.06
Night sleep quality, within-person fluctuations	γ_{03}	-0.05		0.03	.09
Weekend effect	γ_{04}	-0.03		0.07	.70
Change at Burst 2, compared to Burst 1					
Change in level	γ_{10}	-0.08		0.11	.45
Change in time slope	γ_{11}	-0.16		0.20	.42
Change in effect of night sleep quality (between-person)	γ_{12}	0.18	*	0.08	.03
Change in effect of night sleep quality (within-person)	γ_{13}	0.02		0.04	.57
Change in weekend effect	γ_{14}	0.02		0.09	.79
Change at Burst 3, compared to Burst 1					
Change in level	γ_{20}	-0.32	**	0.12	.008
Change in time slope	γ_{21}	0.10		0.21	.63
Change in effect of night sleep quality (between-person)	γ_{22}	0.12		0.09	.21
Change in effect of night sleep quality (within-person)	γ_{23}	0.05		0.04	.25
Change in weekend effect	γ_{24}	-0.08		0.10	.40
Control variables					
Gender	γ_{30}	0.20		0.17	.26
Age[b]	γ_{31}	-0.0008		0.02	.96
ADHD medication[c]	γ_{32}	0.06		0.29	.85
Intervention	γ_{33}	-0.02		0.17	.92
Random Effects & Covariances		Estimate			*p*[d]
Level 2 (between-person)					
Intercept: initial level	$SD(u_{0i})$	0.68	***		<.001
Time slope	$SD(u_{1i})$	0.56			.42
Sleep quality within-person fluctuations	$SD(u_{2i})$	0.08	*		.01
Intercept and time	$r(u_{0i}, u_{1i})$	-0.60	***		<.001
Intercept and sleep quality fluctuations	$r(u_{0i}, u_{2i})$	-0.72	***		<.001
Time and sleep quality fluctuations	$r(u_{1i}, u_{2i})$	0.87			.90
Level 1 (within-person)					
Residual	$SD(\varepsilon_{it})$	0.64			
Autocorrelation[d]	ρ	0.37	***		<.001

Note. N = 49 children, n = 1212 considered observations, *p < .05, **p < .01, ***p < .001
[a] Time is coded 0 = study day 1 within a measurement burst, 1 = study day 18 within a measurement burst, with equal intervals for the intervening study moments

[b] We were not able to collect data on one child's age. To avoid it falling out from data analysis, we set its age to the sample mean age
[c] We were not able to collect data on one child's medication status. To avoid it falling out from data analysis, we assume it was not receiving ADHD medication
[d] The respective p-values for the random effect estimates were obtained by testing in pairs a model that includes the parameter in question against a model missing just this parameter via likelihood ratio tests.

Table 7

Multilevel model to test the between- and within-person association between children's daytime sleepiness and concurrent ADHD symptom levels.

Fixed Effects		Estimate		*SE*	*p*
Burst 1					
I Intercept: initial level	γ_{00}	0.94		1.38	.50
Time slope[a]	γ_{01}	-0.21	*	0.10	.03
Daytime sleepiness, between-person differences	γ_{02}	0.13		0.12	.28
Daytime sleepiness, within-person fluctuations	γ_{03}	-0.05	*	0.02	.01
Weekend effect	γ_{04}	0.09	*	0.04	.03
Change at Burst 2, compared to Burst 1					
Change in level	γ_{10}	-0.16	*	0.06	.01
Change in time slope	γ_{11}	-0.06		0.11	.58
Change in effect of daytime sleepiness (between-person)	γ_{12}	-0.12		0.06	.05
Change in effect of daytime sleepiness (within-person)	γ_{13}	0.02		0.02	.47
Change in weekend effect	γ_{14}	-0.13	*	0.05	.03
Change at Burst 3, compared to Burst 1					
Change in level	γ_{20}	-0.30	***	0.07	<.001
Change in time slope	γ_{21}	0.07		0.12	.55
Change in effect of daytime sleepiness (between-person)	γ_{22}	-0.06		0.07	.38
Change in effect of daytime sleepiness (within-person)	γ_{23}	0.02		0.02	.38
Change in weekend effect	γ_{24}	-0.13	*	0.06	.04
Control variables					
Gender	γ_{30}	0.08		0.12	.54
Age[b]	γ_{31}	0.006		0.01	.57
ADHD medication[c]	γ_{32}	0.48	*	0.20	.02
Intervention	γ_{33}	-0.006		0.12	.96
Random Effects & Covariances		Estimate			*p*[d]
Level 2 (between-person)					
Intercept: initial level	$SD(u_{0i})$	0.53	***		<.001
Time slope	$SD(u_{1i})$	0.40	***		<.001
Sleepiness within-person fluctuations	$SD(u_{2i})$	0.08	***		<.001
Intercept and time	$r(u_{0i}, u_{1i})$	-0.26			.11
Intercept and sleepiness fluctuations	$r(u_{0i}, u_{2i})$	-0.80	***		<.001
Time and sleepiness fluctuations	$r(u_{1i}, u_{2i})$	0.13			.69
Level 1 (within-person)					
Residual	$SD(\varepsilon_{it})$	0.64			
Autocorrelation	ρ	0.31	***		<.001

Note. N = 49 children, n = 4636 considered observations, *p < .05, **p < .01, ***p < .001
[a] Time is coded 0 = study day 1 within a measurement burst, 1 = study day 18 within a measurement burst, with equal intervals for the intervening study moments

[b] We were not able to collect data on one child's age. To avoid it falling out from data analysis, we set its age to the sample mean age

[c] We were not able to collect data on one child's medication status. To avoid it falling out from data analysis, we assume it was not receiving ADHD medication

[d] The respective p-values for the random effect estimates were obtained by testing in pairs a model that includes the parameter in question against a model missing just this parameter via likelihood ratio tests.

4.2. Manuscript 2: Temporal Fluctuations and Associations of Self-Regulation Skills and Working Memory Performance in the Daily Life of Schoolchildren

Submitted as:

Buhr, L., Schmiedek, F., & Gawrilow, C. (2023). Temporal Fluctuations and Associations of Self-Regulation Skills and Working-Memory Performance in the Daily Life of Schoolchildren. *Submitted to: Journal of Educational Psychology*.

Abstract

Children differ in their self-regulation ability as well as in their working-memory performance. These differences can not only be found between persons, but also within individuals, whose performances fluctuate across different time points. Research suggests that these fluctuations might have a serious impact on children's development and academic performance. Therefore, the present study investigates the temporal fluctuation of self-rated self-regulation ability and performance in two load levels of a working memory updating task in 10-12-year-old German schoolchildren (N = 70, including eight children with an ADHD diagnosis) three times a day via ambulatory assessment, as well as the respective association of self-regulation ability and working-memory performance on the between- and the within-person level. As expected, we found substantive fluctuations of self-regulation abilities and working-memory performance on different time scales (i.e., moment-to-moment, occasion-to-occasion, day-to-day). The variance of self-regulation was divided relatively even over the occasion and day timescale and working memory was split relatively equally over the moment, occasion, and day-to-day timescale levels. A positive association between self-regulation and the lower load level of the working memory task was found on the within-person level, suggesting that in moments of high working-memory performance a child also reports higher self-regulation. For the between-person level, a significant effect was found for the higher load level of the working memory task on self-regulation, however, this could be accounted for by a shared association with ADHD-related medication. Future analyses should further investigate the association between executive functions and self-regulation on a between- and within-person level.

Keywords: Self-Regulation, Working Memory, Temporal Fluctuations, Ambulatory Assessment, Children

Introduction

Most children would probably like to be constantly attentive and cognitively active to succeed in school. However, children and adolescents experience ups and downs in self-regulation skills and working-memory performance – with times when they are more attentive and times when they feel that they cannot concentrate at all. These fluctuations might have serious consequences on children's development and education. In this study, we investigate these temporal fluctuations as well as the respective association of two important concepts for task-related behavior: self-regulation and working memory.

Self-Regulation of Behavior and Cognition

Self-regulation refers to the ability to pursue a specific goal by monitoring and changing one's behavior as well as internal states to lead toward the desired aim (Carver & Scheier, 2011). This ability can thereby have many positive effects on individual life outcomes. It is considered to be related to educational and vocational success, internalizing and externalizing psychological disorders, substance dependencies, and abuse (Barkley, 2002; Caye et al., 2016; Kuriyan et al., 2013; Loe & Feldman, 2007; Sciberras et al., 2009). For example, one meta-analysis found that children who are more self-regulated at the age of 8 years show less depressive symptoms, aggressive and criminal behavior, obesity, cigarette smoking, and drug use at the age of 13 years (Robson et al., 2020).

People differ substantially in their ability to self-regulate (Bauer & Baumeister, 2011). In contrast to people with profound capacities to self-regulate, some people show significant impairment in self-regulation that might even lead to the diagnosis of a psychological disorder (Swanson et al., 2012). A well-known example of a lack of self-regulation skills is present in children with attention deficit hyperactivity disorder (ADHD). According to theory, the typical symptoms of inattention, impulsivity, and hyperactivity are supposed to stem from an internal deficit in self-regulation capacities (Barkley, 2011). The appearance of these symptoms in individuals does therefore give an indication of a lack of self-regulation abilities. If children have a higher ability to self-regulate their cognition and behavior, they should experience less inattention, impulsivity, and hyperactivity than children with lower self-regulation (Duckworth & Tsukayama, 2015). However, not only differences between people (i.e., *inter*individual, between-person differences) might explain particular observations in self-regulation, but also differences within an individual might play an important role for their performance (i.e., *intra*individual, within-person differences).

Fluctuation of Self-Regulation

While between-person differences have been extensively researched and shown significant correlations to positive life outcomes, within-person changes have rarely been investigated yet (Ludwig et al., 2016). Within-person changes can be separated into longer-term developmental change trajectories and short-term within-person variability around these (Nesselroade, 1991). The ability to self-regulate one's behavior develops during childhood. Longitudinal studies show that children differ in their development of self-regulation, meaning that we can find between-person differences in within-person developmental change (Raffaelli et al., 2005). Larger short-time fluctuations where children at one occasion are better able to self-regulate than at other occasions have been defined as within-person variability (Muraven et al., 1998). This within-person variability of self-regulation might be related to internal or external factors, such as, for instance, quantity and quality of sleep, social interactions, or physical activity (e.g., Becker et al., 2014; Gruber et al., 2012; Ludwig et al., 2016).

The examination of within-person fluctuations therefore requires repeated measurements of the concepts of interest in one individual (Nesselroade, 1991). This demand can best be met using ambulatory assessment, where identical questionnaires are administered repeatedly with help of technical devices like smartphones or tablets (Ebner-Priemer et al., 2009). Smartphone-administered ambulatory assessment measures in the daily life of children can differ substantially from controlled lab measures of the same constructs, probably since they measure the respective constructs more realistically and therefore enhances ecological validity (Sliwinski et al., 2018). Additionally, the measurement technique reduces memory bias since it asks about the current state or a short timeframe instead of several weeks or months as single questionnaires typically do. When planning an implementation of the method, researchers have to decide how often, with how much interim time, and for which longer period of time the constructs should be measured (Sliwinski et al., 2018). To enhance the levels of analysis even further, developmental researchers suggested to combine so-called “bursts” of measurement repeated over longer measurement periods (Nesselroade, 1991). With help of such a measurement burst design, not only between-person differences and within-person variance can be depicted, but also within-person change (in different aspects of within-person variability) over longer timeframes. In conclusion, if we want to better understand the between- as well as within-person fluctuations of cognitive and behavioral constructs and their respective associations, specific measurement methods like ambulatory assessment with measurement burst designs may allow investigation going beyond existing research evidence.

Although scientific evidence for fluctuations of self-regulation (for example manifested in ADHD symptoms) exists (Schmid et al., 2016), there is as of yet no consensus on the timeframes in which these changes occur. Sampling frequency as well as study durations are diverse in scientific studies, between twice per hour and twice per day for four to 30 days (Koch et al., 2021). This issue is related to crucial considerations in designing a study, where too few measurements overlook important changes in self-regulation, or conversely, too many measurements increase participant burden and research costs (Trull & Ebner-Priemer, 2019). Thus, targeting the appropriate temporal resolution of fluctuations is a necessary antecedent for research projects examining correlates of self-regulation or developing interventions.

Next to the question of the temporal resolution of the fluctuations, we might also ask about the reasons for the ongoing change. Observed fluctuations might be influenced by other internal or external factors relating to an individual. Given theoretical considerations in cognitive, differential, and developmental psychology, an underlying and necessary antecedent for self-regulation might be executive functions (Kofler et al., 2014).

Executive Functions and Working Memory Updating

For the execution of complex tasks, elaborate cognitive functions are necessary. Those are generally summarized under the umbrella term of executive functions (Miyake et al., 2000). One of those executive functions is working memory, which refers to the capability of an individual to keep information in mind while at the same time manipulating it according to task demands (Swanson & Alloway, 2012). For example, when doing mental math, one must remember different terms and operations, apply the operations, and update intermediate results in working memory. In this manner, working memory is important for numerous everyday and educational activities (Blume et al., 2022; Lechuga et al., 2016).

Individual working-memory performance is usually measured with working memory updating tasks (e.g., Dirk & Schmiedek, 2016; Riediger et al., 2011). Originally, objective working memory updating tasks were conducted in the laboratory, where external factors like noise and distractions are highly standardized. In their daily life, however, people need working memory in very different contexts with varying external factors, which might influence the working-memory performance that individuals show. Although working memory has been extensively researched in cognitive psychology and neuroscience, its expression in daily life still needs more empirical attention (Dirk & Schmiedek, 2016).

The combination of highly standardized objective measures with external influences of real-life measurements might therefore show advantages to assessments in the laboratory. With help of very short tasks that are administered repeatedly in the daily lives of participants through technical devices (e.g., smartphones, tablets), a more comprehensive picture of typical performance can be drawn. Instead of measuring the maximal performance, where participants in the laboratory try to show the best output they can, ambulatory assessment measures the typical performance in the daily life of participants which might be influenced much more by internal and external factors (Cronbach, 1960). This also includes the variance of repeated measures for every single person. Consequently, working memory updating tasks have already been administered in research in a daily life setting (e.g., Dirk & Schmiedek, 2016; Riediger et al., 2011).

Another advantage of these daily measures is the recording of fluctuation over time. Like studies concerning self-regulation, research on working memory is increasingly taking the fluctuations within individuals into account – adding a state perspective to the traditional conceptualization of working memory as a trait-like ability (Ilkowska & Engle, 2010a) and highlighting the importance of depicting temporal changes (McKinney et al., 2020). The magnitude of fluctuation seems to decrease over the adult lifespan, with older adults showing more stable working-memory performance across different days than younger adults (Schmiedek et al., 2013). Interestingly, in a study with 8-11-year-old children, the fluctuations of working memory could be further separated into similarly large components of day-to-day, occasion-to-occasion (i.e., within day), and moment-to-moment (i.e., within occasion) variability (Dirk & Schmiedek, 2016; Galeano-Keiner et al., 2022). Therefore, it seems that low working memory at one time point might be "a combination of bad moments, bad occasions, and, at least for some children, bad days" (Dirk & Schmiedek, 2016).

To explain antecedents and consequences of these fluctuations, the association between working memory and other external and internal alterations is investigated. Working memory performance has been shown to be related to other psychological constructs, like affect or sleep (Brose et al., 2012, 2014; Könen et al., 2015). According to theory, working-memory performance should also be linked with self-regulation; however, the empirical support for this hypothesis is varied.

Association of Self-Regulation and Working Memory

Several theories hypothesize that self-regulation and working memory are interrelated within individuals. For instance, Blair and Ursache (2011) propose a bidirectional model of executive functions and self-regulation. According to them, self-regulation seems to rely on executive functions, but over time, the development of executive functions seems to be highly influenced by self-regulation abilities. Building on this and additional theories, Bailey and Jones (2019) developed an integrated model in which executive functions constitute the core processes that allow core regulatory mechanisms to develop (Schmidt et al., 2022). Despite these theoretical indications, research has found mixed evidence concerning the association between self-regulation and executive functions. In some studies, people with an ADHD diagnosis performed worse than healthy controls on executive function tasks (Alderson et al., 2013; Campez et al., 2020; Willcutt et al., 2005). However, effect sizes were medium and several studies did not find any associations between the constructs (Willcutt et al., 2005). One explanation for this heterogeneity of findings might lie in the implementation of the conducted studies. As described earlier, most studies either measure executive functions with cognitive tasks in the laboratory or rely on self-reports of executive functions in daily life, which needs very high introspection and might lead to memory bias (Barkley & Murphy, 2011). Research has for example shown that executive functions do correlate more strongly with ADHD symptoms when reported in parent-rated questionnaires than when measured with standardized tests in the laboratory (Barkley & Murphy, 2011).

Additionally, most studies did compare self-regulation and executive functions on the trait level of between-person differences. However, the association between the constructs might be more pronounced on the within-person level, where on occasions with better executive functions a person might also show better self-regulation. The differentiation of the between- and the within-person level in research has theoretical and practical implications (Wang & Maxwell, 2015). Associations between two variables might differ in size, but even in direction on the two levels (Molenaar & Campbell, 2009). Theory suggests that working memory might assist self-regulation in a specific situation in several ways: goals that require self-regulation are represented more actively, attention is focused on the goal by top-down processes, interference is concealed, hindering thoughts are suppressed, and unwanted emotions are controlled (Hofmann et al., 2012). Furthermore, working memory and self-regulation are assumed to use the same resources, namely the active control of attention, and, therefore, might be highly interrelated (Ilkowska & Engle, 2010b). This situational perspective argues for an examination of the association between self-regulation and working-memory performance on the within-person level. As discussed above, the requirements of real-life measurement and separating the between- and within-person associations of variables can be met using ambulatory assessment. To our knowledge, only one previous study has yet examined the association of self-regulation and working-memory performance on the between- as well as the within-person level in real-life (Blume et al., 2022). In a sample of 198 schoolchildren of Grade 4 and 5, they found no association between self-regulation and working-memory performance on one of the levels. However, self-regulation was assessed with only one item, asking about the current focus ("I am concentrating right now"). Because such a narrow operationalization may have limited reliability and validity, we consider it important to continue the research with more comprehensive measurements of self-regulation. Taken together, measuring self-regulation and working memory with help of ambulatory assessment might help to further investigate their respective associations on the between- and within-person level (Koch et al., 2021).

Current Study

To investigate the temporal fluctuations of self-regulation ability and working-memory performance as well as their respective association, data from an ambulatory assessment study with measurement burst design of schoolchildren aged 10-12 years was analysed. Two overarching research questions are addressed in the current study: (1) What is the underlying temporal composition of the fluctuations of the two constructs self-regulation and working memory in children? We expect significant fluctuations on the occasion-to-occasion and day-to-day level in self-regulation and on the moment-to-moment, occasion-to-occasion, and day-to-day level in working-memory performance with substantial differences between participants. (2) How are self-regulation and working memory associated with each other on the between- as well as on the within-person level in children? We expect to find positive correlations between the two constructs on both levels.

Method

Data for these analyses stems from the project "AttentionGO! - Adaptive dynamics of cognitive and behavioural variability in children in their symptoms of ADHD", an intensive longitudinal study that was conducted at the Department of School Psychology at the University of Tübingen in cooperation with Goethe University, Frankfurt. The project was funded by the German Research Foundation (project number GA 1277/9-1) and approved by the ethics committee of the German Psychological Society (DGPs, CG 102018_amd_112013). The Ministry of Culture, Youth, and Sport in Baden-Württemberg, Germany, approved recruitment in schools (file number 31-6499.20/1087).

Design

The study was conducted employing an intensive longitudinal measurement burst design with multiple informants. Children participated in up to three bursts of 18 days each between autumn 2017 and autumn 2018, with a gap of several months in between the bursts. Each day contained three occasions of assessment, in the morning when waking up, in the afternoon after school and in the evening before going to bed. In addition to ambulatory assessments on each day, measures included questionnaires before and after each burst. Besides the children, parents and teachers were additionally surveyed, with the present study focusing on the data contributed by the children. The project aimed at depicting the dynamics of developmental changes in attention, impulsivity, and hyperactivity with a special focus on self-regulation. Additionally, external and internal factors conceivably influencing self-regulation were investigated (e.g., time spent in nature, affect, social interactions).

Recruitment

We contacted public schools in South-Western Germany and invited them to participate in our study. The German school system relies mainly on teaching different performance levels in different schools. To prevent a skewed distribution of cognitive abilities (e.g., intelligence, self-regulation), especially schools that teach intermediate or all levels were approached. Seven schools agreed to participate, five secondary schools with intermediate level ("Realschule") and two comprehensive schools

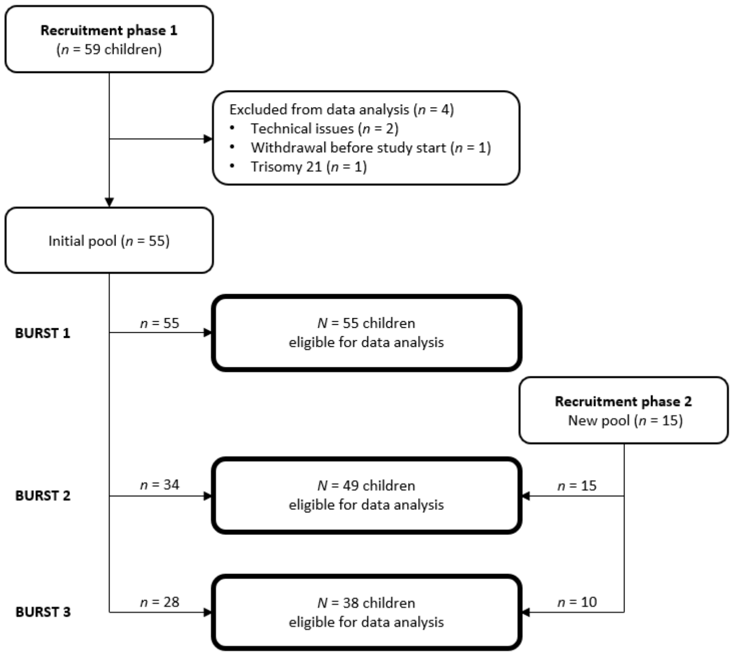
("Gesamtschule"). In the participating schools, all children from Grade 5 as well as their parents were informed about the study. Participants were only included in the study procedure if a consent form signed by parents and children was received. Participants were informed that they could drop out of the study at any time point without explanation and any negative consequences. All personal data of the participants was pseudonymized and stored with password protection on internal servers of the University of Tübingen. The study protocol was administered in accordance with the Basic Data Protection Regulation (DSGVO). Participants were informed beforehand about their rights concerning data inspection and deletion. All participating children and their families received a €40 voucher for an excursion to a regional theme park or zoo after completion of the study.

Participants

A total of 70 children participated in the study (30 male). At the beginning of the study, all children attended Grade 5 and the age range of the children was 10-12 years (*M* = 10;9 years, *SD* = 5.7 months). Parents indicated an ADHD diagnosis for eight of the participating children. All children with an ADHD diagnosis were receiving medical treatment with methylphenidate. Diagnosis of another psychological disorder led to exclusion from the study. Due to dropout, a second recruitment process was undertaken before the second burst of the study. An overview of recruitment, dropout, and retention throughout the study is depicted in Figure 5.

Figure 5

Recruitment Process and Retention of Participants (Buhr et al., 2022)



Procedure

Background questionnaires were answered by parents at the start of each burst, filled out by children at the start and end of each burst, and by teachers at a convenient time during each burst. At the start of each burst, children were handed out smartphones (Moto G5 plus©, Motorola, Libertyville, Illinois) in school. Smartphones were prepared in a way that all other activities than study participation (e.g., texting, using the internet, gaming) were prevented. It was made sure that children were trained to use the smartphone and understood the working memory updating task. The ambulatory assessment period of each burst lasted for 18 days, including weekends. Smartphones rang three times per day, in the morning after waking up, directly after school, and in the evening before going to bed. Children and parents could indicate their preferred time slots within pre-defined ranges for all three measurement occasions. After every signal of the smartphone, children had 30 min to answer the questions, otherwise, the occasion was indicated as missing. Responses of the children could therefore be analysed at the timescale level of moments (within each occasion), occasions (morning, noon, evening), days, and the three measurement bursts.

The three bursts were conducted approximately half a year apart and consisted of a similar study protocol for each burst. However, before the second burst, an intervention to enhance self-regulation was implemented. Half of the children received the full intervention called mental contrasting with implementation intentions (MCII), where they imagined a personal wish and the positive outcomes that it would lead to, followed by defining possible obstacles and formulating an if-then plan to overcome these impediments (e.g., Duckworth et al., 2013). The other half of the children received a reduced intervention including just visualizing their wish and the positive outcomes. This experimental design was chosen due to ethical reasons to not deny children in the control group potential positive effects of a self-regulation intervention. Analyses showed that in total children reported an improvement in their self-regulation abilities with no significant difference between the two self-regulation conditions (Schwarz & Gawrilow, 2019).

The study incorporated additional measurements, which will not be reported here, for example, physical activity via actigraphs. Interested readers can find additional analyses in further publications (Buhr et al., 2022; Eppinger Ruiz de Zarate & Gawrilow, in preparation; Hilger et al., 2020; Moschko et al., 2022; Reuter et al., 2020, 2021; Schwarz & Gawrilow, 2019).

Measures

Self-Regulation Scale

At each measurement occasion, children were asked to indicate their self-regulation capacity at that specific point in time. For that aim, the questions of the Conners C3-AI (Lidzba et al., 2013), which measures symptoms of ADHD (i.e., inattention, impulsivity, hyperactivity), and the short form of the Self-Control Scale (SCS-KD; Bertrams & Dickhäuser, 2009) were adapted for daily measurement. Instead of asking how the specific items were matched with the experiences in the last month, children were asked to indicate how the description conformed with their experience since the last measurement occasion. The original questionnaires were answered on 4-point (Conners C3-AI) and 5-point (SCS-KD) Likert scales. To enable a wider range of answer possibilities and thereby better depict the variance of the daily experiences as well as to prevent the tendency to choose the mean value, both answer scales were changed to 6-point Likert scales ranging from *not at all* to *exactly*. For easier interpretation, all answers were recoded before the analyses, so that low values indicate low self-regulation and high values indicate high self-regulation.

Table 8

List of Items comprising the Self-Regulation Scale.

Item (English translation)	Item (German version)
Since the last time I filled out the form…	Seit dem letzten Ausfüllen…
… I talked too much.	… habe ich zu viel geredet[1]
… I occasionally forgot what I was supposed to do.	… habe ich zwischendurch vergessen, was ich eigentlich tun sollte[1]
…I had too much energy to sit still.	… habe ich zu viel Energie gehabt, um still zu sitzen[1]
… I could hardly concentrate.	… habe ich mich schlecht konzentrieren können[1]
… I did something that I regretted afterwards.	… habe ich was gemacht, was ich danach bereut habe[2]
… I was lazy.	… war ich faul[2]
… I was able to pull myself together *(values reverse coded for analysis)*	… konnte ich mich gut zusammenreißen[2]

Note. All items were rated on a 6-point Likert scale from (1) not at all (trifft gar nicht zu) to (6) exactly (trifft ganz genau zu); [1] item adapted from Conners C3-AI (Lidzba et al., 2013); [2] item adapted from SCS-KD (Bertrams & Dickhäuser, 2009).

Spatial Working Memory Updating Task

Working memory was measured on each measurement occasion with a spatial working memory updating task, developed for ambulatory assessment (Dirk & Schmiedek, 2016). Introduced as a game with a background story, children had to indicate the positions of little cartoon monsters, which were updated throughout the task on a 4x4 grid. In the encoding phase, two or three monsters of different colors were shown for 3000 ms and the participants had to remember each of their positions. After an interstimulus interval (ISI) of 250 ms, arrows of the corresponding color indicated in which direction the monsters were moving (updating phase). The arrows were shown in the middle of the grid for 1500 ms with an ISI of 250 ms between them. Every monster's position was updated at least once. After the sequence of updating operations, children had to indicate via touch the final positions of the monsters. For this recall phase, participants were given 2000 ms, after which they received feedback on whether their answers were correct. In total, 8 blocks of the game were presented for each study occasion, 4 blocks with 2 monsters (2-Load condition) and 4 blocks with 3 monsters (3-Load condition). Per occasion, the spatial working memory updating task took approximately 8 min to finish.

Accuracy scores were calculated for blocks and loads by dividing the number of correct answers by the number of total answers. When participants had not touched any field within the 2000ms, the answer was either indicated as wrong or as missing according to the following system: If within one block only one answer out of two in the 2-Load condition or two answers out of three in the 3-Load condition were missing, these were scored as a wrong answer. Only when all answers within the block were missing, this block was indicated as incomplete. If only one block out of the four in a load condition was complete, the other three incomplete blocks were indicated as missing. If two or three blocks within a load condition were complete, the answers in the remaining one or two (incomplete) blocks were scored as wrong. Accuracy values were multiplied by 100 to facilitate interpretation (percentage of correct answers with respect to total answers).

Control Variables

We included age, gender, and ADHD medication in the multilevel models as control variables since we expected them on basis of previous research to influence self-regulation. Self-regulation has been found to increase with age (Biederman et al., 2000; Faraone et al., 2006; Raffaelli et al., 2005). Girls seem to be more self-regulated than boys (Matthews et al., 2009). Children with an ADHD diagnosis and respective medication might show significantly reduced self-regulation abilities (Faraone, 2009). Age (in months), gender, and ADHD medication (categorical variable: yes/no) were inquired about via parent questionnaires directly before each burst. All eight children with ADHD diagnosis received medication with methylphenidate. Ravens' Standard Progressive Matrices (SPM; Horn, 2009) were included in the validity analyses as a measure of fluid intelligence. Children were assessed on Ravens' Standard Progressive Matrices in the beginning of the first participation in the study (before Burst 1 or Burst 2).

Analyses

We analyzed the data using R Version 1.4.1717 (R Core Team, 2020) and Mplus (Muthén & Muthén, 2021). To avoid bias resulting from extreme individual reports, all individual data points concerning self-regulation or working memory that were lying three standard deviations above or below the individual mean of a participant were excluded from further analysis.

For the estimation of individual variance components for the self-regulation scale and both load levels of the working memory updating task, we individually fitted one multilevel model per child for each outcome variable (based on analyses in Dirk & Schmiedek, 2016; Galeano Weber et al., 2018). Self-regulation or working-memory performance at the respective load level was used as the outcome variable, while running trial number was included as independent variable. Random intercepts were allowed for the nested time variables day, occasion, and moment (only for working memory). Consequently, total daily variance was split into components of day-to-day, occasion-to-occasion, and moment-to-moment fluctuations. Results of all children were then averaged to investigate the mean variability across components.

To investigate the association of self-regulation and working memory on the between- and the within-person level, multilevel models were fitted using data on the occasion level, combining data from the three measurement bursts to increase statistical power, and using the nlme package (version 3.1-155) in R. We fitted separate models for Load 2 and Load 3 conditions of the working memory task. Raw scores were split into two components, indicating the trait-like between-person differences (WM_B) that differentiate persons in terms of their working-memory performance on the one hand and the state-like fluctuations (WM_W) that account for better or worse performance than usual on specific occasions. Models were built in an iterative process, where the main variables were implemented in the model first and then autocorrelation parameters and control variables were added later. In the first step, a multilevel model, which included fixed and random intercepts and fixed and random slopes for time and within-person fluctuations of working memory was used. Time was integrated into the model in three different ways: one term indicating a linear change in self-regulation across measurement occasions within each burst (occasion), as well as two dummy coded variables indicating a change in self-regulation intercepts between Burst 1 and Burst 2 ($Burst_2$) and Burst 1 and Burst 3 ($Burst_3$), respectively. Moreover, we modelled two-way interactions between occasion and study burst, to test whether and how a linear change in self-regulation across measurement occasions might differ between study bursts. Possible autocorrelation due to the repeated measures structure of the data was accounted for by modelling the time dependence of the residuals with a first-order autoregression (AR1) in a second step. In a third step, we included age, gender, and ADHD medication as control variables, due to previous research indicating an influence of these variables on self-regulation (e.g., Faraone, 2009; Faraone et al., 2006; Matthews et al., 2009). Formally, this model is described by the following equations:

Level 1:

$$SR_{ij} = b_{0i} + b_{1i}occasion_{ij} + b_{2i}Burst2_{ij} + b_{3i}Burst3_{ij} + b_{4i}WMW_{ij} + b_{5i}occasion_{ij} * Burst2_{ij} + b_{6i}occasion_{ij} * Burst3_{ij} + \varepsilon_{ij}$$

Level 2:

$$b_{0i} = \beta_0 + \beta_1 WMB_i + \beta_2 age_i + \beta_3 gender_i + \beta_4 medication_i + \gamma_{0i}$$

$$b_{1i} = \beta_5 + \gamma_{1i}$$

$$b_{2i} = \beta_6$$

$$b_{3i} = \beta_7$$

$$b_{4i} = \beta_8 + \gamma_{2i}$$

$$b_{5i} = \beta_9$$

$$b_{6i} = \beta_{10}$$

The models were calculated using maximum likelihood estimation (ML). All analyses used an alpha level of $\alpha = .05$.

Transparency and Openness

Following the Journal Article Reporting Standards (JARS; Kazak, 2018), we report our recruitment, handling of outliers, all measures used for the analyses and all analyses. Materials and analysis code for this study are available under https://osf.io/bm6h2/?view_only=2b18f66259a64ace97582ec53c6e1883.

This study's design and its analysis were not pre-registered.

Results

Descriptives

To prevent inflating the dropout rate due to the divided recruiting process, possible observations were calculated for three bursts multiplied by the 18 days and 3 occasions for the 55 children that were recruited in November 2017 and calculated for two bursts for the 15 children who started in April 2018. Consequently, a maximum amount of 10,530 occasions could have been collected for all participants. Due to technical failure, the working memory task was not presented on the last day (three occasions) at the end of Burst 1, resulting in a total of 10,035 possible observations for working memory. Children indicated their self-regulation on 5,764 occasions, resulting in a participation rate of 54.7%. For Load 2 of the working memory task, 5,414 observations were available (54.0%) while Load 3 was available for 5,411 occasions (53.9%). Since the working memory task was presented in four blocks per occasion for each load level, on the block level there were 40,140 observations possible. Children reacted to a total of 21,584 blocks in Load 2 (53.8%) and 21,586 blocks in Load 3 (53.8%).

Before conducting multilevel analyses, 83 self-regulation, 69 working memory Load 2, and 18 working memory Load 3 observations were excluded from the analyses because they were found to be more than three standard deviations away from the individual participant mean and therefore defined as outliers. All outliers were found on the extreme lower end of the personal values, none was found to be extremely high.

Table 9

Descriptive Statistics of Self-Regulation and Working Memory.

item	*M(SD)*	*ICC*	*M ISD (SD)*	*MSSD*
SR	5.32 (0.73)	.42	0.5 (0.25)	0.35 (0.37)
WM L2	81.34 (26.91)	.32	21.33 (9.86)	639.04 (478.35)
WM L3	58.02 (29.98)	.38	23.58 (5.39)	772.46 (382.05)

Note. SR = Self-regulation (1 = low/6 = high, items see Table 8), WM = working memory (in percent accuracy), ICC = Intraclass Correlation Coefficient, ISD = Intraindividual Standard Deviation, MSSD = Mean Square Successive Differences

In total, the children indicated a mean level of self-regulation of M = 5.32 (SD = 0.73), which constitutes a high average on a scale from 1 to 6. The mean accuracy of the working memory task was 81.34 (SD = 26.91) on Load 2 and 58.02 (SD = 29.98) on Load 3. Descriptive statistics of the constructs used for hypothesis testing can be found in Table 9. The intraclass correlation coefficient (ICC), which indicates the proportion of total variance that can be explained by between-person difference had values between .32 and .42 for the constructs. Consequently, 42% of the total variance in self-regulation can be attributed to between-person differences, the remaining 58% are explained by systematic within-person fluctuation and measurement error. Correlation between the two load conditions of the working memory tasks was r = .59 ($p < .05$) and between self-regulation and Load 2/Load 3 r = .10 / r = .15 (both $p < .05$) respectively on the within person-level. On the between-person level, correlations between the working memory loads were r = .89 ($p < .05$) and between self-regulation and Load 2/Load 3 r = .25 / r = .31 (both $p < .05$)

Psychometric Properties of the Self-Regulation Scale and Working Memory Task

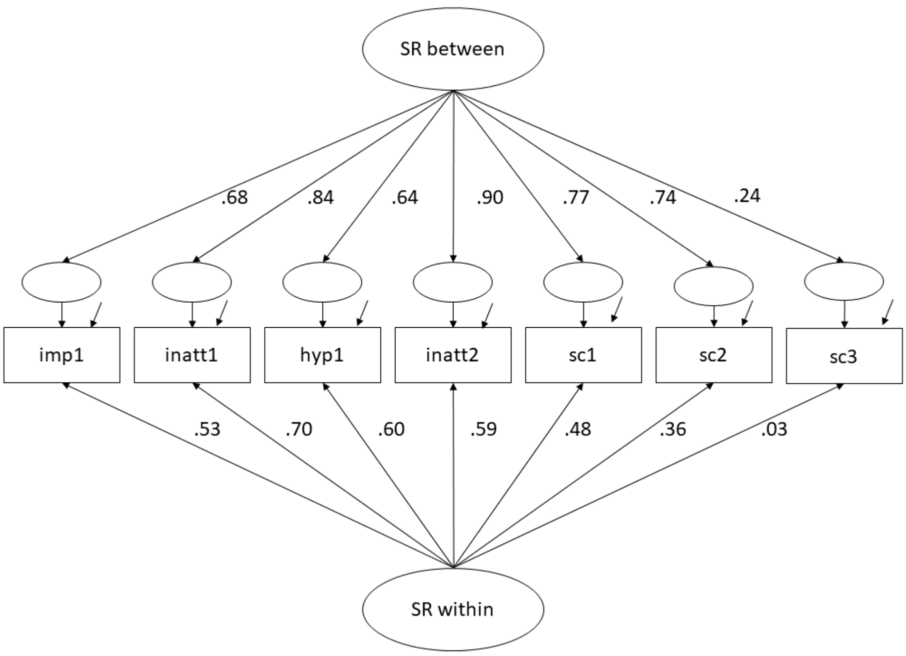
Reliability

In advance of answering our research questions, we investigated the psychometric properties of the self-regulation scale and the working memory task. Based on a two-level confirmatory one-factor model, we assessed systematic between-person and within-person reliabilities (Geldhof et al., 2014). Within these analyses, we calculated the proportion of latent variation to total variation on each level.

For self-regulation, all seven items served as indicators of latent factors at the between- as well as the within-person level of variation (Figure 6). For all items except the third self-control item, standardized loadings were high on the between-person level (.64 - .90) and moderate on the within-person level (.36 - .70). All standardized loadings were significant ($p < .001$), except for the third self-control item, which was non-significant on the between- and the within-person level. The third item of self-control was only weakly associated with the between-person latent factor (.24, $p < .001$) and did not load on the within-person latent factor (.03, $p = .40$). Overall, the composite reliability calculated from the factor loadings was good (between $\omega = .783$, within $\omega = .625$). Model fit indices indicated good fit, except for the CFI and TLI, which failed to meet conventional standards of good fit ($CFI = .892$, $TLI = .837$, $SRMR_{between} = 0.06$, $SRMR_{within} = 0.039$, $RMSEA = .027$). Allowing correlated residuals on the within-person level between the item imp1 with inatt1, inatt2 and hyp1 as well as between items hyp1 and inatt1, as proposed by the modification indices, enhanced CFI and TLI to meet these criteria ($CFI = .990$, $TLI = .983$, $SRMR_{between} = 0.054$, $SRMR_{within} = 0.017$, $RMSEA = .009$). Composite reliability on the within person level, however, lowered to $\omega = .555$ when these modifications were included.

Figure 6

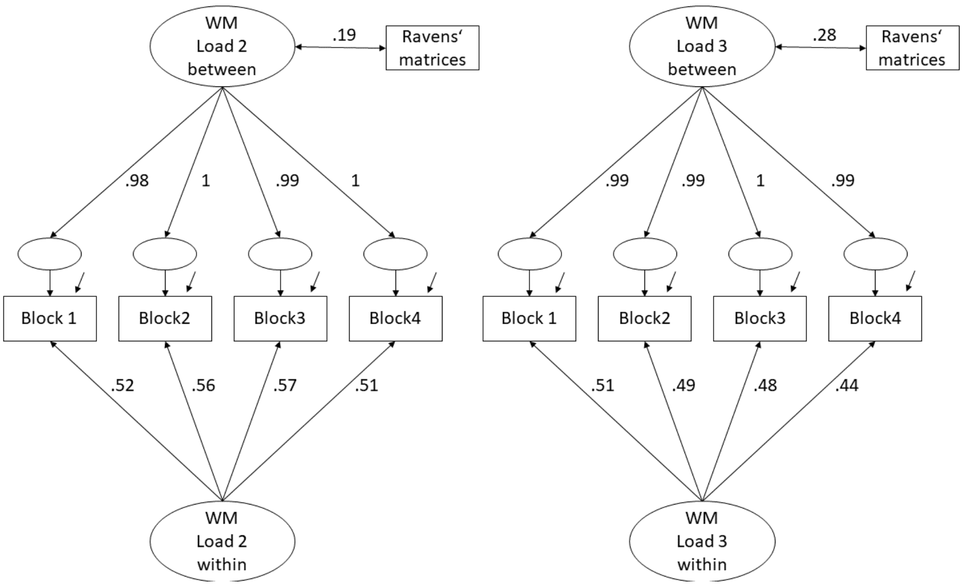
Twolevel Factor Model of Self-Regulation Items



To test for systematic between- and within-person variance in the two load conditions of the working memory task, we fitted one model per load, with the blocks within each occasion as indicator variables. Performance on Raven's Standard Progressive Matrices was included in the model on the between-person level to test for the validity of the measure, expecting a substantial positive correlation (Kyllonen & Christal, 1990). Performance on Raven's Standard Progressive Matrices was not significantly related to the working memory factor in Load 2 (r = .19, 95%-CI = -.04 - .41, p = .11) but moderately to the working memory factor in Load 3 (r = .28, 95%-CI = .08 - .48, p = .01). In Load 2 of the working memory task, composite reliability was high on the between-person level (ω = .997) and moderate on the within-person level (ω = .62) [3]. The model fit met the criteria by Hu and Bentler (1999) very well (CFI =.99, TLI =.97, $SRMR_{between}$ = 0.015, $SRMR_{within}$ = 0.013, RMSEA =.018). Similar reliabilities were found for Load 3 of the working memory task (between ω = .997, within ω = .546). The model fit criteria were also met in the model for Load 3 (CFI =.99, TLI =.98, $SRMR_{between}$ = 0.012, $SRMR_{within}$ = 0.01, RMSEA =.014).

Figure 7

Between- and Within-Person Factor of Working-Memory Performance



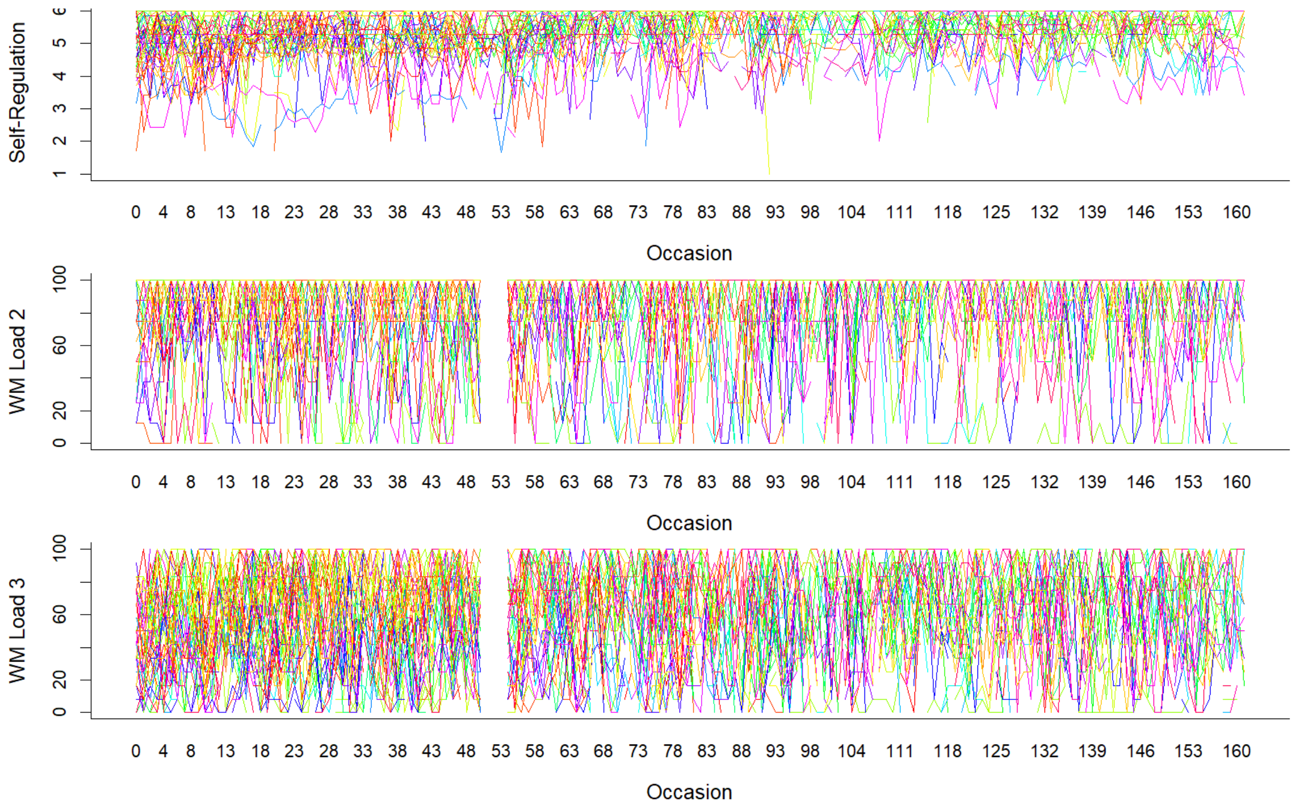
[3] As proposed by Geldhof and colleagues (2014), we included the term $2\sum_{i=2}^{k}\sum_{j=1}^{k}\theta_{ij}$ when calculating omega to account for covariances between the indicators within one measurement occasion.

Fluctuations over Time

To get a descriptive picture of within-person fluctuations of self-regulation as well as working memory, we depicted the fluctuations over time of each individual. Here again, all time points were included in the graphs without consideration of the different bursts. The individual lines show significant fluctuations within individuals for the mean of the self-regulation items and both load conditions of the working memory task (Figure 8). Self-regulation seems to be high in general for most children. Still, there is a considerable amount of time points, where children indicated low self-regulation. For working memory, a difference between the load conditions can be seen in the graph. While there is indication for ceiling effects in the Load 2 condition, accuracy in Load 3 is much more variable.

Figure 8

Fluctuations of Self-Regulation and Working Memory over Time



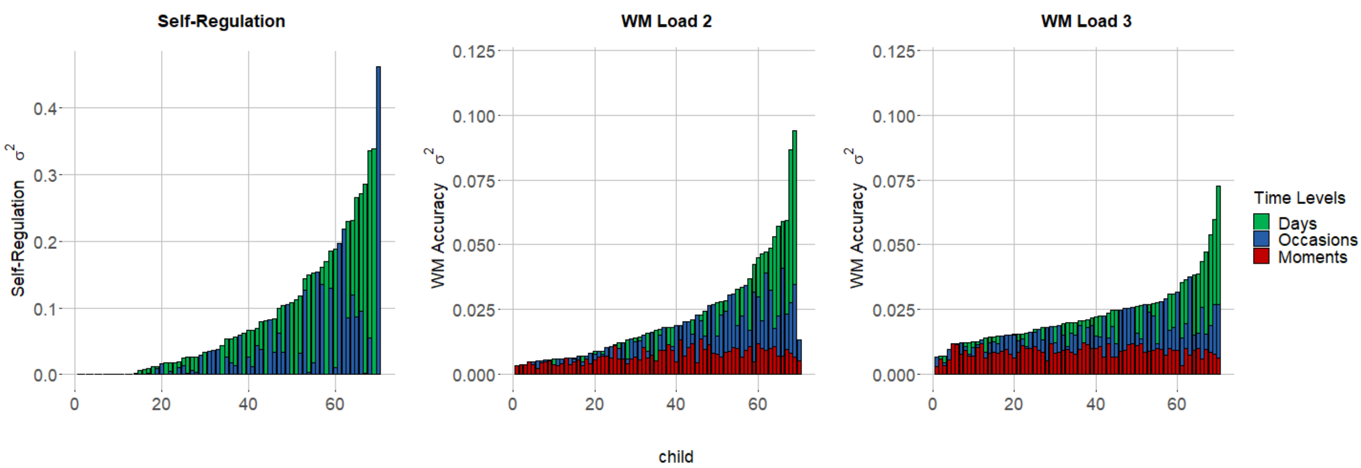
Note. WM = working memory; data on the working memory task was not available for one day (three occasions) due to technical failure, indicated by the white space in the graph.

Fluctuations of Self-Regulation and Working Memory at Different Timescales

Our first overarching research question concerned the temporal structure of the fluctuation of self-regulation and working memory. To test which of the time levels of the study (moments, occasions, and days) the fluctuations occur on, we decomposed the total daily variance of each child by time level using multilevel models (based on analyses in Dirk & Schmiedek, 2016; Galeano-Keiner et al., 2022). The moment level was only investigated for working-memory performance since the self-regulation scale was only administered once per occasion. Figure 9 shows great differences between the individual participants, both in their total amounts of variance as well as in the partition of the variance over the different timescale levels. Individuals differed substantially in their self-regulation, with some children not depicting any variance at all, some only depicting variance on the occasion level, some on both levels, and some only on the day level. All children show at least some variance in their working-memory performance on the moment level, while some children additionally show fluctuations over occasions and days and others do not.

Figure 9

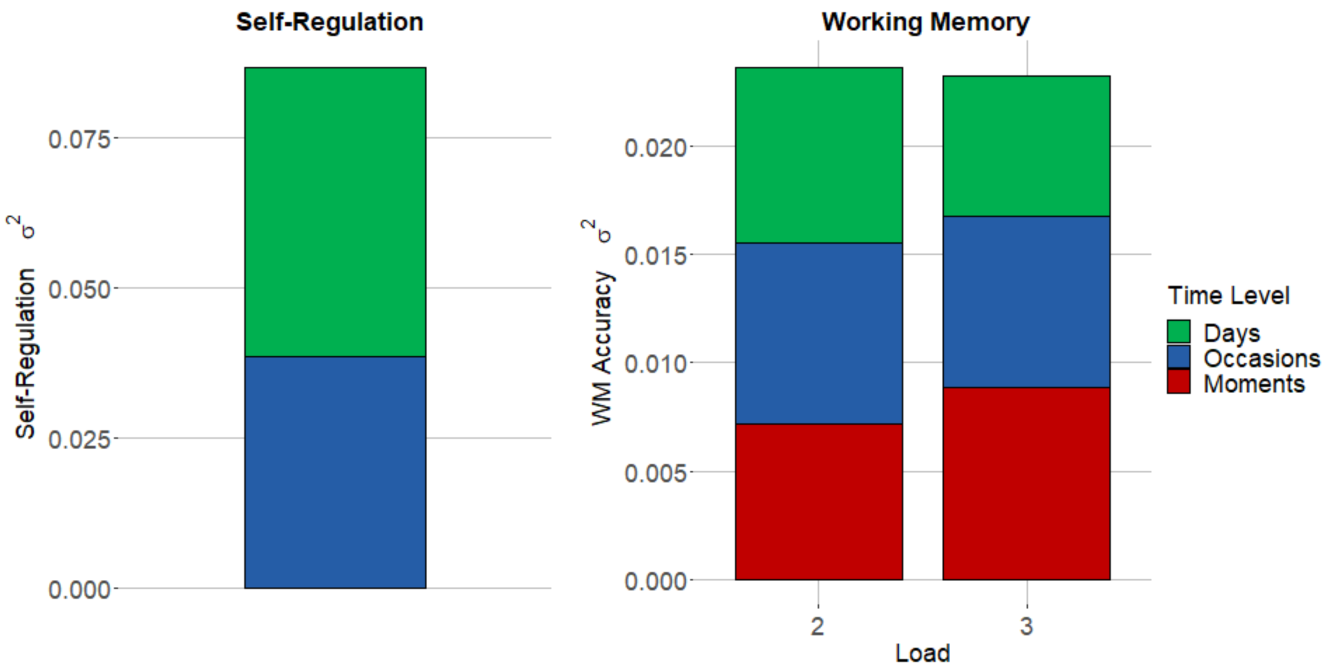
Fluctuations of Self-Regulation and Working Memory over Different Timescale Levels



We averaged the variances over all children to estimate the timescales of fluctuations within children in total (Figure 10). As can be seen in Figure 10, variance was relatively equally divided over the respective timescales. Variability in self-regulation was composed of a similar amount in occasion-to-occasion and day-to-day fluctuation. Working-memory variance was divided relatively equally in moment-to-moment, occasion-to-occasion, and day-to-day fluctuations.

Figure 10

Average Variance Components of the Fluctuations of Self-Regulation and Working Memory at Different Timescale Levels



The Association of Self-Regulation and Working Memory Between and Within Individuals

To answer our second research question, we tested for associations between self-regulation and working memory on the between- and the within-person (occasion-to-occasion) level using multilevel analyses. Models were built stepwise and model fit was compared with likelihood ratio tests. Separate models were generated for each of the two load conditions of the working memory task. In the first step, only the time variables and the between- and within-level accuracy of working memory were included in the models. The results of the models are shown in Table 10.

Table 10

Multilevel Model of the Association of Time and Working Memory with Self-Regulation.

	WM Load 2			WM Load 3		
Predictors	*Estimates*	*95% CI*	*p*	*Estimates*	*95% CI*	*p*
(Intercept)	5.135	5.017 – 5.253	<.001	5.144	5.027 – 5.260	<.001
Occasion	0.005	0.003 – 0.007	<.001	0.005	0.002 – 0.007	<.001
Burst 2	0.205	0.130 – 0.280	<.001	0.202	0.126 – 0.277	<.001
Burst 3	0.256	0.175 – 0.337	<.001	0.247	0.165 – 0.329	<.001
WM between	0.191	-0.011 – 0.392	.06	0.200	0.020 – 0.379	.03
WM within	0.031	0.004 – 0.057	.03	0.031	0.002 – 0.059	.03
Occasion * Burst 2	-0.001	-0.003 – 0.002	.48	-0.001	-0.003 – 0.002	.46
Occasion * Burst 3	-0.002	-0.005 – 0.001	.16	-0.002	-0.005 – 0.001	.18
Random Effects	*Estimates*	*Residual*		*Estimates*	*Residual*	
Intercept	0.45			0.45		
Occasion	0.007			0.007		
WM within	0.068	*0.53*		0.079	*0.53*	

Note. *N* = 70 subjects, 5100 observations

Occasion was positively associated with self-regulation in both models, indicating that children reported increasingly higher levels of self-regulation across study occasions. However, estimates of the effects were very small (both $\beta = 0.005$, $p < .001$). Burst 2 and 3 were as well positively associated with self-regulation in both models ($\beta = 0.202 – 0.256$, all $p < .001$). This means that, compared to Burst 1, children indicated higher self-regulation in the following bursts. The interaction of occasion and Burst 2 or Burst 3 was not significant in the models, indicating that the within-burst trends were of comparable size across bursts. In the Load 2 condition, working memory performance was significantly associated with self-regulation on the within-person level ($\beta = 0.031$, $p = .03$) but not reliably associated with self-regulation on the between-person level. Adding fixed and random effects of Load 2 of the working memory task allowed to explain an additional 1.2% of variance at the within-person level and 6.5% of variance at the between-person level. Load 3 of the working memory task showed a significant positive association with self-regulation on the between-person level ($\beta = 0.200$, $p = .03$) and on the within-person level ($\beta = 0.031$, $p = .03$). The inclusion of fixed and random effects of Load 3 of the working memory task allowed to explain an additional 1.4% of variance at the within-person level and 7.9% of variance at the between-person level. In a second step, we tested for the effect of autocorrelation in the data. Including a first-order autocorrelation of the residuals in the model significantly improved the model fit (χ^2 [df=1] = 348/328, both p < .001). Unfortunately, the models with the autocorrelation included did not show any within-person association between the Load 3 condition of working-memory performance and self-regulation. The other effects remained similar to the previous model.

Table 11

Multilevel Model of the Association of Time and Working Memory with Self-Regulation, Accounted for Autocorrelation.

	WM Load 2			WM Load 3		
Predictors	*Estimates*	*95% CI*	*p*	*Estimates*	*95% CI*	*p*
(Intercept)	5.131	5.010 – 5.252	<.001	5.137	5.017 – 5.257	<.001
Occasion	0.005	0.002 – 0.008	<.001	0.005	0.002 – 0.007	<.001
Burst 2	0.198	0.104 – 0.293	<.001	0.197	0.103 – 0.292	<.001
Burst 3	0.257	0.155 – 0.359	<.001	0.251	0.149 – 0.353	<.001
WM between	0.179	-0.019 – 0.377	.08	0.195	0.019 – 0.372	.03
WM within	0.033	0.006 – 0.060	.02	0.022	-0.002 – 0.047	.08
Occasion * Burst 2	-0.001	-0.004 – 0.002	.63	-0.001	-0.004 – 0.002	.61
Occasion * Burst 3	-0.002	-0.005 – 0.001	.22	-0.002	-0.005 – 0.001	.24
Random Effects & Autocorrelation	*Estimates*	*Residual*		*Estimates*	*Residual*	
Intercept	0.44			0.44		
Occasion	0.006			0.006		
WM within	0.071	*0.53*		0.061	*0.53*	
Autocorrelation	.27			.26		

Note. N = 70 subjects, 5100 observations

Finally, age, gender, and ADHD-related medication were stepwise included in the models as control variables (Table 12). For each control variable, a likelihood ratio test was conducted to examine whether the inclusion improved the model fit. No association was found between age or gender and self-regulation. Model fit was not improved by the integration of age or gender, so they were not included in the final model. Medication was significantly related to self-regulation in both load conditions of working memory performance (Load 2: β = -0.416, p = .02; Load 3: β = -.398, p = .03). According to likelihood ratio tests, including medication improved both models significantly (Load 2: χ^2 [df=1] = 5.3, p = .02; Load 3: χ^2 [df=1] = 4.9, p = .03). Including medication in the model lowered the association between working memory and self-regulation. Consequently, the between-person effect of working memory was not detectable anymore in the Load 3 condition. The only remaining significant association between self-regulation and working memory performance was a within-person effect in the Load 2 condition ($\beta = 0.03$, $p = .02$)

Table 12

Multilevel Model of the Association of Time, Working Memory and Medication with Self-Regulation, Accounted for Autocorrelation.

	WM Load 2			WM Load 3		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	5.171	5.049 – 5.293	<.001	5.176	5.054 – 5.297	<.001
Occasion	0.005	0.002 – 0.008	<.001	0.005	0.002 – 0.007	<.001
Burst 2	0.198	0.103 – 0.293	<.001	0.197	0.103 – 0.291	<.001
Burst 3	0.258	0.156 – 0.360	<.001	0.252	0.150 – 0.354	<.001
WM between	0.149	-0.043 – 0.342	.13	0.167	-0.005 – 0.339	.06
WM within	0.033	0.006 – 0.060	.02	0.022	-0.002 – 0.046	.08
Medication	-0.416	-0.769 – -0.063	.02	-0.398	-0.750 – -0.046	.03
Occasion * Burst 2	-0.001	-0.004 – 0.002	.64	-0.001	-0.004 – 0.002	.61
Occasion * Burst 3	-0.002	-0.006 – 0.001	.22	-0.002	-0.005 – 0.001	.23
Random Effects & Autocorrelation	*Estimates*	*Residual*		*Estimates*	*Residual*	
Intercept	0.42			0.42		
Occasion	0.006			0.006		
WM within	0.071	*0.53*		0.061	*0.53*	
Autocorrelation	.27			.26		

Note. N = 70 subjects, 5100 observations

Discussion

Children experience occasions of high and low self-regulation abilities as well as high and low working-memory performance during their daily lives. Regular test batteries, administered in the laboratory, or one-time assessments via questionnaires cannot capture these fluctuations and therefore might not be suitable to find connections between the constructs. In the present study, we therefore administered an intensive longitudinal design and ambulatory assessment to measure the fluctuations and interconnections of self-regulation abilities and working-memory performance in the daily life of schoolchildren at the ages between ten and twelve years.

Our first research question concerned the temporal structure of the fluctuations in self-regulation as well as working memory. Therefore, we first analysed whether we could indeed find systematic variation on the between- as well as on the within-person level. Although the within-person reliabilities were only moderate, self-regulation and working memory had significant factor loadings on common factors at the between- and within-person level – indicating systematic differences between persons as well as systematic fluctuations across days. Visual inspection of the individual variations of self-regulation and working memory over time supported the notion of large fluctuations within children. These findings are in accordance with previous research, which found substantial fluctuations over time within children and adults for self-regulation (Ludwig et al., 2016) as well as for tasks of executive functions, like working memory (e.g., Brose et al., 2012; Dirk & Schmiedek, 2017; Schmiedek et al., 2013).

As expected, we found that the variances could be decomposed into moment-to-moment (only for working-memory), occasion-to-occasion, and day-to-day components. This indicates that the response to a question about self-regulation or the accuracy on a working-memory task at one point in time might result from any combination of fluctuations at the moment, occasion, or day level. A child with high self-regulation indication might be self-regulated at the time of assessment because one caught a good time during the day, or because self-regulation is relatively high throughout the whole day. Working-memory performance could even be split into three time levels, with variance being equally divided over moment-to-moment, occasion-to-occasion, and day-to-day variance.

These findings correspond well to the evidence found by Dirk and Schmiedek (2016). They applied the same working memory updating tasks to a sample of elementary schoolchildren three times a day over four weeks. As in the present study, they found that variance in working memory could be composed into moment-to-moment, occasion-to-occasion, and day-to-day variance. Galeano-Keiner et al. (2022) could reproduce these results, when investigating children in the fifth grade twice a day.

Our second research question concerned the relation between self-regulation and working memory on the between- as well as the within-person level. In our multilevel analyses, we found significant associations between self-regulation and the different bursts of the study as well as the occasion within each burst. In Burst 2 and Burst 3, the self-reported self-regulation was significantly higher than in Burst 1. Theoretical assumptions and empirical findings point to a development of the regulation of emotions and behavior over childhood and adolescence (Biederman et al., 2000; Faraone et al., 2006; Raffaelli et al., 2005). This might explain why self-regulation increases significantly across the bursts separated by about half a year each. Additionally, all children had started fifth grade in the beginning of the study, which is linked with a change of schools in Germany. It is possible that excitement about the new school environment might have lowered self-regulation in the first burst and with increasing habituation to the new school, self-regulation might have enhanced. Previous research has shown a significant effect of schooling on executive functions and working memory which might have gradually increased self-regulation ability in the participating children (Morrison et al., 2019).

Another potential explanation for the increase in self-regulation after the first burst might be found in the study protocol. Within the project, an intervention with the aim to enhance self-regulation was conducted before the second burst of the study. Children were divided into two groups, with one group receiving the full intervention and the other group a similar but reduced intervention (Schwarz & Gawrilow, 2019). However, since children in the experimental group did not differ significantly from children in the control group concerning their self-reported self-regulation abilities, we would not expect any influence on the association between working memory and self-regulation on the between- or the within-person level.

The occasion of the study within each burst was also significantly related to the report of self-regulation, indicating an average trend of improving self-regulation. The association between occasion and self-regulation might be explained by the initial-elevation bias. Ambulatory assessment research has shown that people tend to answer more extreme in the first assessments of a study and then regress more toward an average evaluation (Shrout et al., 2018). In the present study, this would mean that children reported their self-regulation as worse in the first assessments of each burst and then lower their assessment towards a more realistic evaluation (for continuative analyses see Appendix 1). Future research should investigate how this initial response bias occurs and how it might be prevented.

Concerning our main hypothesis, we could find a positive between-person association between Load 3 of the working memory updating task and self-regulation, which vanished when medication was included in the model. This indicates that children with better average working memory also indicate higher self-regulation. This fits nicely in the theoretical perspective that executive functions like working memory are associated with self-regulation difficulties, as they are for example experienced by children with ADHD diagnosis (Barkley, 2011). Several studies comparing people with ADHD diagnoses and healthy controls in their performances on executive function tasks report similar findings (Alderson et al., 2013; Campez et al., 2020; Willcutt et al., 2005). What differentiates this study from the examples mentioned before is the general population sample. We hypothesize that self-regulation is a dimensional ability that is present in every individual with a different level. Our results indicate that these dimensional differences in self-regulation in a general population sample of German schoolchildren are in fact related to working memory updating as an executive function task.

No between-person effect could be found for the Load 2 condition of the working memory task, which can probably be explained by a ceiling effect since the task might have been too easy for the age group in the study. The two load levels had been included in the study to enhance comparability to previous research, especially to the study of Dirk and Schmiedek (2016), who used the same working memory updating task in a sample of 8-11-year-old children. However, in the future, the working memory updating task should be adapted more thoroughly to the specific age group of participants to avoid ceiling effects.

In line with our hypothesis, we found a significant within-person association of the Load 2 condition of the working memory task and self-regulation. This implies that on occasions when a child shows relatively high working-memory performance, it tends to also indicate relatively high self-regulation. In the Load 3 condition, the within-person association was only present in the model when only time and working-memory performance were included, the addition of autocorrelation and medication covered the effect. This finding of within-person associations fits nicely into the theory that working memory assists self-regulation in daily life (Hofmann et al., 2012).

Age and gender were not significantly associated with self-regulation. Since we found an effect of burst on self-regulation, it is surprising that age at the start of the study does not seem to be interrelated with self-regulation. Our sample was relatively homogeneous in age, which resulted in a low variance. Another possible interpretation is that grade in school is more predictive of self-regulation than actual age. Previous research has shown that the schooling effect on executive functions and working memory exceeds age effects (Morrison et al., 2019). With the beginning of school, abilities in executive functions and working memory grow significantly, probably due to the enriching environment and training opportunities. Since all children had just started in Grade 5 at the beginning of Burst 1, and therefore all had four years of training their self-regulation within school behind them, this same amount of schooling might level out individual age effects. We would also have expected that gender had a significant association with self-regulation, which we did not find in this sample. ADHD related medication was significantly negatively related to self-regulation. All children in the sample who had a diagnosis of ADHD were also prescribed medication (i.e., Methylphenidate). Although this medication is supposed to enhance self-regulation (e.g., Faraone & Buitelaar, 2010), the symptoms of inattention and hyperactivity/impulsivity seem to be so pronounced in children with an ADHD diagnosis that they override the positive association of working memory performance and self-regulation.

Limitations

The present study holds some significant advantages in comparison to more classical laboratory studies and one-time assessments, for example, the assessment in real-life and the separation of the between-person and the within-person levels. However, some limitations have to be discussed here that might have influenced our findings.

Creating an ambulatory assessment study is challenging (Mehl & Conner, 2012). The participant burden is already very high because of the repeated measurements. Most norm-based scales that are usually used in psychological research are not advisable for repeated measurement – first because of their wording that is mainly referring to longer timeframes and second because of the length of the scales. To reduce participant burden and enhance compliance, the self-report scales in this study were shortened significantly. In a previously conducted pilot study, only those items that showed significant variance were selected. This led to a more concise study protocol where assessment time on each occasion was below 10 min. Unfortunately, this procedure lowered the comparability of the scales with previous research and established norms. It might have also lowered the reliability of the scales on the between- and the within-person level. Since research shows an increasing interest in ambulatory assessment studies, specific scales that meet these specific requirements should be developed and tested.

Additionally, the use of self-report questionnaires for children at the age of 9-12 years can be debated as to whether children are reliable informants of their self-regulation abilities. However, this study was specifically concerned with the individual experiences of children during their daily life. Furthermore, researchers usually consider school-aged children to be reliable informants about their health and affect (e.g., Leonhardt et al., 2016; Riley, 2004). Finally, the scales we adapted for our study are, according to the authors, appropriate for children between the age of 8-18 years (Lidzba et al., 2013). Future research should nonetheless compare children's self-report of their self-regulation abilities to observer reports of parents and teachers (van der Ende et al., 2012), or even more objective measurement methods, for instance controlling for hyperactivity with actigraphy (Gawrilow et al., 2014).

When inspecting our data, we found ceiling effects for the self-regulation scale as well as the Load 2 condition of the working memory task. It seems to be a positive discovery that in a general population sample, children report few problems with self-regulation during their daily life. However, this ceiling effect lowers the variance of the data available and therefore might lead to underestimation of the expected associations. Future studies might therefore implement scales that are depicting a normal distribution of the data to measure self-regulation capacities (e.g., Schulz-Zhecheva et al., 2017). The working memory task we used has already been successfully administered in previous research with younger children aged

8-10 (Dirk & Schmiedek, 2016). In children, working-memory performance grows with increasing age (Gathercole et al., 2004; Neubeck et al., 2022). This might explain why we found a ceiling effect for Load 2 of the working memory task but not for Load 3, the first one might have been just too easy. Future research might therefore heighten the level of difficulty with increasing age of the participants.

Another potential limiting aspect of our measurements is the low association of fluid intelligence and working memory. We did find no association of the fluid intelligence measure of Ravens' Progressive Matrices with the Load 2 condition of the working memory task and only a low correlation for the Load 3 condition. A previous study found a high correlation between working-memory performance assessed with the same paradigms and fluid intelligence (e.g., Dirk & Schmiedek, 2016). However, they used the Culture Fair Intelligence Test (CFT) to measure fluid intelligence. Although the Ravens' Standard Progressive Matrices (SPM; Horn, 2009) that were used in the current study, are supposed to be suitable for children starting at the age of six years, they might not have been adequate to depict fluid intelligence in our sample. Therefore, we would suggest that future studies should further investigate the relationship of working-memory performance and different measures of intelligence.

Finally, the timeframes used for assessments are a limitation of the study. Due to instruction from the regional school board, all assessments had to take place outside of school in the children's free time. Therefore, children responded to the smartphones in the morning before school, in the afternoon after school, and in the evening before going to bed. Since self-regulation ability and working-memory performance might be especially important in an educational context, it would have been preferable to conduct the study within the children's schools (Dirk & Schmiedek, 2016; Ehm et al., 2016; Loe & Feldman, 2007). Working together with teachers and political stakeholders, researchers should therefore try to find a way to assess self-regulation and working memory in the school context without disturbing educational and learning processes.

Implications and Future Analyses

The current study has shown that we need more research concerning self-regulation and working memory, especially on the within-person level. We have found that both self-regulation and working memory fluctuate over time within individuals. Therefore, future studies should take temporal variability into account when looking at antecedents, correlates, and consequences of self-regulation and working memory. To make this possible, reliable and valid scales and tasks have to be developed that depict the daily variance of individuals' experiences. For example, dimensional rating scales might be better equipped to depict a general population sample and capture the fluctuation of strengths and weaknesses in self-regulation (Blume et al., 2020; Schulz-Zhecheva et al., 2017).

Conclusion

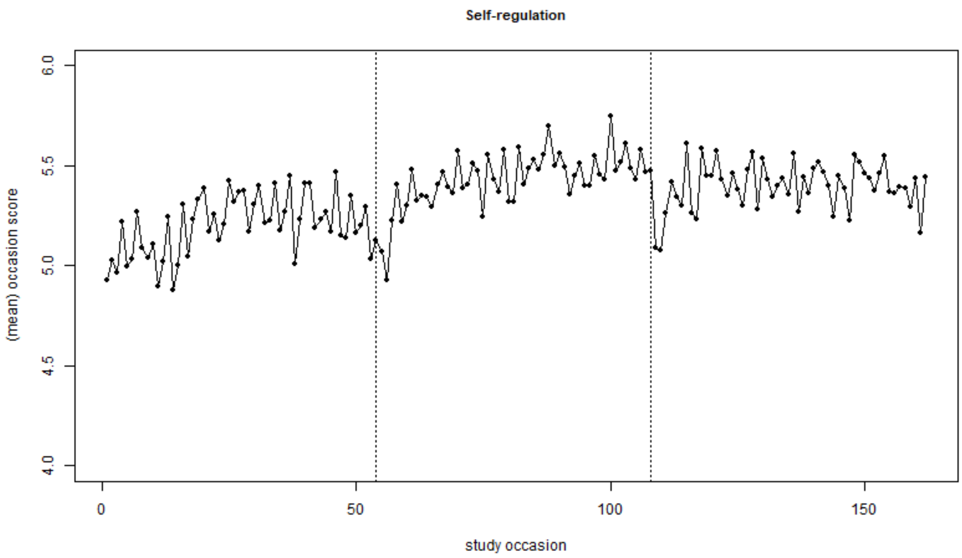
Children experience occasions of high and low self-regulation skills and working-memory performance in their daily lives. In the present study, these fluctuations of self-regulation ability could be equally divided over two and working-memory performance could be equally divided over three temporal frames: moment-to-moment (only working memory), occasion-to-occasion, and day-to-day fluctuations. That means that measured self-regulation ability or working-memory performance in a specific situation are influenced by the interplay of individual performance at the moment, occasion, and day. We found an association of self-regulation and working-memory performance on the between-person level for the difficult condition of the working memory task, but not for the easier condition, probably due to a ceiling effect. A significant association was found between self-regulation and within-person working-memory performance in the easier condition of the working memory scales. Future research should develop scales that are reliable and valid in ambulatory assessment and might better depict (within-person) associations between self-regulation and working memory. The study shows that we should further pursue approaches of measuring self-regulation and working memory repeatedly in intensive longitudinal studies and investigate the multitude of potentially influential factors at the moment, occasion, and day level.

Appendix 1

To test for the occurrence of an initial elevation bias, where participants answer more extreme in the beginning of an assessment period and then regress more towards the mean in later assessments (Shrout et al., 2018), we first inspected the data visually. Figure 11 shows the mean self-regulation of all children per study occasion. The dashed lines represent the beginning of a new study burst. It seems that the first three occasions of every burst indicate lower mean self-regulation than the subsequent occasions.

Figure 11

Visual Inspection of Initial Elevation Bias



We then tested whether including this possible effect changed the findings from the multilevel models. Dummy codes were generated for each of the first three days per burst. Only the inclusion of the first day in the model significantly enhanced the fit, day two and three were not associated to self-regulation. Children reported significantly lower self-regulation on the first day of each burst compared to the other days (β = - 0.18, $p < .001$). However, the associations between self-regulation and all other explanatory variables did not change due to the inclusion of a variable describing the first day into the model (Table 13).

Table 13

Multilevel Models with Day 1 included.

	WM Load 2			WM Load 3		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	5.217	5.093 – 5.340	<.001	5.221	5.099 – 5.344	<.001
Day 1	-0.181	-0.254 – -0.107	<.001	-0.183	-0.257 – -0.110	<.001
Occasion	0.004	0.001 – 0.006	.01	0.003	0.001 – 0.006	.01
Burst 2	0.207	0.112 – 0.301	<.001	0.206	0.112 – 0.300	<.001
Burst 3	0.258	0.156 – 0.360	<.001	0.253	0.151 – 0.355	<.001
WM between	0.147	-0.046 – 0.339	.13	0.164	-0.009 – 0.336	.06
WM within	0.032	0.006 – 0.059	.02	0.020	-0.004 – 0.044	.11
Medication	-0.413	-0.766 – -0.060	.02	-0.395	-0.748 – -0.043	.03
Occasion * Burst 2	-0.001	-0.004 – 0.002	.56	-0.001	-0.004 – 0.002	.53
Occasion * Burst 3	-0.002	-0.005 – 0.001	.24	-0.002	-0.005 – 0.001	.26
Random Effects & Autocorrelation	*Estimates*	*Residual*		*Estimates*	*Residual*	
Intercept	0.42			0.42		
Occasion	0.006			0.006		
WM within	0.070	*0.53*		0.060	*0.53*	
Autocorrelation	.27			.26		

Note. N = 70 subjects, 5100 observations

4.3. Manuscript 3: Dimensional Self- versus Observer Report of Inattention, Hyperactivity/Impulsivity, and Emotion Regulation in a General Adult Sample

Submitted as:

Buhr, L., Greiner, F., Schwarz, U., & Gawrilow, C. (2023). Dimensional Self- versus Observer Report of Inattention, Hyperactivity/Impulsivity, and Emotion Regulation in a General Adult Sample. *Submitted to: Journal of Attention Disorders*.

Abstract

Objective

This study aims at the dimensional assessment of symptoms of inattention, hyperactivity/impulsivity, and emotion regulation in the general population and the exploration of differences and commonalities between self- and observer ratings.

Method

142 adult dyads (i.e., target person and a significant other) completed a dimensional ADHD questionnaire (SWAN-DE), extended by an emotion regulation scale.

Results

Analyses show moderate correlation between self- and observer reports (r = .41 - .61). Self-reports are significantly more negative, with evidence for more extreme answer tendencies on strengths and weaknesses. Partial strict measurement invariance of the subscales inattention and hyperactivity/impulsivity was confirmed, but not for the emotion regulation subscale. Differences in informant reports are associated with age of the target person, age and gender of the observer, and symptom strength.

Conclusion

The dimensional SWAN-DE scales can be applied for self- and observer report in the measurement of symptoms of inattention, hyperactivity/impulsivity, and emotion regulation.

Keywords: ADHD, Emotion Regulation, Dimensionality, SWAN, Self-and-other Ratings

Introduction

People differ in their ability to focus on a task at hand as well as regulating their behavior and impulses. When these symptoms of inattention and/or hyperactivity/impulsivity are pronounced, an Attention Deficit-Hyperactivity Disorder (ADHD) can be diagnosed. However, inattention and hyperactivity/impulsivity occur also in the general population in different intensities. Therefore, scientific evidence has hinted for some time towards a dimensional definition of psychological disorders in general (Coghill & Sonuga-Barke, 2012) including ADHD (Frazier et al., 2007).

Dimensional Measurement of ADHD-Symptoms

In a dimensional view, people with a disorder "differ in degree, not in type" (Coghill & Sonuga-Barke, 2012, p.469) of their symptoms, with some people in the general population showing high strengths in regulating their attention and behavior and others with intense problems in self-regulation. To account for this dimensional view, the "Strengths and Weaknesses of ADHD and Normal Behavior" (SWAN; Swanson et al., 2012) scales have been developed for parents and teachers to report the symptoms of children and has been translated and evaluated in several languages (Lai et al., 2011; Lakes et al., 2012; Robaey et al., 2007; Schulz-Zhecheva et al., 2017). Recently, Blume et al. (2020) developed and tested a German self-report version of the SWAN scales for adults (SWAN-DE-SB).

In contrast to other ADHD questionnaires, the items of the SWAN are neutrally formulated instead of stating (negative) symptoms and can be judged on a 7-point Likert scale from *far below average* to *far above average* in comparison to other people. The scale has shown to generate a normal distribution of mean scores in the general population, with individuals with an ADHD diagnosis scoring far on the *far below average* side (Blume et al., 2020; Polderman et al., 2007). The questionnaire has also shown good reliability and validity (Brites et al., 2015).

Association of Emotion Regulation Deficits with Inattention and Hyperactivity/Impulsivity

Research has found increasing evidence that the symptoms of inattention and hyperactivity/impulsivity in adults often co-occur with deficits in emotion regulation (Retz et al., 2012).

The term emotion regulation describes the attempt to control, the occurrence, experience and expression of emotions (Christiansen et al., 2019). Emotion regulation can be administered at different stages: through selection of external factors (e.g., through the choice of environment and situation), by changing the emotion when it occurs (e.g., thinking of something else, breathing slowly), or by adapting the response (e.g., withholding an emotional outburst). A study of adults with and without an ADHD diagnosis has shown that emotion regulation moderates the relationship between inattention, hyperactivity/impulsivity symptoms and functional outcomes (Bodalski et al., 2019). Theory suggests three models to describe the connection between inattention, hyperactivity/impulsivity, and emotion regulation (Shaw et al., 2014). The first includes deficits in emotion regulation as a third core symptom in ADHD next to inattention and hyperactivity/impulsivity, the second distinguishes between two groups of individuals with ADHD, one with and the other without deficits in emotion regulation and the third sees emotion regulation as a factor that is distinct but significantly correlated to ADHD. In a factor analysis of data from a clinical sample of 213 adults, Hirsch and colleagues (2018) found evidence for emotion regulation as a core symptom of ADHD next to inattention and hyperactivity/impulsivity, but they caution that much more research is necessary for a final verdict on the topic.

Some scales to assess ADHD symptoms in adults have been integrating emotion regulation as subscales, for example the *Conners Adult ADHD Rating Scales* (CAARS; Christiansen et al., 2013) or the *Wender Utah Rating Scale* for retrospective diagnosis (WURS; Ward et al., 1993), although the symptoms are currently not integrated in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-V; American Psychiatric Association, 2013). However, taken the scientific evidence, we think it important to assess strengths and weaknesses in emotion regulation corresponding to inattention and hyperactivity/impulsivity when assessing ADHD symptoms to inform about abilities of an individual. Additionally, since the association of emotion regulation with inattention and hyperactivity/impulsivity is mainly assessed in clinical samples, the examination in a general population sample might enable new insights. For example, a dimensional association of emotion regulation with inattention and hyperactivity/impulsivity in the general population might give evidence for a systematic relationship of emotion regulation and ADHD symptoms.

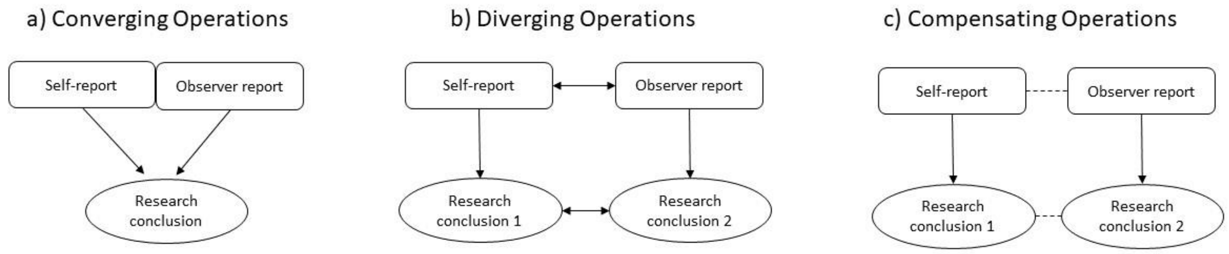
Higher ADHD symptoms are associated with negative life outcomes and strengths in attention and regulation of behavior and cognition with positive consequences (Barkley, 2002; Caye et al., 2016; Kuriyan et al., 2013; Loe & Feldman, 2007; Robson et al., 2020; Sciberras et al., 2009; Smithers et al., 2018). Consequently, we consider these strengths and weaknesses important to examine further in the general population. Understanding the distribution of ADHD symptoms in the general population might, for example, help to tailor interventions for individuals with higher symptoms and promote strengths in attention and behavior regulation. However, the reliability and validity of measurement of ADHD symptoms in the general population has to be investigated further (Blume et al., 2020). To enhance validity of the measures, it is general practice in clinical and research assessment to ask several reporters about psychological constructs, like ADHD symptoms (De Los Reyes et al., 2013; Hunsley & Mash, 2007). In adults, this is typically the patient themselves and a significant other, for example a partner, close friend, or family member. However, it has been shown that these self- and observer reports significantly differ from each other. It is therefore difficult to interpret empirical findings, without knowing how to handle differences between reporters. Despite the challenges that these reporter discrepancies pose in research, there is yet no consensus on how to handle inconsistent informant reports.

Self- versus Observer Report

To decide how to handle and interpret differences between reporters, we need to understand the reason for these divergences. For this reason, De Los Reyes and colleagues (2013) have developed the Operations Triad Model to systematically describe and explain differences in observers (Figure 12). According to them, differences in observer reports can be handled by Converging Operations, Diverging Operations, or Compensating Operations. Integrating multiple informants into one research conclusion and interpreting all differences between their ratings as measurement error, is called Converging Operations. The difference between observers can, however, also include meaningful information for diagnosis or research aims. In this case, when despite the divergence there are systematic relations between the observers, we can use Diverging Operations. Finally, if the findings between informants are entirely inconsistent and no systematic relations can be found between them, Compensating Operations apply.

Figure 12

Operation Triad Model (adapted from De Los Reyes et al., 2013)



The Operations Triad Model has been advanced by Alexander and colleagues (2017), who added the question of context and insight to the decision whether and what kind of multiple informants should be used. Applying this theory to ADHD, symptoms of inattention and hyperactivity/impulsivity might be present in different contexts and more or less adaptive according to context (Schmid et al., 2016). This would imply that different observers might have different information about an individual. For example, a work colleague might see more inattentive and/or hyperactive/impulsive symptoms than a friend with whom the individual meets for sportive activities or partying. Additionally, in some cases individuals with higher ADHD symptoms might lack the insight of the intensity of their symptoms (Smith et al., 2000). On the other hand, observers might sometimes not see what happens inside the individual and therefore not detect instances of high inattention. Hence, the individual itself might be better equipped to report about internal symptoms like inattention, while an observer might have more ability to report about external symptoms like hyperactivity/impulsivity. Evidence for this notion has already been found in previous research, where self- and observer reports were more similar on external than on internal symptoms (Belendiuk et al., 2007).

Literature review about self- and observer report in ADHD questionnaires

Studies on informant agreement in ADHD symptoms differ on whether past or current symptoms are examined whether the reporter is a clinician or a personal contact of the participant, on the inspected sample (clinical, referred for diagnosis, general population) and the questionnaire applied. Most studies until now have examined agreement between teacher- and parent-report on ADHD symptoms in children (e.g. Narad et al., 2015). Others have compared self-report questionnaires of adults with clinician diagnoses after a structured interview (Abrams et al., 2018). Here, we focus on research of current ADHD symptoms in adults, assessed with a self-report and an observer report by a significant other person, like a partner, a family member or a close friend.

In a study with 281 college students with educational difficulties, Zucker and colleagues (2002) found moderate correlations between self- and observer ratings on the subscales inattention and hyperactivity/impulsivity (r = .55 - .57) on the ADHD Behavior checklist for adults. Significant others, which were partners, friends, or family members, reported significantly more inattentive symptoms than participants did in their self-report. Van Voorhees and colleagues (2011) assessed 349 adults referred for ADHD diagnosis with help of the Conners ADHD rating scale (CAARS) and found weak to moderate correlations between self- and observer report (r = .24 - .46 on the item level, r = .33 - .39 on the scale level). Self-reports were in general more negative than observer reports and there was no effect of the type of observer (partner, family member, friend). In another sample of adults referred for ADHD assessment, Kooij and colleagues (2008) found a wide range of weak to strong correlations between observers, depending on the questionnaire used (r = .3 - .7). Here as well, the participants rated themselves as having more severe ADHD symptoms than did their significant others.

A very different sample was used in the study by Belendiuk et al. (2007), who assessed 69 mothers of children with an ADHD diagnosis. They found a weak correlation of r = .29 for hyperactive/impulsive symptoms and a moderate correlation of r = .54 for inattentive symptoms. In contrast to the study by Zucker et al. (2002), participants reported significantly more inattentive symptoms than their significant others. The authors explain the high correspondence between self- and observer reports with the specific sample, which has a) a statistically higher probability to experience ADHD symptoms themselves due to heritability of the disorder and b) an enhanced knowledge about the disorder because of the diagnosis of their children.

Few studies tried to explain the variance in differences between self- and observer ratings by demographic or relationship variables (Barkley et al., 2011). In research with children, we find evidence that girls might have higher concordance with their parents, supposedly because parents seem to talk more with their daughters about their behavior (Belendiuk et al., 2007). Adults did not depict any gender differences concerning the reporter difference in a community sample, but in a clinical sample, women showed greater differences in their self-report to an observer report than men (Barkley et al., 2011). The gender of the rater has, to our knowledge, not been assessed yet as a possible influencing factor for the divergence of self- and observer ratings. Older adults have higher divergence in their self- to observer ratings when reporting about ADHD related impairments (Barkley et al., 2011). We could not find any studies investigating whether the age of the observer might have an influence on the congruence of their ratings to self-reports. No effect could be found according to the type of observer, whether the observer was a family member, a partner or a friend (Barkley et al., 2011; Van Voorhees et al., 2011). According to the Operations Triad Model, context and insight are important to reliably report about behavior (Alexander et al., 2017). For example, when people live together, they might see each other in many different contexts and thus show higher agreement in their ratings but to our knowledge, no study had yet assessed whether target person and observer were living in one household. Finally, insight might be more difficult for a person when they experience higher symptoms of ADHD (Smith et al., 2000), and a diagnosis of ADHD seems associated with lower agreement between self- and observer reports (Zucker et al., 2002). These demographical (gender & age of target person and observer), relationship (partner, friend, family member, living together) and disorder variables (strength of the ADHD symptoms), therefore, need to be inspected more carefully on their association with differences in observer ratings.

Based on the above-described research, we would conclude that self- and observer ratings on ADHD scales for adults correlate moderately. More specifically, agreement appears to be higher on the externalizing symptoms like hyperactivity/impulsivity than on internalizing symptoms like inattention. In general, individuals seem to report more ADHD symptoms than observers do, thus their judgment of their own behavior seems to be more negative than it is seen from outside. Due to a lack of previous research, explorative investigations seem advisable concerning whether demographic variables (gender & age of the target person and the observer), relationship variables (partner, friend, family member, living in one household), or symptom strength are associated with the difference between self- and observer reports. Furthermore, only few studies on the associations of self- and observer reports focus on ADHD symptoms in a non-clinical sample (Barkley et al., 2011). This, however, might be valuable since the general population would be expected to comprise a wider variance of attention and behavior abilities than a clinical sample. However, typically used categorical ADHD questionnaires are not able to adequately depict this variance, since most people would fall into the non-ADHD category without differentiation of their behavior. A possible solution to this challenge is the application of dimensional scales of ADHD symptoms.

Using Dimensional Scales to compare Self- and Observer Reports

When used for comparing self- and observer report, the dimensional view might result in new findings and innovative research questions. For categorical ADHD scales, interrater agreement is usually higher in general population samples than in clinical samples, since most participants would score on the *zero* part of the scale. In clinical samples, interrater agreement is lower, supposedly because of the higher variance in symptoms that individuals depict (Kooij et al., 2008). When a general population sample is assessed with help of the SWAN, variance should typically be high as well, because it assesses a much wider range between strengths and weaknesses (Swanson et al., 2012).

Another advantage of the dimensional measurement of ADHD symptoms is the possibility to investigate differences on both extremes of the scale, not only analyzing whether self- or observer reports depict more weaknesses but also, whether they depict more strengths. Taken the finding, that self-reports state more ADHD symptoms than observer reports (Kooij et al., 2008; Van Voorhees et al., 2011), we can investigate with a dimensional scale, whether they would report more or less strengths concerning attention and impulse-control.

The child-versions of the SWAN have previously been examined on observer differences between parents and teachers. Correlations between parent and teacher ratings on the SWAN scales were moderate to strong (r = .55 - .66). Low agreements could on the one hand be explained by target/observer specific characteristics. For example, when a child had a lower language ability, teachers reported more inattentive symptoms and agreement with parent report was therefore lower (Gooch et al., 2017). The differences between parent and teacher ratings on the SWAN might on the other hand be due to different underlying constructs, as has been shown by a measurement invariance analysis (Jungersen & Lonigan, 2021). This means that parents and teachers seem to interpret the items on the SWAN scales when assessing a child differently.

Applying these research questions to adults, we would like to assess what characteristics of the target person and the significant other might impact the discrepancies between the reporters. Furthermore, it seems necessary to examine whether the self- and observer report scales show an identical factor structure and therefore measure the same constructs.

Current study

In the current study, we assessed a general population sample of 142 adults on their self-reported and observer reported regulation of attention and behavior, operationalized as ADHD symptoms. The prevalent research question concerned the correspondence of self- and observer reports on a dimensional ADHD symptom questionnaire. Based on the above stated literature, we formulated the following hypotheses:

H1: The SWAN-DE scales can be used in a version for self- as well as for observer reports to measure ADHD symptoms in a general population sample.

H1a: Emotion regulation can be assessed dimensionally in a third subscale next to inattention and hyperactivity/impulsivity in the SWAN-DE scales.

H1b: The self- and observer versions of the SWAN-DE both show a normality distribution and a high reliability.

H1c: The self- and the observer report version of the SWAN-DE measure the same underlying constructs.

H2: Self- and observer report on the SWAN scales for adults correlate moderately.

H2a: Correlation between observers is higher for the subscale of hyperactivity/impulsivity than for the subscale of inattention.

H2b: Target persons evaluate themselves as more negative than their significant others do.

H2c: Target persons tend to more extreme self-reports as well in their strengths as in their weaknesses than the observer reports of the significant others do.

H3: Differences between self- and observer reports are associated with specific demographics of the target person or the significant other (gender & age), their respective relationship (partner, friend, family member, living in one household) or the strength of ADHD symptoms.

H3a: Demographic, relationship and disorder variables influence the magnitude of difference between self- and observer reports.

H3b: Demographic, relationship and disorder variables influence the direction of difference between self- and observer reports.

Method

Data was collected with an online questionnaire on SoSci Survey (Leiner, 2019). The study was approved by the psychological ethics committee of the University of Tuebingen (Az Gawrilow_2020_0717_196).

Participants

Participants were recruited via the general Mailinglist of the University and personal contacts. People who agreed to participate filled out the self-report questionnaire and then received a personalized link to forward to a significant other (partner, family member, friend) of their own choice. Self-reports were collected for 284 individuals, full information was available for 142 dyads. Individuals without observer report had a higher probability of having an ADHD diagnosis, but the groups of people with and without observer report did not differ significantly in terms of age, gender, education and mean self-report ADHD symptoms on the SWAN-DE total or the three subscales inattention, hyperactivity/impulsivity, and emotion regulation.

Only complete dyads of a target person and their significant other were included in the analyses. The target individuals who provided self-report of their symptoms were mainly female (101 female, 39 male, 2 divers) and mean age was 24.53 years (*SD* = 7.74 years, range 18 – 61 years). Education level of the target individuals was very high, 101 indicated to have a secondary school degree (German: *Abitur*), 36 accomplished a University degree (German: *Bachelor* or *Master*) and 4 had successfully finished a vocational training (German: *Berufsausbildung*). Significant others, who provided observer report about the target person, had a higher proportion of men (85 female, 54 male, 1 divers) and were slightly older (mean age 28.9 years, *SD* = 11.57 years, range 18 – 61 years) than the target group. The relationship of the dyads was indicated by the significant other individual. Most participants asked their partner to provide observer-report of their symptoms (*N* = 53), others a friend (*N* = 44) and some a family member (significant other person indicated as parent of target person: *N* = 19, siblings: *N* = 18, significant other is the child of a target person: *N* = 2). In addition, 28 dyads were living in the same household. Only target individuals were asked for current and past ADHD diagnoses as well as medication. Two participants indicated a current and two a childhood diagnosis of ADHD. Surprisingly, twelve participants disclosed that despite a lack of diagnosis they suspected having ADHD themselves. None of the participants indicated taking medication.

Procedure

Measures

SWAN-DE

The Strengths and Weaknesses of ADHD and Normal Behavior (SWAN; Swanson et al., 2012) scales have been developed to account for a normal distribution of ADHD symptoms in the general populations. The 18 neutrally formulated items depict the symptom list of the Diagnostic and Statistical Manual of Mental Disorders (DSM - 5; American Psychiatric Association, 2013). The German version for self-report of adults has been developed and tested previously (SWAN-DE-SB; Blume et al., 2020). For the observer report (SWAN-DE-FB), all items were reformulated from the first person singular (*I am sitting still*) into the third person singular (*S/he is sitting still*). For better comprehension, the significant other was asked the name (or a nickname) of the target person in the beginning of the study and this name was automatically included within all items (e.g., *Max is sitting still*) and immediately deleted after completion of the study due to data security reasons. An overview of the scale can be seen in Table 20 (Appendix).

Responses were coded from 0 (*far below average*) to 6 (*far above average*), with lower values indicating higher ADHD symptoms and higher values indicating better self-regulation of attention and behavior. As outcome variable, the mean score of the total score can be used. Additionally, the total ADHD scale can be separated into two subscales inattention (items 1-9) and hyperactivity/impulsivity (items 10 -18).

Emotion Regulation

To account for scientific evidence of emotion regulation being associated with inattention and hyperactivity/impulsivity, we included a self-generated subscale of emotion regulation into the SWAN-DE scales. Nine neutrally formulated items were created, by reformulating the deficit-oriented items in literature (e.g., Corbisiero et al., 2017) and categorical scales for diagnosis of ADHD (e.g., ADD; Brown, 2006; CAARS; Christiansen et al., 2013; WURS; Ward et al., 1993) into a neutral wording (see Appendix, Table 20). Comprehension and correlation of the new items with the original deficit-oriented items was assessed in a pilot study including 46 adult participants. One item was excluded due to low comprehensiveness. All items showed low to moderate negative correlation with the deficit-oriented items (r = -.33 - -.63).

Analyses

All analyses were implemented in R (version 2022.02.0). Self- as well as observer reports were inspected separately concerning their normality distribution visually via histograms and statistically through a Kolmogorov-Smirnov-Test, skew and kurtosis. Reliability was tested with help of Cronbachs Alpha and item-whole correlations for each item with the total ADHD scale and the subscales were analyzed. Pearson Correlations were calculated to examine the association between self- and observer report. To test for statistical differences, t-tests were conducted on the total and both subscales. Linear regression of the z-standardized values from the self- on the observer report were graphically depicted to investigate whether the self-ratings were more extreme than the observer ratings. A confirmatory factor analysis (CFA) was conducted on the self-report scale with help of the R package *lavaan*. Fit indices by Hu & Bentler (1999) were used as a rule of thumb to evaluate model fit, but not as strict cut-off values, since these values are very sensitive to measurement conditions. After a satisfactory model fit was accomplished, measurement invariance between the self- and the observer report scale was tested in a stepwise approach on the configural, metric, scalar and residual invariance (Putnick & Bornstein, 2016). Each model was compared with the previous one using a chi-square difference test. When a significant difference in model fit was suggested, measurement invariance was not confirmed. The model then was adapted by allowing single differences between the groups, until no significant difference in model fit was found to the previous level. Difference scores were calculated by subtracting the means of the observer-ratings from the means of the self-ratings. The *magnitude* of the difference, is represented by the absolute values (distance from zero, without negative or positive sign). The *direction* of the difference, on the other hand, is indicated by the real number of the difference (with negative of positive sign). For the direction of difference, high difference values indicate more positive self-ratings than observer ratings and low difference values indicate more negative self-ratings than observer ratings. To explain the variance in magnitude and direction of difference, multiple regression analyses were performed for the total scale, the inattention, hyperactivity/impulsivity, and emotion regulation subscales with demographic variables (age and gender of target person and observer), relationship (partner, friends, family, living in one household) and ADHD symptom strength as the explanatory variables and the difference scores of the total and the subscales as the dependent variable. All analyses were tested against $\alpha = 0.05$.

Results

Normality Distribution and Reliability of the SWAN-DE-SB and SWAN-DE-FB

Our first research question concerned the normal distribution and reliability of the self- and observer report scales (H1b). The total scales of the self-report and the observer report questionnaires, as well as their respective subscales inattention, hyperactivity/impulsivity, and emotion regulation all showed normal distributions which were slightly shifted towards the above average/strengths part of the scale (Figure 13). Cronbachs alpha was α = .64/.77 for the total self-report and the observer report scale respectively and α = .75/.78 for the inattention, α = .8/.84 for the hyperactivity/impulsivity, and .82/.89 for the emotion regulation subscales.

Figure 13

Normality distributions of the Total ADHD scale and the subscales inattention, hyperactivity/impulsivity, and emotion regulation of the self-report version and the observer report of the SWAN-DE

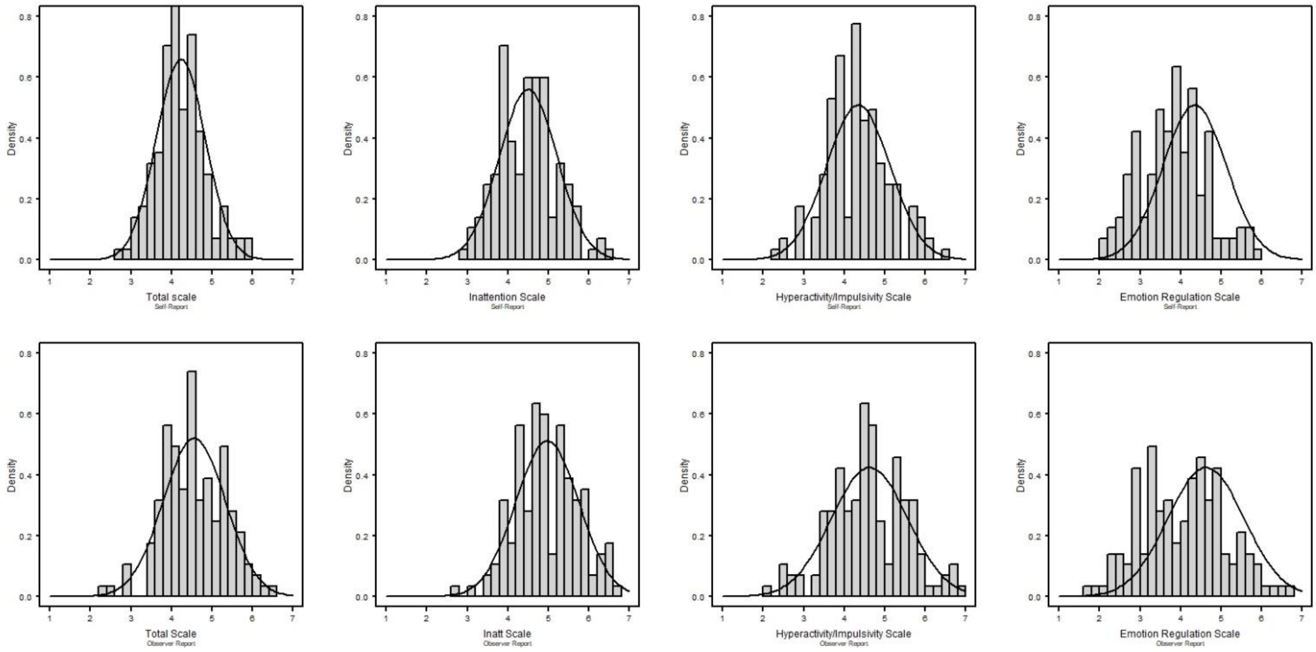


Table 14

Distribution Analysis of the SWAN-DE self-report and observer Report.

	SWAN-DE-SB				SWAN-DE-FB			
	D	*p*	Skew	Kurtosis	*D*	*p*	Skew	Kurtosis
Total scale	1	<.001	.25	.25	0.99	<.001	-.09	.04
Subscale Inattention	1	<.001	.21	-.19	1	<.001	.01	-.33
Subscale Hyperactivity/Impulsivity	0.99	<.001	.18	.26	0.99	<.001	-.07	.06
Subscale Emotion Regulation	0.98	<.001	.19	-.25	0.97	<.001	.10	-.52

Note. SWAN-DE-SB = self-report of German SWAN scale, SWAN-DE-FB = observer report of German SWAN scale, D = Kolmogorov-Smirnov-Test.

The items all correlated moderately with the total scale and their respective subscales (Table 15). The subscales showed very strong correlations with the total scales (self-report: inattention r = .82 / hyperactivity r = .76 / emotion regulation r = .77; observer report inattention r = .80 / hyperactivity/impulsivity r = .84 / emotion regulation r = .86).

Table 15

Part-whole corrected item-whole correlation.

	SWAN-DE_SB		SWAN-DE-FB	
	r_{item}	r_{mean}	r_{item}	r_{mean}
Total scale	0.17 - 0.44	0.32	0.32 - 0.55	0.45
Subscale Inattention	0.21 - 0.59	0.43	0.18 - 0.58	0.46
Subscale Hyperactivity/Impulsivity	0.32 - 0.57	0.49	0.32 - 0.67	0.56
Subscale Emotion Regulation	0.22 - 0.72	0.52	0.34 - 0.83	0.63

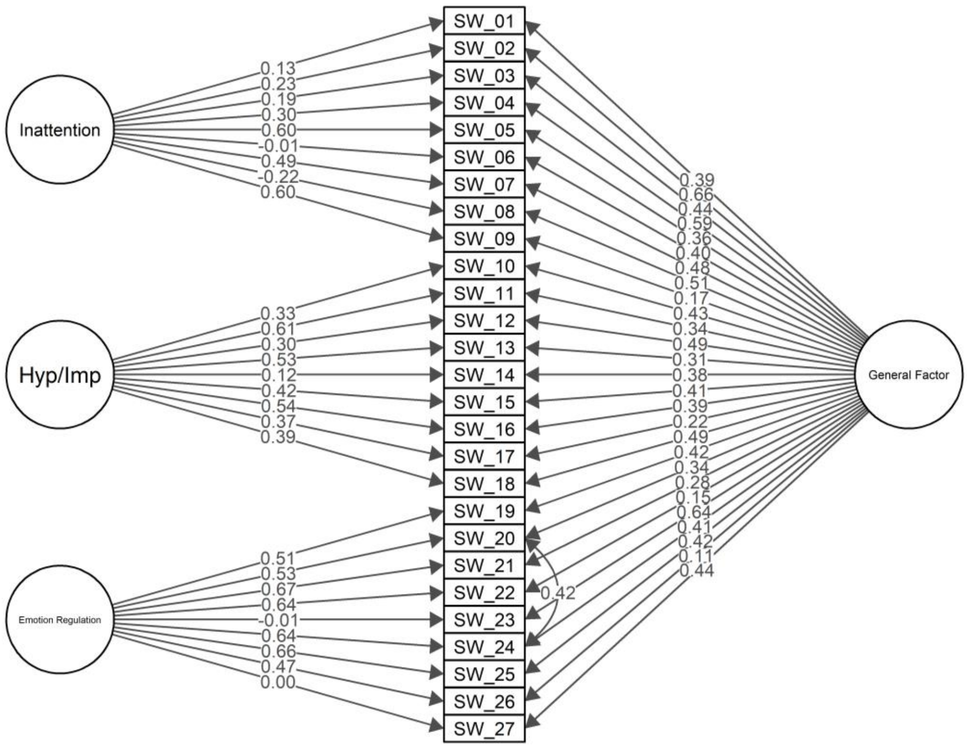
Note. SWAN-DE-SB = self-report of German SWAN scale, SWAN-DE-FB = observer report of German SWAN scale, r_{item} = range of the item whole correlations, r_{mean} = mean of the item whole correlations

Measurement Invariance

To examine, whether the SWAN-DE-SB and the SWAN-DE-FB measure the same underlying constructs, a measurement invariance analysis was conducted (H1c). As a first step to test measurement invariance between the self- and the observer report scales of the SWAN, a confirmatory factor analysis (CFA) was conducted on the self-report scale. Three models were built stepwise and model fit was tested against each other: a non-hierarchical model with the first nine items loading on the latent factor inattention, nine loading on the latent factor of hyperactivity/impulsivity, and nine loading on the factor emotion regulation and a bifactor model with a general ADHD factor and the three subscales as latent factors. The bifactor model showed significantly better model fit. Inspection of modification indices suggested to allow correlated measurement errors between items 20 (*Stay calm in difficult situations*) and 24 (*Stay calm on the inside in stressful situations*). Due to the similarity of the items, correlated measurement errors were allowed. The modified bifactor model (Figure 14) showed adequate model fit (CFI = .79, TLI = .75, RMSEA = .08, SRMR = .08). Measurement invariance was not given on the metric (or weak) level, indicating that factor loadings were significantly different for self- and observer reports.

Figure 14

Structual Equation Model of the SWAN-DE



To test whether this low model fit and the lack of measurement invariance was influenced by the newly developed scale of emotion regulation, the same analyses were conducted only with the items of the subscales inattention and hyperactivity/impulsivity included. Again, the bifactor model with a general ADHD factor and two subscales as latent factors fit the data better than a model with only the two subscales as latent factors (Figure 15). Model fit was significantly higher when emotion regulation was not included (CFI = .87, TLI = .83, RMSEA = .07, SRMR = .07). An overview of the results of the measurement invariance analyses can be seen in Table 16. The configural invariance model showed adequate model fit, providing evidence that the individual items load on the same factors as well in the self-report as in the observer report data. Similarly, metric (or weak) measurement invariance was confirmed, indicating equivalent factor loadings between the observer groups. The significant difference between the model for metric and the model for scalar (or strong) invariance suggest a difference in intercepts of one or more items between self-ratings and observer ratings. Three constraints were released stepwise according to suggestions by modification indices until the fit indices were not significantly different from the unconstrained model. Compared to this partial scalar invariance model, the strict invariance model showed no significant change in model fit, suggesting similar residual variances in the two groups.

Figure 15

Structural Equation Model of the SWAN-DE-SB without the Emotion Regulation Subscale

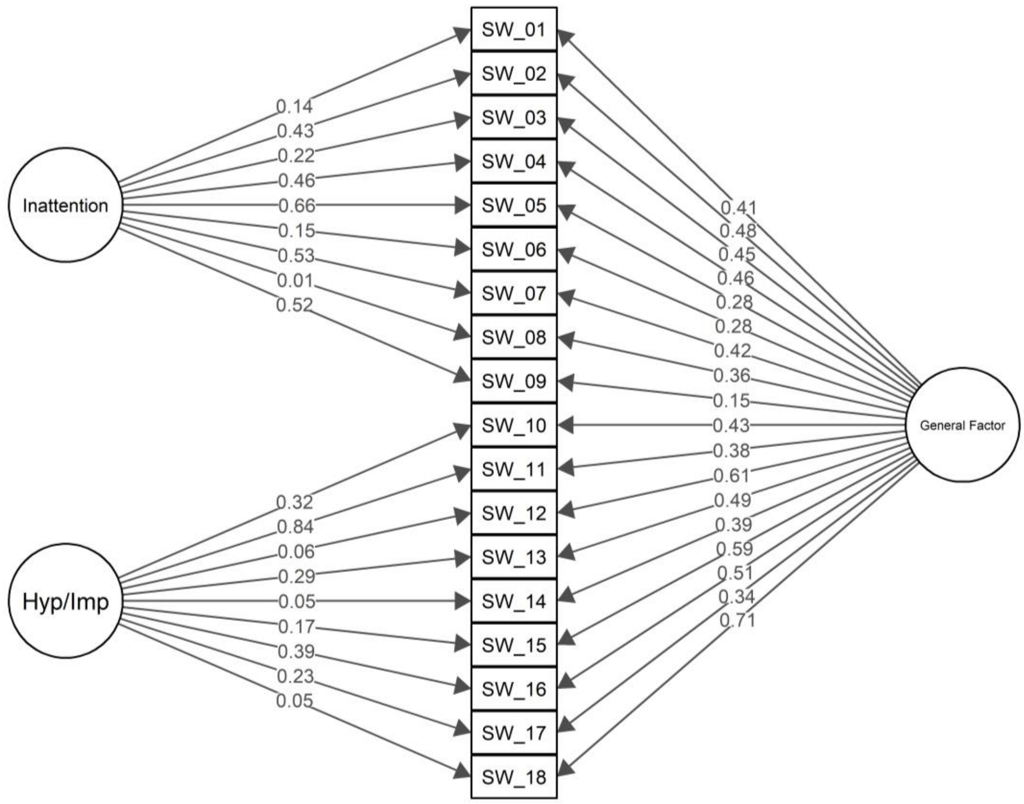


Table 16

Measurement Invariance of the SWAN-DE-SB versus SWAN-DE-FB.

	χ^2	Model *df*	*p*	CFI	TLI	RMSEA	SRMR	χ^2 diff (df)	*p* diff
Configural Invariance	556.115	240	0	.839	.795	.080	.093	NA	NA
Metric (weak) Invariance	600.597	273	0	.834	.813	.077	.099	44.48 (33)	.087
Scalar (strong) Invariance	641.206	288	0	.821	.809	.078	.104	40.61 (15)	.000
Partial scalar Invariance	618.547	285	0	.831	.818	.076	.101	17.95 (12)	.117
Strict Invariance	640.217	303	0	.829	.827	.074	.100	21.67 (18)	.247

Note. Subscale emotion regulation excluded from the model, CFI = comparative fit index, TLI = Tucker-Lewis fit index, RMSEA = root mean square error of approximation, SRMR = standardized root mean square residual

Association of Self- and Observer Report

Our second research question concerned the association of self- and observer reports. Mean scores as well as differences of self- and observer report can be seen in Table 17. Correlations between self- and observer reports were moderate (r = .41 - .61) (H2). Contrary to our expectations, the correlation of hyperactive/impulsive symptoms was not higher than of inattentive symptoms (H2a). Difference between self- and observer report was significant on the total- as well as on all three subscales, self-reports were significantly more negative than observer reports (H2b).

Table 17

Descriptives and Associations of the SWAN-DE self-report and observer report.

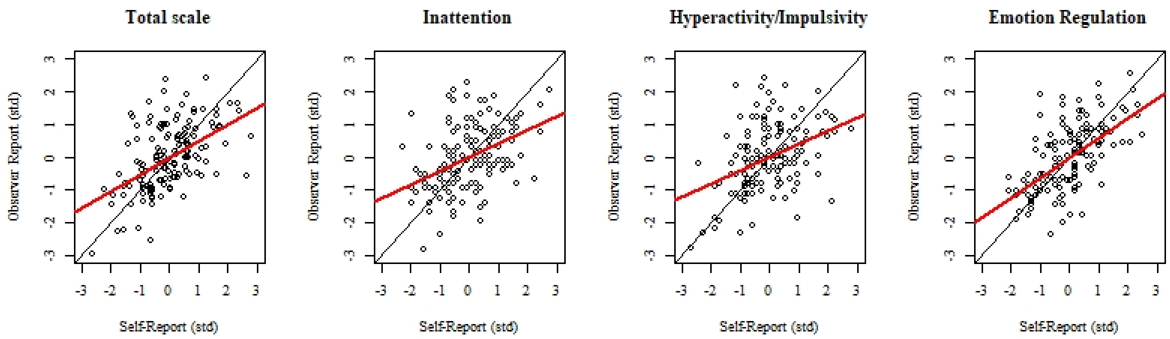
	SWAN-DE-SB	SWAN-DE-FB	Correlation SB & FB	t-test SB vs FB
Total scale	M = 4.23	M = 4.56	r = .51	$t(141)$ = -5.74, p < .001
Subscale Inattention	M = 4.50	M = 4.97	r = .41	$t(141)$ = -6.92, p < .001
Subscale Hyperactivity/Impulsivity	M = 4.35	M = 4.62	r = .41	$t(141)$ = -3.49, p < .001
Subscale Emotion Regulation	M = 3.84	M = 4.09	r = .61	$t(141)$ = -3.53, p < .001

Note. SWAN-DE-SB = self-report of German SWAN scale, SWAN-DE-FB = observer report of German SWAN scale

Regression of the self-report values on the observer report values indicates a tendency for more extreme answers on the positive as well as on the negative side for self-report compared to observer report (Figure 16) (H2c). For equal answer tendencies, slopes of the regression would be expected to have a value close to 1. In contrast to that, the slope for the total ADHD scale is indicated as $b_1 = 0.51$, for the subscales inattention and hyperactivity/impulsivity both $b_1 = 0.41$ and for the subscale emotion regulation $b_1 = 0.61$.

Figure 16

Simple Linear Regression of Self-Report on Observer Report



Note. black line = regression line for complete agreement, red line = regression of self-report on observer report

Divergence Analysis

Magnitude of difference

To test for associations of the absolute value of the difference between self- and observer reports with demographic, relationship, and disorder variables, four regression models were generated, one with total difference on the total scale, and one with total difference on each subscale (inattention, hyperactivity/impulsivity, and emotion regulation) as dependent variables (H3a). All explanatory variables were included at the same time. Results can be seen in Table 18. Magnitude of difference on the total scale was significantly associated with observer reported symptom strength ($\beta = 0.28$, $p < .001$). Only observer reported strength of inattention symptoms was significantly associated with the magnitude of difference on the inattention scale (β = 0.27, p < .001). Age of the target person (β = 0.03, p < .001) as well as observer reported strength of hyperactivity/impulsivity symptoms (β = 0.26, p < .001) were significantly associated with the subscale hyperactivity/impulsivity. Age of the significant other (β = -0.01, p = .02) as well as observer reported symptom strength (β = 0.14, p = .01) were significantly related with magnitude of difference on the subscale emotion regulation. The models explain between 9.7 and 22.8% of variance in the absolute value of the difference between self- and observer reports (total ADHD scale *adjusted* R^2 = 0.228, subscale inattention *adjusted* R^2 = 0.124, subscale hyperactivity/impulsivity *adjusted* R^2 = 0.217, subscale emotion regulation *adjusted* R^2 = 0.097).

Table 18

Regression on Magnitude of Difference between Self- and Observer Ratings.

	Total scale			Inattention			Hyperactivity/Impulsivity			Emotion Regulation		
Predictors	*Est*	*95% CI*	*p*	*Est*	*95% CI*	*p*	*Est*	*95% CI*	*p*	*Est*	*95% CI*	*p*
(Intercept)	-1.09	-1.74 – -0.45	**.001**	-0.90	-1.70 – -0.11	**.026**	-1.01	-1.78 – -0.25	**.010**	-0.06	-0.78 – 0.65	.864
Gender Self	0.05	-0.13 – 0.24	.572	-0.03	-0.25 – 0.19	.761	0.02	-0.22 – 0.26	.864	0.00	-0.23 – 0.24	.988
Gender Observer	0.05	-0.12 – 0.22	.580	0.05	-0.16 – 0.26	.623	-0.02	-0.25 – 0.21	.877	0.09	-0.12 – 0.30	.412
Age Self	0.01	-0.00 – 0.02	.057	0.01	-0.01 – 0.02	.445	0.03	0.01 – 0.04	**<.001**	0.01	-0.01 – 0.02	.323
Age Observer	-0.00	-0.01 – 0.00	.223	-0.00	-0.01 – 0.01	.878	-0.01	-0.02 – 0.00	.063	-0.01	-0.02 – -0.00	**.020**
Partner	0.14	-0.21 – 0.48	.443	0.09	-0.33 – 0.51	.672	0.05	-0.41 – 0.51	.829	0.35	-0.08 – 0.78	.110
Friend	0.12	-0.24 – 0.49	.498	0.11	-0.33 – 0.55	.617	0.12	-0.36 – 0.60	.627	0.11	-0.33 – 0.56	.617
Family Member	0.32	-0.06 – 0.71	.101	0.26	-0.20 – 0.73	.263	0.26	-0.25 – 0.77	.310	0.42	-0.06 – 0.89	.084
Living together	0.13	-0.08 – 0.35	.220	0.03	-0.23 – 0.29	.823	0.09	-0.19 – 0.37	.531	0.04	-0.23 – 0.30	.789
Observer Report	0.28	0.17 – 0.38	**<.001**	0.27	0.15 – 0.39	**<.001**	0.26	0.15 – 0.37	**<.001**	0.14	0.03 – 0.24	**.009**
Observations	132			132			132			132		
R^2 / R^2 adjusted	0.281 / 0.228			0.184 / 0.124			0.271 / 0.217			0.159 / 0.097		

Direction of difference

The direction of the difference between self- and observer ratings of the scales was examined equivalently in four different models, one for the total ADHD scale and one for each of the subscales inattention, hyperactivity/impulsivity, and emotion regulation (Table 19, H3b). For the total ADHD scale, gender of the significant other (β = -0.20, p = 0.04) and observer rated symptom strength (β = - 0.55, p < .001) were significantly negatively associated with the difference of self- and observer report. Observer rated symptom strength was the only explanatory variable which was significantly associated with the differences between self- and observer ratings for the subscale for inattention (β = -0.58, p < .001). The difference on the hyperactivity/impulsivity subscale was significantly associated with both, age of the target person (β = -0.02, p = 0.39) and observer reported symptom strength of hyperactivity/impulsivity (β = -0.61, p < .001). Gender of the significant other (β = -0.33, p = .01) as well as observer reported symptom strength (β = -0.48, p < .001) were significantly associated with rater differences on the emotion regulation scale. The models explain between 36.4 and 47.2% of the difference in self- and observer reports (total ADHD scale *adjusted* R^2 = 0.472, subscale inattention *adjusted* R^2 = 0.364, subscale hyperactivity/impulsivity *adjusted* R^2 = 0.461, subscale emotion regulation *adjusted* R^2 = 0.432).

Table 19

Regression on Direction of Difference between Self- and Observer Ratings.

	Total scale			Inattention			Hyperactivity/Impulsivity			Emotion Regulation		
Predictors	*Est*	*95% CI*	*p*	*Est*	*95% CI*	*p*	*Est*	*95% CI*	*p*	*Est*	*95% CI*	*p*
(Intercept)	2.57	1.84 – 3.30	**<.001**	2.77	1.83 – 3.70	**<.001**	3.19	2.28 – 4.09	**<.001**	1.87	1.03 – 2.71	**<.001**
Gender Self	-0.03	-0.24 – 0.18	.776	0.09	-0.17 – 0.35	.509	-0.06	-0.35 – 0.22	.666	-0.07	-0.35 – 0.20	.597
Gender Observer	-0.20	-0.40 – -0.01	**.042**	-0.07	-0.32 – 0.17	.559	-0.22	-0.49 – 0.04	.101	-0.33	-0.58 – -0.08	**.011**
Age Self	-0.01	-0.02 – 0.01	.243	-0.01	-0.02 – 0.01	.490	-0.02	-0.04 – -0.00	**.039**	0.00	-0.01 – 0.02	.857
Age Observer	0.01	-0.00 – 0.02	.181	0.00	-0.01 – 0.01	.745	0.01	-0.00 – 0.02	.107	0.01	-0.00 – 0.02	.250
Partner	-0.09	-0.49 – 0.30	.648	-0.06	-0.56 – 0.43	.803	-0.18	-0.72 – 0.37	.515	-0.00	-0.50 – 0.50	.995
Friend	-0.20	-0.61 – 0.21	.327	-0.38	-0.89 – 0.14	.150	-0.12	-0.69 – 0.44	.671	-0.12	-0.64 – 0.41	.662
Family member	-0.38	-0.81 – 0.06	.091	-0.43	-0.97 – 0.12	.126	-0.53	-1.13 – 0.08	.086	-0.17	-0.72 – 0.39	.552
Living together	-0.16	-0.40 – 0.09	.207	-0.08	-0.38 – 0.23	.628	-0.24	-0.58 – 0.09	.152	-0.17	-0.48 – 0.14	.291
Observer Report	-0.55	-0.67 – -0.43	**<.001**	-0.58	-0.72 – -0.43	**<.001**	-0.61	-0.74 – -0.48	**<.001**	-0.48	-0.60 – -0.36	**<.001**
Observations	132			132			132			132		
R^2 / R^2 adjusted	0.508 / 0.472			0.408 / 0.364			0.498 / 0.461			0.471 / 0.432		

Discussion

People differ in their strengths and weaknesses in attention and behavior regulation, which can be measured in dimensional scales of ADHD symptoms. However, the magnitude of symptoms is often rated differently by self- than by observer reports. In this article we examined the self-report version of the German Strengths and Weaknesses of ADHD and Normal Behavior Scales (SWAN-DE-SB) and the same items reformulated into third-person singular to create the observer version of the scales (SWAN-DE-FB) concerning the internal structure and association of self- and observer reports of ADHD symptoms in adults. Additionally, we newly developed a neutrally formulated subscale of emotion regulation to add to the SWAN-DE scales. Our main research questions concerned the usability of the scales and the concordance of the internal structure of the questionnaires (H1), the agreement of self- and observer report (H2), and the explanation of these differences in agreements (H3).

In our first hypothesis, we expected that the SWAN-DE scales could be used in a version for self- as well as for observer reports to measure ADHD symptoms in a general population sample (H1). We hypothesized that emotion regulation could be assessed dimensionally in a third subscale next to inattention and hyperactivity/impulsivity in the SWAN-DE scales (H1a). We found that the emotion regulation subscale correlated highly with the total scale, showed very high reliability and a normal distribution. Therefore, we would conclude that emotion regulation can very well be assessed dimensionally as a subscale of the SWAN-DE questionnaires. Furthermore, we assessed the distribution and reliability of the self- and observer report data (H1b). As expected, the SWAN-DE scales depicted a normal distribution of both the self- and the observer report data. Reliability of the SWAN-DE-SB was slightly lower than in a previous sample (Blume et al., 2020). This previous sample, however, included a general population and a clinical sample, which might have enhanced the reliability by widening the heterogeneity of the values (Blume et al., 2020). In the current sample, we found higher homogeneity of the values, which might lead to lower reliability. Nevertheless, both reliabilities of the self- as well as the observer report were acceptable to good in all three scales of our sample.

Moreover, we hypothesized that the self- and observer report version of the SWAN-DE measure the same underlying constructs (H1c). Since we newly developed the observer version of the SWAN scales and added a subscale assessing emotion regulation to both versions, we tested whether the internal structure of the self- and the observer report scales were comparable. Building a model with the emotion regulation scale did not fit the data very well. A previous meta-analysis concluded that emotion regulation problems are only found in 30-70% of adults with an ADHD diagnosis (Shaw et al., 2014). Therefore, it might be that emotion regulation problems do not load on the same general ADHD factor as inattention and hyperactivity/impulsivity. Additionally, the measurement invariance analysis showed that target persons and observers interpreted the items differently. This might reflect a problem in the wording of the items. Constructing neutrally worded items measuring emotion regulation was a challenge, since all existing emotion regulation scales we found are deficit oriented (e.g., irritable, throw tantrums, unpredictable mood; Christiansen et al., 2012). Future studies should try to develop more appropriate neutrally formulated measures of emotion regulation to investigate whether our lack of findings was indeed grounded on unsuitable items or on a lack of association between emotion regulation and inattention, hyperactivity/impulsivity in the general adult population. Next, we explored whether excluding the emotion regulation scales would enhance the comparability of the internal structure of the SWAN-DE scales in self- and observer report. Measurement invariance analyses confirmed the good psychometric properties of the originally developed SWAN-DE scales with only items of inattention and hyperactivity/impulsivity included. The results suggest that target persons and significant others seem to understand the items similarly and describe similar behavior by them. Metric invariance was confirmed, giving evidence to equivalent factors loadings in the two groups. Only the partial scalar invariance hints towards different intercepts in single items. Even partial strict measurement invariance could be achieved, suggesting equal residual variance between self-report and observer report data. These results are positively surprising, since measurement invariance could not be confirmed between parent- and teacher ratings on the SWAN (Jungersen & Lonigan, 2021).

Our second hypothesis expected a moderate correlation of the self- and the observer report of the SWAN-DE (H2). Like in previous studies which compare self- and observer reports of ADHD symptoms, the reports correlated moderately with each other (e.g., Belendiuk et al., 2007; Kooij et al., 2008; Van Voorhees et al., 2011; Zucker et al., 2002). From earlier research we would have expected the inter-rater correlation to be higher for hyperactivity/impulsivity than for inattention (H2a), since hyperactivity/impulsivity symptoms seem to be easier detectable from an outside perspective (Belendiuk et al., 2007). This hypothesis, however, could not be confirmed, correlation between self- and observer report for inattention was $r = .41$ for both subscales. In the previous study (Belendiuk et al., 2007), mothers of children with an ADHD diagnosis were assessed. The authors suspect that this very specific kind of sample has a high knowledge of ADHD symptoms and their effects in daily life and therefore might be especially observant when it comes to their own hyperactivity/impulsivity symptoms. In our study, however, we sampled a general population sample which might have the previous knowledge and therefore the correlation between self- and observer ratings is the same for the inattention as well as the hyperactivity/impulsivity scale. Self-report of ADHD symptoms was significantly more negative than observer report (H2b), which had also been found in previous research (Murphy & Schachar, 2000).

A new research question that we could examine because of the dimensional structure of the SWAN-DE scales was the question whether target persons rated themselves only more extreme on the negative end of the dimension (weaknesses) or also more extreme on the positive end (strengths). We expected the target persons to tend to more extreme self-reports as well in their strengths as in their weaknesses than the observer report of the significant others would (H2c). Regression analyses suggest that, indeed, self-reports tend to be more extreme on both sides of the continuum than observer reports. The finding that, on average, the self-reports are more negative than observer reports can be explained with a greater difference between self- and observer reports on the negative end than on the positive end. This might be explained by social biases and the feelings of individuals that they are more distinct to other people, than they actually are. When persons detect strengths in their behavior, they interpret them as very distinct to other people around them, a bias called illusory superiority (Hoorens, 1993). On the other hand, when persons experience weaknesses, they would imply that other people do not have comparable difficulties, which might be attributed to the negativity bias (Vaish et al., 2008). The more pronounced difference between self- and observer ratings on the negative end might be explained by the social desirability bias in the observers (Krumpal, 2013). This theory implies that individuals feel more comfortable giving socially accepted answers and might therefore euphemize weaknesses in the target persons but not their strengths. Other possibilities might be the lack of insight and context that significant others have or the fundamental attribution error, where behavior of others is attributed more on their personality than on the situation (Malle, 2006). However, we cannot test these suggestions with the current study and additional research is needed to better understand the reasons for the difference between self- and observer ratings.

Besides analyzing the psychometric qualities of the SWAN-DE-SB and the newly developed SWAN-DE-FB, in our third hypothesis we attempted to explain differences between the ratings with demographic (gender and age), relationship (partner, friend, family member, living in one household) and disorder specific characteristics (strength of the symptoms) of the target person as well as the observer (H3). Due to a lack of systematic research findings, we performed an explorative analysis, including all possible explanatory variables into the regression models at the same time. To explain both, magnitude and direction of the difference, regression models were first built with the total value of the difference (only allowing positive values, H3a) and then with the difference itself (allowing positive and negative values, H3b). Separate models were built for the total ADHD scale, and for each of the subscales inattention, hyperactivity/impulsivity, and emotion regulation.

We investigated whether the demographic characteristics like age and gender of the target person or the significant other were associated with magnitude and direction of the difference of their ratings. Age of the target person was a significant predictor of magnitude of the difference for the hyperactivity/impulsivity scale. A previous longitudinal analysis has found that differences in reports of internalizing problems increase with age (van der Ende et al., 2012). In our study, this seems to also hold for externalizing symptoms. Concerning the direction of the difference, we found a negative association of the age of the target person. That means that within our study sample, with increasing age older target persons reported to have more hyperactivity/impulsivity symptoms themselves than their significant other reported. Age of the observer was significantly negatively associated with the magnitude of difference between self- and observer ratings on emotion regulation, indicating that with increasing age of the observer the ratings were more similar. We could think of several explanations for this finding, for example that older people might know each other for a longer time and therefore have aligned their impressions of each other more closely. Unfortunately, since we have not inquired about the duration of the relationship, we are not able to test this interpretation. When the observer indicated to be female, they evaluated the target person significantly more positive than the self-report, compared to male observers. This might be explained by gender effects in the social desirability bias, because women tend to conform more to social rules and therefore might be more polite in their judgment of others (Dalton & Ortegren, 2011).

None of the relationship variables we included in the models was significantly associated with the magnitude or direction of difference between self- and observer ratings. Even though previous research did not find any difference between ratings from partners, friends and families (Van Voorhees et al., 2011), we would have expected that at least living together would have an impact on the contexts in which the target person and significant other see each other and the insight that the observer has. We will discuss below possible explanations for the lack of findings.

A high association with both, the magnitude and the direction of the difference between self- and observer report, on the total as well as the three subscales is found for observer reported ADHD symptom strength. According to our data, the higher the significant other reported the strength of the symptoms, the bigger the difference between self- and observer ratings. Contrary to our expectations, the target persons with higher ADHD symptoms rated themselves significantly more negative than the significant other.

These findings imply that even when a person experiences very high inattention, hyperactivity/impulsivity, and difficulties in emotion regulation, significant others would report a much lower symptomatology. Different to what we expected from Smith and colleagues (2000), the lack of insight does not seem to lie at the target person but at the significant other, who does not seem to see the symptoms as gravely. One explanation for these diverging results might be the fact, that Smith and colleagues (2000) examined adolescents with an ADHD diagnosis, while we investigated inattention, hyperactivity/impulsivity and emotion regulation in adults of the general population. Research has shown, that adolescents have a more positive view on themselves than their parents have (e.g., Hughes et al., 2009). Additionally, it might be that with growth and maturation, humans better learn to hide or compensate their symptoms of inattention and hyperactivity/impulsivity. Our finding that with increasing age of the target person, the difference between self- and observer report widens, would fit into that assumption. However, this interpretation should be taken cautiously since our sample was mainly composed by young adults. Due to the low number of older adults, this finding could be coincidental.

As explained above, according to the Operations Triad Model, multiple observer reports can either be handled by Converging Operations, where all differences are taken as measurement errors, by Diverging Operations, where it is assumed that the difference between the ratings incorporate meaningful information, or by Compensating Operations, with the premise that one of the observers lacks insight or context to correctly evaluate behavior (see Figure 12; Alexander et al., 2017; De Los Reyes et al., 2013). When relating our results to the Operations Triad Model, we would conclude that the SWAN-DE scales do not represent Converging Operations, since self- and observer report differ significantly from each other. According to our data, these discrepancies seem to stem more from a lack of insight that the observer has into the symptoms of the target person, than from a different understanding of the disorder. However, we would need more research to confirm these findings and infer reliable rules of conduct concerning the diagnosis of ADHD.

Limitations and Future Directions

The current study has helped to give some further insight into the associations of self- and observer reports of ADHD symptoms in adults and has confirmed the SWAN-DE as a reliable scale. Although this study broadens our understanding of the association of self- and observer ratings of ADHD symptoms in adults, we acknowledge that future studies should try to avoid some of the limitations our research possesses. Additionally, much research is still needed until we can completely understand how self- and observer reports are composed. First, like many psychological studies conducted in a university context, participants in our sample showed to be mainly female, on average comparably young and highly educated. Especially the influence of age on the difference between self- and observer reports should be investigated more thoroughly in a sample with a more balanced age range in future studies.

To our surprise, the type of relationship did not show any significant association with the difference between self- and observer reports. Disregarding the probability that such an effect indeed does not exist, the lack of findings could be explained by two reasons: construction of the relationship variables and statistical analyses. First, we could have defined the type of relationship differently, for example asking how much time the target person and significant other spend together, how long they know each other or in which contexts they see each other. These relationship variables might differ highly in partnerships, friendships, families and people living together but might have a big impact on the insight into ADHD symptoms. Second, the high influence of the strength of ADHD symptoms on observer difference might have covered smaller effects of relationship on these differences. Additionally, it is possible that the strength of ADHD symptoms influences the number and quality of close relationships an individual has. Research has shown that adults with higher ADHD symptoms experience more difficulties in their social and romantic relationships (Sciberras et al., 2009). Especially difficulties in emotion regulation might hinder positive interpersonal contact with other individuals (Bodalski et al., 2019). This might then also influence how close the target person is with the significant other that is providing the observer report. To our knowledge, it has not yet been investigated whether adults with higher ADHD symptoms live on their own more often than adults with lower ADHD symptoms, but it might have influenced our results. Thorough investigation of the closeness and quality of relationship between target person and significant other might therefore significantly enhance future research.

Furthermore, the association of subjective and objective measurements of ADHD symptoms should be investigated more thoroughly. Previous research has found that self- and observer report only moderately predict objective measures of intelligence and executive functions and that neither of the reporters was significantly better in this prediction (Alexander & Liljequist, 2016). It might be interesting to investigate whether alternative objective measures of inattention and impulsivity (e.g.; continuous performance tasks; Emser et al., 2018; Hall et al., 2016) or hyperactivity (e.g., measures of movement via actigraphy; Boonstra et al., 2007) might show higher association with subjective scales. Additionally, future research should examine whether the higher variance of the SWAN-DE, compared to the previously used categorical scales, might have higher potential to be associated with objective measures.

Another interpretation of the usefulness of self- and observer reports might be their predictive value of future outcomes. If either self- or observer report of ADHD symptoms better predict outcomes like educational/vocational success, social relationships, health or substance abuse, these should be considered preferentially when conducting research about interventions.

The SWAN-DE as a newly developed scale should also be investigated more thoroughly concerning its stability over time, for example with daily measures (e.g., ambulatory assessment/ecological momentary assessment) and longitudinal studies. With help of these measurement techniques, short-term fluctuations, as well as long-term developmental changes can be depicted (Nesselroade, 1991). Additionally, the SWAN-DE scales might be analyzed more thoroughly for use with clinical samples (Blume et al., 2020). To use them in a clinical population, the discriminant validity of the scales should be examined thoroughly to show that the SWAN-DE is specifically useful to diagnose ADHD but no other psychiatric disorders like major depression.

Conclusion

Measuring ADHD symptoms dimensionally in the general population has many statistical and practical advantages. With the presented study we could show that the SWAN-DE scales seem to be reliable and valid scales, both in the self- as in the observer-report version. The association of emotion regulation with ADHD symptoms must be investigated more thoroughly, however. A novelty that the dimensional SWAN-DE scales enable is the finding that self-reports are more extreme on both the strengths as well as the weaknesses side of ADHD symptoms. We also found evidence for the influence of demographic variables (age & gender) as well as symptom strength on the concordance of self- and observer reports. Future studies should investigate more thoroughly possible biases leading to diverging ratings on the SWAN-DE scales.

Appendix

Table 20

SWAN-DE-SB and SWAN-DE-FB: German items and English translation.

	Self-Report	Observer-Report	English Translation
1	Ich beachte Details genau und vermeide Flüchtigkeitsfehler.	%name% betrachtet Details genau und vermeidet Flüchtigkeitsfehler.	Give close attention to detail and avoid careless mistakes.
2	Ich halte die Aufmerksamkeit bei Aufgaben und Aktivitäten aufrecht	%name% hält die Aufmerksamkeit bei Aufgaben und Aktivitäten aufrecht.	Sustain attention on tasks or activities.
3	Ich höre zu, wenn ich direkt angesprochen werde.	%name% hört zu, wenn er/sie direkt angesprochen wird.	Listen when spoken to directly.
4	Ich bringe angefangene Tätigkeiten zu Ende.	%name% bringt angefangene Tätigkeiten zu Ende.	Follow through on instructions and finish chores.
5	Ich organisiere meine Aufgaben und Aktivitäten.	%name% organisiert eigene Aufgaben und Aktivitäten.	Organize tasks and activities.
6	Ich beschäftige mich freiwillig mit Aufgaben, die anhaltende geistige Anstrengung erfordern.	%name% beschäftigt sich freiwillig mit Aufgaben, die anhaltende geistige Anstrengung erfordern.	Engage in tasks that require sustained mental effort.
7	Ich behalte den Überblick über Gegenstände, die für meine Aktivitäten erforderlich sind.	%name% behält den Überblick über Gegenstände, die für die eigenen Aktivitäten erforderlich sind.	Keep track of things necessary for activities.
8	Ich ignoriere äußere Reize.	%name% ignoriert äußere Reize.	Ignore extraneous stimuli.
9	Ich behalte alltägliche Aktivitäten im Gedächtnis.	%name% behält alltägliche Aktivitäten im Gedächtnis.	Remember daily activities.
10	Ich sitze still.	%name% sitzt still.	Sit still.
11	Ich bleibe sitzen, wenn es Regeln oder soziale Konventionen erfordern.	%name% bleibt sitzen, wenn es Regeln oder soziale Konventionen erfordern.	Stay seated when required by rules or social conventions.
12	Ich reguliere meine motorische Aktivität.	%name% reguliert eigene motorische Aktivität."	Modulate motor activity.
13	Ich halte einen der Situation angemessenen Geräuschpegel.	%name% hält einen der Situation angemessenen Geräuschpegel.	Keep noise level reasonable.
14	Ich komme zur Ruhe und ruhe mich aus.	%name% kommt zur Ruhe und ruht sich aus.	Settle down and rest.
15	Ich denke über Fragen nach (bevor ich mit einer Antwort herausplatze).	%name% denkt über Fragen nach (bevor er/sie mit einer Antwort herausplatzt).	Reflect on questions (control blurting out answers).
16	Ich warte geduldig, bis ich an der Reihe bin.	%name% wartet geduldig, bis er/sie an der Reihe ist.	Await turn.
17	Ich steige in laufende Gespräche ein, ohne diese zu unterbrechen und zu stören.	%name% steigt in laufende Gespräche ein, ohne diese zu unterbrechen und zu stören.	Enter into conversations without interrupting or intruding.

18	Ich reguliere meine verbale Aktivität und kontrolliere übermäßiges Reden.	%name% reguliert eigene verbale Aktivität und kontrolliert übermäßiges Reden.	Modulate verbal activity and control excess talking.
19	Ich habe meine Laune unter Kontrolle.	%name% hat seine/ihre Launen unter Kontrolle.	Keep emotions under control.
20	Ich bleibe auch in schwierigen Situationen ruhig und bedacht.	%name% bleibt auch in schwierigen Situationen ruhig und bedacht.	Stay calm in difficult situations.
21	Meine Stimmung ist über verschiedene Situationen hinweg stabil.	%name%s Stimmung ist über verschiedene Situationen hinweg stabil.	Stable mood in different situations.
22	Kurze Phasen, in denen ich traurig oder entmutigt bin, habe ich selten.	Kurze Phasen, in denen %name% traurig oder entmutigt ist, hat er/sie selten.	Rarely short phases of sadness.
23	Ich kann mich lange mit einer Sache beschäftigen.	%name% kann sich lange mit einer Sache beschäftigen.	Keep engaged with a task for a long time.
24	Auch in stressigen Situationen bleibe ich innerlich ruhig.	Auch in stressigen Situationen bleibt %name% innerlich ruhig.	Stay calm on the inside even in stressful situations.
25	Mit Alltagsstress komme ich gut zurecht.	Mit Alltagsstress kommt %name% gut zurecht.	Cope well with daily stressors.
26	Wenn etwas nicht funktioniert, wie ich es mir vorgestellt habe, ist das für mich in Ordnung.	Wenn etwas nicht funktioniert, wie %name% es sich vorgestellt hat, ist das für ihn/sie in Ordnung.	Cope well and accept when something does not work out.
27	Phasen, in denen ich sehr aufgedreht bin, habe ich selten.	Phasen, in denen %name% sehr aufgedreht ist, hat er/sie selten.	Rarely phases of overexcitement.

Note. All items were rated on a 7-point Likert scale ranging from (0) *far below average* (weit unterdurchschnittlich) to (6) *far above average* (weit überdurchschnittlich); German instruction: "Menschen unterscheiden sich in ihren Fähigkeiten, Aufmerksamkeit zu fokussieren, Aktivität zu kontrollieren und Impulse zu unterdrücken. Wie schätzen Sie sich selbst im Vergleich zu anderen Erwachsenen Ihres Alters bezüglich jeder der unten aufgelisteten Aussagen ein? Bitte wählen Sie die beste Einschätzung basierend auf Ihrem Erleben und Verhalten im vergangenen Monat. Falls Sie aufgrund einer ADHS-Diagnose Medikamente einnehmen (z.B. Methylphenidad, Atomoxetin, Dexamphetamin, Lisdexamfetamin, Guanfacin), geben Sie bitte Ihre Einschätzung, wie Sie sich verhalten, wenn Sie keine Medikamente einnehmen. Verglichen mit anderen Erwachsenen in ihrem Alter, wie gut können Sie Folgendes:

5. General Discussion

When we think about how to best measure self-regulation, we should first think about *why* we want to measure self-regulation. Do we want to diagnose a psychological disorder, like ADHD, and use this diagnosis to decide who is eligible for support and treatment? Do we want to measure whether interventions and trainings have an effect on the self-regulation ability of the participants? Do we want to investigate differences in self-regulation within the population or within individuals to understand the concept of self-regulation better? Or do we want to understand how self-regulation is related to different life outcomes, like educational success, health behavior, social relationships, or life satisfaction? All of these goals are very important in scientific research and clinical practice to enhance our understanding of human behavior and ultimately support individuals in their personal life course.

The findings that I presented in the three manuscripts lead to the conclusions that integrating differences between as well as fluctuations and variance within individuals, investigating associations with internal and external factors, measuring self-regulation in a dimensional way, and involving different reporters are promising tools to better depict self-regulation of humans in research and clinical practice. However, one overarching result of the different manuscripts within this thesis is that we need much more research to better define how to measure self-regulation. Another finding is, that the best way to measure self-regulation depends heavily on the research question, the aim of the study, the context in which it is measured, and the characteristics of the sample measured. Below, I will describe these conclusions more specifically.

5.1. Inter- vs. Intraindividual Differences in Self-Regulation

The first two of the depicted manuscripts investigated self-regulation and ADHD symptoms on the inter- and the intraindividual level. These two levels are essential to distinguish in psychological research, because they imply different theoretical, practical, and methodological interpretations and can even reveal contrary effects of the same constructs (Curran & Bauer, 2011; Molenaar & Campbell, 2009; Wang & Maxwell, 2015).

In manuscripts one and two, we have distinguished the inter- and the intraindividual level of self-regulation or ADHD symptoms to account for this issue. In both studies we have calculated the intraclass correlation coefficients (ICC) to determine the proportion of the total variance that can be explained by interindividual differences. The ICCs for self-regulation were 0.43 in manuscript one and 0.42 in manuscript two respectively. Thus 43/42% of the variance in self-regulation could be explained by interindividual differences between the participants and the remaining 57/58% of the variance were composed of intraindividual fluctuations and measurement error. This indicates that the inter- and the intraindividual level contribute significantly to the variance in measurements and should therefore both be included in the investigation of self-regulation abilities.

In the second manuscript, we investigated the fluctuations of self-regulation on the intraindividual level and the timescales on which these fluctuations occur more thoroughly. Graphical depictions of the fluctuations in self-regulation showed a ceiling effect, where most responses indicated that children did not experience symptoms of inattention and/or hyperactivity/impulsivity. However, variance in self-regulation is still detectable in the visual representations. In chapter 6, I will discuss in more detail, how this ceiling effect might be a consequence of the scales that were used and how this might be prevented in future by implementing dimensional scales. Decomposing daily variance according to timescale levels revealed that approximately half of the variance could be found on the occasion level (morning, afternoon, evening) and the other half of the variance on the day level. For children, this means that both, the self-regulation level on a specific occasion and the self-regulation on a specific day seem contribute to the indicated self-regulation. Self-regulation ability might thus fluctuate both, relatively fast from occasion to occasion and more slowly from day to day. This is especially important to know, when we consider different timescales on which to measure self-regulation in the daily life of children via ambulatory assessment.

When we investigated associations of self-regulation with internal factors like sleep and executive functions, analyses revealed different effects on the inter- and the intraindividual level. First, when depicting the association of daytime sleepiness and ADHD symptoms in manuscript one, we found no effect on the interindividual level but a negative association of daytime sleepiness on ADHD symptoms on the intraindividual level. This indicates that children who were sleepier in general did not differ in their overall ADHD symptoms from children that indicated feeling more awake, but in moments when children felt sleepier during the day, they indicated less ADHD symptoms. Explorative analyses separating the two core symptoms of inattention and hyperactivity/impulsivity indicated that only hyperactivity/impulsivity was negatively related to daytime sleepiness on the intraindividual level. On the interindividual level, however, daytime sleepiness was significantly positively related with symptoms of inattention. This implies that children who feel in general more tired than others also report higher inattention in general.

Second, we investigated the association of self-regulation with working-memory performance on the inter- and the intraindividual level. We found that the easier working memory task (Load 2) was significantly associated with self-regulation on the intraindividual level, indicating that in moments with high performance on the task, children also indicated high self-regulation abilities. For the more difficult version of the working memory task (Load 3), the picture was more ambiguous. When only time variables and working memory were included in the model, the association between self-regulation and working-memory performance was significant on both, the inter- as well as the intraindividual level. However, accounting for autocorrelation only left a significant association on the interindividual level and including medication as a control variable lead to a model were no association between self-regulation and working-memory performance was statistically significant, neither on the inter- nor on the intraindividual level. Possible explanations for and implications from these findings will be discussed in more detail in the next chapter.

To conclude, distinguishing the inter- from the intraindividual level in our analyses investigating the associations of self-regulation and internal factors like night sleep, daytime sleepiness, and working-memory performance helped us to gather more information about the relation of the constructs on the trait as well as on the state level. We could show that some associations were only found on one of the two levels. This implies that future theoretical considerations should specify more carefully, whether they assume mechanisms to be present only on the interindividual, only on the intraindividual, or on both levels. Theoretical models which already incorporate fluctuations of ADHD symptoms and self-regulation are for example the state regulation model, where symptoms of inattention and hyperactivity/impulsivity stem from an attempt of the individual to outbalance a state of under- or over-arousal (Van der Meere, 2005), and the strength model of self-control which implies that self-control fluctuates depending on internal and external resources (Baumeister et al., 2007). Nevertheless, the findings from manuscripts one and two show that both levels, the inter- as well as the intraindividual level, should be considered when assessing self-regulation abilities and interrelated constructs. To our knowledge, research has shown growing awareness to the two levels in the recent years by increasingly implementing ambulatory assessment studies in empirical investigations (Koch et al., 2021). In clinical psychological praxis, according to the DSM-5 the diagnosis of ADHD calls for “a persistent pattern of inattention and/or hyperactivity-impulsivity that interferes with functioning or development” (314.01/F90.01; American Psychiatric Association, 2013). This definition accounts for the intraindividual fluctuations of self-regulation and should be incorporated in diagnostic measurement approaches of ADHD. Interestingly, first approaches have been developed to implement investigation of intraindividual fluctuations of different psychological states in clinical practice, for example with help of the clinical application m-Path (*M-Path*, 2023), where data from patients is gathered in daily life via smartphones and directly forwarded to the therapist. This might lead to more informed therapy occasions where contextual factors are exceedingly included in the planning and evaluation of interventions. From the methodological perspective, multilevel models already allow to distinguish inter- and intraindividual effects statistically (Mehl & Conner, 2012). However, depicting inter- and intraindividual differences of psychological constructs in the daily life of participants still poses many challenges on the analyses. For example, the differentiation of inter- and intraindividual effects might be complicated by growth effects (Wang & Maxwell, 2015). Research has initiated to include new methodological approaches within the framework of multilevel models, for example Bayesian statistics (Curran & Bauer, 2011). These new developments might in future help to better evaluate differences between and fluctuations within individuals in self-regulation, ADHD symptoms and internal and external factors like sleep and executive functions.

5.2. Correlates and Antecedents of Self-Regulation

Like any psychological state, the ability to self-regulate one’s cognition and behavior does not just appear *out of nowhere*. Antecedents might trigger high or low self-regulation, correlates underlying the same basic mechanisms might occur at the same time, and consequences might follow from high or low self-regulation. Measuring self-regulation repeatedly in the daily lives of individuals allows for a better understanding about the associations of self-regulation with different internal and external factors. Next to correlations of interindividual differences, systematic intraindividual associations of the fluctuations are providing useful evidence for a dependency of the constructs.

5.2.1. Sleep

The investigation of the association of ADHD symptoms, night sleep, and daytime sleepiness we conducted in manuscript one provided some interesting insight. We did not find any of the hypothesized inter- or intraindividual associations ADHD symptoms with night sleep or daytime sleepiness, but instead we found a negative intraindividual association of ADHD symptoms and daytime sleepiness. Contrary to our expectations, children indicated to experience less ADHD symptoms in moments when they felt sleepier during the day. In an explorative attempt to further examine these findings, we then separated the ADHD symptoms into two subscales, one for inattention and one for hyperactivity/impulsivity. Rerunning the analyses with each of the subscales as explanatory variables showed a negative intraindividual association of inattention and night sleep (children indicated lower inattention after a night of good sleep). No association was found between hyperactivity/impulsivity and night sleep, neither on the inter- nor on the intraindividual level. Inattention was significantly positively associated with daytime sleepiness on the interindividual level, implying that children who indicated on average higher inattention than other children also in general felt sleepier. In contrast to that, we found a negative intraindividual association of hyperactivity/impulsivity with daytime sleepiness. This suggests that in moments when children reported less hyperactivity/impulsivity they indicated higher daytime sleepiness, than in moments where they felt more awake.

This finding in the explorative analyses, where we separated the analyses on basis of the core symptoms of ADHD, is especially interesting, since it suggests that inattention and hyperactivity/impulsivity might be differently associated with internal or external factors like night sleep or daytime sleepiness. Each core symptom might therefore, according to our findings, have different precursors, correlates and consequences. This fits into the theory of ADHD as a disorder with multiple causations and different phenotypes (Kooij et al., 2019).

Theory nowadays assumes that ADHD symptoms stem from an interaction of biological, psychological and social origins (Döpfner et al., 2020; Nigg et al., 2020). Additionally, the DSM-5 discriminates three subtypes of ADHD, the prevalently inattentive, prevalently hyperactive/impulsive, and the combined subtype (American Psychiatric Association, 2013). Therefore, it might be possible that the subtypes of ADHD have different etiologies. Previous research has found that individuals with different ADHD subtypes also experience different patterns of sleep problems and daytime sleepiness (Mayes et al., 2008). It might be interesting to investigate in future studies, whether the symptoms inattention and hyperactivity/impulsivity fluctuate differently during the daily life of individuals and which internal and external factors are associated with only one or both core symptoms of ADHD.

Concerning the temporal sequence, we only investigated the association of night sleep with ADHD symptoms during the following school day, and daytime sleepiness with ADHD symptoms at the same time. However, the association of ADHD symptoms and sleep might be much more varied – researchers have hypothesized that ADHD symptoms might lead to worse night sleep, that medication might hinder or even improve sleep or that higher ADHD symptoms might influence daytime sleepiness (Hvolby, 2015). Therefore, sleep quality could be a precursor, correlate, or consequence of self-regulation, and there might be bidirectional interactions. Additionally, this association might be mediated and moderated by other internal and external factors like medication, social interactions, and physical activity. To sum up, the association of self-regulation with night sleep and daytime sleepiness should be further investigated on the inter- and the intraindividual level, to better understand the mechanisms underlying the relationships as well as their directions.

5.2.2. Working Memory

In manuscript two, we investigated the relationship of self-regulation with another internal factor – working-memory performance as a measure of executive functions – on the inter- and the intraindividual level. The associations of self-regulation and working memory were diverging on the two levels. On the interindividual level, we could not find any association of self-regulation and the easier load (Load 2) of the working memory task. Visual inspection of the data showed a pronounced ceiling effect of the working memory task with most children scoring on the upper level of the scale. Therefore, variability between children was very low which might have disenabled the detection of any association between self-regulation and working-memory performance on the interindividual level. We concluded that, because of the age of the participants, the task might have been too easy for them. Previous research had used the working memory updating task in ambulatory assessment studies with younger children, where adequate variability was found on both load levels (e.g., Dirk & Schmiedek, 2017). Future studies should therefore adapt the degree of difficulty more carefully to the age of the children. Since we found an interindividual effect on self-regulation for Load 3 of the working memory task (when only time and working memory were included in the model), we conclude that this task might have better represented the interindividual differences of the participants. However, this effect became insignificant when ADHD related medication was included in the model. All children in the sample that had an ADHD diagnosis received medication with methylphenidate. Medication (and correspondingly ADHD diagnosis) was negatively related to self-regulation. This implies that the children with an ADHD diagnosis indicated significantly lower self-regulation (i.e., higher inattention and hyperactivity/impulsivity) and this effect was so strong that it covered the weaker association of general working-memory performance and self-regulation.

On the intraindividual level, we could find an association of self-regulation and the performance in the easier load of the working memory task (Load 2). This finding hints towards the association of self-regulation abilities and working-memory performance within individuals and therefore fits well in the theoretical considerations where working memory supports self-regulation through the active representation of goals, focusing attention on these goals, suppress intrusive interference or thoughts and controls undesired emotions (Hofmann et al., 2012). The finding might also be taken as indicator for an analogous reliance of self-regulation and working-memory performance on the active control of attention as a resource (Ilkowska & Engle, 2010b). The intraindividual effect on self-regulation we found for Load 3 of the working memory task disappeared as soon as we included autocorrelation into the model. Autocorrelation indicates that data from measurement timepoints which were close to each other were more similar than data from measurement timepoints which were far apart. Since we have found that fluctuations are composed of variance on different time levels (days and occasions), autocorrelation might indicate that the fluctuation on the higher time level (i.e., days) covers the association we found on the lower time level (i.e., occasions). In the daily life of children this would imply that when a child shows relatively high self-regulation and working-memory performance in the morning, both, self-regulation and working-memory performance are presumably also high in the afternoon of the same day. Taken together, self-regulation and working-memory seem to be associated on the inter- as well as on the intraindividual level, however, these associations seem to be influenced heavily by additional factors, for example an ADHD diagnosis and respective medication. Future analyses should try to replicate the found associations and investigate more carefully on which time levels the intraindividual fluctuations should best be measured.

5.3. Dimensional versus Categorical Perspective

Psychological disorders have originally been defined as being categorically distinct from normal psychological functioning. However, for some time now, theory and empirical evidence have increasingly hinted towards a dimensional structure of many disorders, including ADHD (e.g., Coghill & Sonuga-Barke, 2012; Larsson et al., 2012; Marcus & Barry, 2011). The dimensional view of ADHD makes it possible to position individuals on a continuous variable ranging from high self-regulation of cognition and behavior to high inattention and hyperactivity/impulsivity (Arnett et al., 2013; Swanson et al., 2012). In our general population sample in manuscript three, we found normal distributions of the general ADHD scale as well as all subscales of the SWAN-DE, as well for the self- as for the observer report data.

Next to confirming the normal distribution of inattention and hyperactivity/impulsivity in the general population, in manuscript three we assessed how a dimensional scale of emotion regulation might be related to the core symptoms of ADHD. Three models about the relationship between emotion regulation and ADHD symptoms are currently discussed (Shaw et al., 2014). The first model hypothesizes that emotion dysregulation might be interpreted as a third core symptom of ADHD. Previous research has found evidence for emotion regulation being a core symptom next to inattention and hyperactivity/impulsivity (Hirsch et al., 2018). However, since the confirmatory factor analysis in our study indicated considerably worse model fit for the model of the SWAN-DE including emotion regulation, than for the model with only inattention and hyperactivity/impulsivity, we would interpret this against the hypothesis that emotion regulation might be a core symptom of ADHD. However, this conclusion has to be taken cautiously, since the finding might also be based on the newly developed items, a limitation which will be discussed more thoroughly in chapter 6. Future studies with revised items might come to a different conclusion. The second model, according to Shaw and colleagues (2014), distinguishes two groups of individuals with ADHD, one group with and one without emotion regulation problems. We did not test this hypothesis in our study. However, we found that the emotion regulation subscale produced a normal distribution in the general population. Therefore, we would imply that a categorical differentiation of individuals with and without emotion regulation difficulties should not be applied, but, like the core symptoms of ADHD, emotion regulation should be assessed dimensionally in the population.

Finally, the third model sees emotion regulation as an individual factor which is strongly associated with ADHD symptoms but still distinct (Shaw et al., 2014). Our data seems to best apply to this third model. The emotion regulation subscale in our sample is significantly correlated to the general ADHD scale, but the confirmatory factor analysis shows that the subscale does seem to be distinct from the general ADHD factor. To conclude, we could not convincingly confirm any of the three models about the association of emotion regulation with ADHD symptoms. Our data hints more towards the third model, where emotion regulation is a distinct but highly related factor in ADHD. Future research might investigate this question more closely with help of reliable and valid (in the best case dimensional) emotion regulation scales.

Dimensional measurement has many advantages in research and clinical practice. Most statistical tests rely on the assumption that data is normally distributed. Categorical scales, however, produce a highly skewed distribution in the general population, where most individuals fall into the *zero* or *no symptoms* category. A dimensional scale depicts data with a normal distribution since variance can equally be detected on the positive as on the negative end of the scale. Another advantage is, that the 7-point Likert scale of the SWAN-DE is much more fine-grained than previous scales. This enables a more detailed distinction between individuals and allows for more precise investigations of individual differences. The higher differentiation also makes it possible to assess changes in symptoms within a person over time more accurately. Especially if these changes, for example after an intervention, are too small to provoke a change in category (from *ADHD* to *no ADHD*), the SWAN-DE scales depict these changes more reliably.

It might be argued that we need categorical data for decisions about diagnoses and treatment in clinical practice. Initial *Receiver Operating Characteristic* (ROC) Analyses have shown that it is possible to determine cutoff scores within the SWAN-DE, which distinguish

ADHD symptoms from normal behavior equivalently to other ADHD scales (Blume et al., 2020). These cutoff scores certainly must be confirmed with help of comprehensive validation studies in clinical and general population samples, before they can be used for diagnostic purposes. However, we argue that these dimensional ADHD scales are a promising tool to protect the self-esteem of individuals during the diagnostic process, since they allow to report individual strengths next to the weaknesses in self-regulation of cognition and behavior.

5.4. Self-Report vs Observer Report

Who is eligible to report about symptoms of inattention and hyperactivity/impulsivity in a diagnostic process? How much do different observers agree in their ratings and how can differences be explained? These questions are very important for empirical research concerning self-regulation but also for the diagnostic process in clinical practice, however, they have not been researched exhaustively in the past. In my third manuscript, I have been trying to examine these questions a bit further. A sample of 142 adults provided self-ratings of their ADHD symptoms on the dimensional SWAN scales and recruited a significant other (partner, friend, family member) to contribute an observer report on the same scale adapted for observer report. Data showed a moderate but significant correlation between reporters, with self-reports being more negative on average than observer reports. Further analyses, however, showed that self-reports were more extreme, both at the positive as well as on the negative side of the dimensional scales. The finding of more negative self-reports on average is explained by a higher discrepancy between self- and observer report the more negative the observer report is.

This finding is striking and represents a novelty in the research of observer agreements. Since previous scales of ADHD symptoms were focusing on the weaknesses and did not investigate the strengths in self-regulation of cognition and behavior, it was not possible to investigate whether individuals would also see their strengths more positive than a significant other. Due to the dimensional scale depicting strengths as well as weaknesses in ADHD symptoms and self-regulation, we were able to show that self-reports were indeed more extreme on both sides of the scale. These findings might lead to the conclusion that individuals perceive themselves as more distinctive from other people both on the positive (i.e., illusory superiority bias; Hoorens, 1993) as well as on the negative side (i.e., negativitiy bias; Vaish et al., 2008) than an external observer. For clinical practice this might lead to the important notion that individuals will probably rate their impairments in daily life higher than an outsider would do, but at the same time that they might have higher faith in their individual resources to cope with these challenges.

To control for the possibility that the difference in ratings is caused by a different understanding of the applied scales, we investigated the self- and the observer reports for measurement invariance. This technique implies fitting structural equation models to the two groups and stepwise increasing the rules in which the models are supposed to match. First, just the assignment of the items to the latent factors is tested (configural invariance), then it is tested whether the factor loadings of the items are the same in both groups (metric or weak invariance). If metric (weak) invariance applies, we can test whether the intercepts are equal in both groups (scalar or strong invariance) and finally, a model is tested where factor loadings, intercepts and residuals are constrained to be equal across groups is tested (residual or strict invariance). Each model is tested against the precedent with a likelihood ratio test. In case of a non-significant difference between the models, invariance applies. When the difference is significant, partial invariance can be tested by alleviating the restrictions for some of the items (Jungersen & Lonigan, 2021).

In our analyses, we first tested a model where the newly developed subscale of emotion regulation was included next to the previously researched subscales inattention and hyperactivity/impulsivity. The data showed weak model fit, not approaching cutoff scores for model fit (Hu & Bentler, 1999). Additionally, measurement invariance of the observers was not confirmed, indicating that target persons and significant others interpreted the items on the emotion regulation scale differently. Two possible interpretations for this lack of model fit and measurement invariance seem plausible. First, emotion regulation might not be a core symptom of ADHD, which might explain the lack of model fit. Second, the newly developed scales for emotion regulation might not have been reliable and valid, a limitation which we will discuss in the next chapter. The self- and observer ratings of the SWAN-DE only including the subscales inattention and hyperactivity/impulsivity indicated a partial strict measurement invariance, where we had to relieve some restrictions to receive partial scalar (strong) invariance, but full configural, metric (weak) and residual (strict) invariance was confirmed. Accordingly, we can assume that target persons and significant others understood and filled in the questionnaire with the same underlying constructs in mind. This is a very important finding, underlining the excellent applicability of the SWAN-DE scales in self- and observer report research.

The final aim of the third manuscript was the attempt to explain differences between self- and observer reports with demographic (gender & age), relationship (partner, friend, family member, living together) and symptom strength variables. This knowledge is important, to understand whether the differences between self- and observer reports contain meaningful information instead of random measurement error. Referring to the Operations Triad Model (De Los Reyes et al., 2013), it helps us to distinguish between the need of diverging and compensating operations. Diverging operations interpret differences in reports as both containing valuable information while compensating operations are based on the notion that one of the reporters does not have sufficient context or insight to validly judge the self-regulation of the target person and therefore the information from the other reporter is supposed to compensate (Figure 1).

Concerning the demographic variables, an association with the difference between self- and observer report was found for the age of the target person. The older the target person was, the bigger was the difference between self- and observer report, with the self-report being increasingly more negative than the observer report. Previous studies have found that the difference in reports of internal problems increases with age (van der Ende et al., 2012). In our study we found this effect for externalizing symptoms of hyperactivity/impulsivity. Several explanations seem plausible for our finding, which should be investigated further in future research. First, the effect could be arbitrary in our study data. We recruited many more young than old adults in our sample, which might lead to a higher weight on the single individual in the older adults. That means, when only a few older adults have higher disagreement with their significant others on the SWAN-DE scales, this might already explain the effect we found. Future studies with more balanced age distributions should try to replicate this effect for a more reliable interpretation. However, if this effect proves to be robust, we could explain it with a higher internalization of the symptoms, a higher habituation to the symptoms from a significant other, and an interpretation of positive life-outcomes as evidence for low ADHD symptoms by the significant other (i.e., fundamental attribution error; Malle, 2006). Interestingly, self-esteem or a positive view on oneself, is reported to increase during adulthood, which would contradict our results (Robins et al., 2002). Comparably, the age of the observer was also negatively related to the magnitude in difference between self- and observer ratings. This indicates that the older the observer in our sample, the more the self- and observer reports agree with each other. We could imagine that this might be because the older the observer, the longer target person and observer know each other and therefore have had more communication with each other about self-regulation abilities or have seen each other in a larger amount of situations and in more varied contexts where self-regulation is required. Since we have not recorded relationship length, we are not able to test for this hypothesis and it might be an interesting starting point for future studies.

Furthermore, gender of the observer was negatively associated with the direction of the difference between self- and observer ratings, indicating that female observers, in contrast to male observers, assessed the target person more positively than the target person evaluated themselves. Since research has found gender differences in social desirability bias (Dalton & Ortegren, 2011), where women are more prone to appear positive in social situations, we hypothesize that this finding indicates that women evaluated the target persons more benevolent than men. Investigating this hypothesis in future research might be challenging, but first ideas would include having men and women assess the same target person with help of the SWAN-DE scales and comparing differences in ratings. However, this might still pose the question whether the target person actually behaves differently in the presence of male or female significant others. Another possible research question might be, whether same or different gender of target person and significant other (e.g., a woman evaluating another woman or a woman evaluating a man) lead to higher or lower accordance in ratings. We did not assess this research question because the relatively small sample size in our study in combination with a high number of variables in the multiple regression models already lowered the power to detect possible effects. However, it might be an interesting starting point to investigate in future studies.

To our surprise, no association was found for any of the relationship variables we investigated in our sample (i.e., type of dyad – partner, friend, or family member, and cohabitation) with the divergence between self- and observer report. We assumed that it made a difference in context and insight whether a partner, a friend, or a family member reported about ADHD symptoms. Also, we assumed that living together would have an impact on interrater agreement, with higher agreement when the dyad lived together. Nevertheless, dynamics within the investigated relationships might be very different. For example, some individuals might have closer relationships to their friends while others rely more on their family. Therefore, it might be important to inspect the relationship between target person and significant other more closely when investigating the correspondence of their reports. Length of the relationship, a measure of personal and non-personal (e.g., talking on the telephone, chatting via digital devices) contact, or companionship might be interesting variables to investigate in future studies.

The most robust association between reporter characteristic and difference in ratings we found in our study concerned the ADHD symptom strength of the target person. The difference between self- and observer reports was higher when observer ratings were more negative. Significant others might not be able to entirely perceive the intensity of ADHD symptoms of a target person, because this target person has learned to hide the symptoms, because the symptoms do not cause much impairment in the contexts when target person and significant other interact, or because the symptoms might simply be seen as a personality trait of the person that significant others habituated to (i.e., halo effect; Forgas & Laham, 2017). Significant others also might interpret the success of the target person in different life domains (e.g. work, health behavior, social relationships) as evidence for lower influence of observed ADHD symptoms (i.e., fundamental attribution error; Malle, 2006). Another possible explanation might be the social effect where humans do not want to say anything negative about another person, especially when it is a significant other they feel emotionally connected to (i.e., social desirability bias; Krumpal, 2013). Finally, individuals might indicate their symptoms as more extreme than they experience them, because they lack the necessary models for comparison, because they do not want to flatter themselves, because they feel they receive more help when they exaggerate their symptoms or because they established their self-image during childhood, when symptoms were more pronounced and they received negative feedback from their environment.

To sum up, demographic variables (i.e., age & gender) and strength of the ADHD symptoms of the target person seem to influence the difference between self- and observer ratings. Surprisingly, we did not find any evidence for an influence of the kind of relationship on this reporter difference. We cannot examine any of the speculations about the underlying reasons for the influence of demographic variables and symptom strength on observer agreement within the current research project. However, this seems like an important area to conduct further research in. Only if we know what information we exactly receive, when we ask for self- or observer reports, we can agree on guidelines for assessment of self-regulation and ADHD symptoms in research and in clinical practice as well as statistical implementations.

6. Limitations & Future Research

Within this dissertation I have investigated and discussed (new) approaches and challenges in the measurement of self-regulation, being aware that the title includes a very broad area of research. I hope to add new insights to the research community and inspire new ideas and hypotheses about measurement techniques. In the three empirical studies I presented, there is only limited capacity to examine the question about the best approach to measure self-regulation. Therefore, in the following section I will discuss some of the topics that remain open and might be investigated in future research.

The first challenge in the question on how to measure self-regulation is the inconsistent definition of self-regulation in the literature. Many authors use the term self-regulation interchangeably with constructs like self-control, executive functions, non-cognitive capacities, ADHD symptoms, and many more (Nigg, 2017). This leads to very diverse theories about associations and dependencies of the constructs within and between individuals. Therefore, it is difficult to formulate concrete hypotheses to test these theories empirically. Even in the three manuscripts of this thesis, we assembled different constructs to operationalize self-regulation. For example, we interpreted self-control as part of self-regulation in manuscript two but left the self-control scale out and focused only on the ADHD scale in manuscript one. In the presented studies, we operationalized ADHD and the core symptoms inattention and hyperactivity/impulsivity as indicators of a deficit in self-regulation. However, underlying self-regulation deficits are not unique in ADHD but are also fundamental symptoms in other psychological disorders (e.g., borderline personality disorder). Investigating the presented measurement techniques with regard to different psychological disorders might enhance our understanding of self-regulation in the general population and (different) clinical samples. Future examinations should therefore establish a common language, even better, an overarching theory that includes and discriminates self-regulation and its related constructs.

When the definition of the concept self-regulation is clear, it might be easier to implement specific questionnaires and scales to measure self-regulation ability in humans. In the three manuscripts, we used different scales to operationalize self-regulation: the Conners C3-AI (Lidzba et al., 2013), the short form of the Self-Control Scale (SCS-KD; Bertrams & Dickhäuser, 2009), and the SWAN-DE-SB (Blume et al., 2020). All of these scales have different advantages and pose individual challenges. The Conners C3-AI and the SCS-KD are valid and reliable measures which have been extensively researched in previous studies (Christiansen et al., 2012; Rauch et al., 2014). However, they are not originally constructed to be used for dimensional research and might not include the whole range of strengths and weaknesses in self-regulation of cognition and behavior in the population. Instead, they were constructed to identify children with inattention, hyperactivity/impulsivity and difficulties in self-control, which might explain why we found ceiling effects in the general population of school children in manuscripts one and two. Participants in the study might have had relatively high self-regulation abilities compared to clinical samples, where more difficulties in self-regulation might exist. Additionally, these scales were constructed for single measurements and originally interpreted self-regulation and self-control as relatively stable psychological traits. However, in the first two manuscripts we implemented them in ambulatory assessment, where we aimed at the examination of intraindividual fluctuations (i.e., states) as well as interindividual differences (i.e., traits). Therefore, another possible reason for the ceiling effects we found lies in the wording of the items which might not allow to indicate slight changes in self-regulation during daily life. Scales specifically constructed for the use in ambulatory assessment should allow for the depiction of smaller shifts in self-regulation during daily life. They might choose specific examples when self-regulation is required and make use of Likert scales with a sufficiently wide range. A promising attempt for this purpose might be the adaptation of the SWAN-DE scales for ambulatory assessment designs.

The SWAN-DE scales used in manuscript three were developed recently and therefore lack the necessary research evidence to (yet) be called reliable and valid. Initial studies have found good reliability and validity (e.g., Blume et al., 2020; Schulz-Zhecheva et al., 2017), however, these findings have to be replicated to confirm the evidence. Additionally, to our knowledge these scales have not yet been investigated on their discriminant validity to other constructs, like internalizing (e.g., depression, anxiety) or externalizing (e.g., oppositional defiant, conduct) disorders. With their dimensional focus, the SWAN-DE scales might be especially suitable to be used in ambulatory assessment, depicting fluctuations in strengths and weaknesses of self-regulation within individuals. However, this use has to my knowledge not been investigated yet. Taken together, more research is needed to develop scales which reliably and validly measure self-regulation as well in clinical as in general population samples, on the trait as well as on the state level and in all age groups.

A similar challenge in the third manuscript was the lack of a reliable and valid scale to measure emotion regulation dimensionally. To our knowledge, no dimensional scale of emotion regulation with neutrally formulated items had been developed before. Therefore, we re-formulated the deficit-oriented emotion regulation items from conventional ADHD scales (e.g., ADD; Brown, 2008; CAARS; Christiansen et al., 2013; WURS; Ward et al., 1993) into neutrally formulated items and integrated the 7-point Likert scale of the SWAN-DE ranging from *far below average* to *far above average*. This re-formulation, however, posed a challenge to us, since many of the conditions (e.g., throw tantrums, unpredictable mood, irritable) do not seem to have a clearly describable neutral counterpart and all items might respectively be described as *stay calm*. In a pilot study with 46 adult participants, we assessed the comprehension of the newly developed items and their correlations to the deficit-oriented items, however, no comprehensive analysis for validity and reliability was conducted. We would thus caution against overarching conclusions from our findings concerning the association of emotion regulation and ADHD symptoms, since the results might also be simply a consequence of inadequate items for the measurement of emotion regulation. Future research should try to develop adequate dimensional scales that show high validity and reliability and might then investigate factor structure and measurement invariance of self- and observer reports of dimensional measures of ADHD symptoms.

The research I have presented in the three manuscripts allows for interesting insights concerning the measurement of self-regulation in different samples. However, the studies all included relatively low sample sizes due to the lack of resources and highly challenging research designs (i.e., ambulatory assessment with measurement bursts). Expanding the research with higher sample sizes might help to enable more general conclusions about the measurement of self-regulation, for example by comparison of different age groups or developmental trajectories, concerning gender differences, socio-economic status, and cultural differences. Our samples were mainly composed of WEIRD (white, educated, industrialized, rich, democratic) participants, which restricts the universal conclusions we can make with our findings (Henrich et al., 2010). We would expect that our findings might be relatively universal in humans, but, since we established that self-regulation is highly influenced by internal and external factors, future research should confirm this universal validity.

Not only the number of participants could be enhanced in future studies but also the number of assessments in the first two manuscripts. When conducting ambulatory assessment studies, the statistical analyses always profit from a high number of data points. Some analyses are only accomplishable when enough data points are present. However, this remains in conflict with the high burden that ambulatory assessment studies impose on the participants. Getting pulled out of one's daily life activity several times during the day to answer the same questions over and over again might lead to a decrease in motivation and dismissal of the study protocol. Additionally, it has been suggested that the repeated measurement, and with it the enhanced attention on the target behavior, might influence the answer tendencies of participants through measurement reactivity (Barta et al., 2012). In the presented ambulatory assessment studies, we tried to weigh out the considerations of enough data points for profound analyses with those of participant burden and reactivity. However, it is possible that future studies with a higher amount of data points finds different temporal fluctuations or associations of self-regulation with additional constructs.

Next to the constructs sleep and working memory which we have examined, we can think of many other internal and external factors, which already have been or should be examined in their connection to self-regulation abilities in individuals. For example, other studies have investigated time in nature (Reuter et al., 2020), affect (Whalen et al., 2006), and parenting behaviors (Bridgett et al., 2018). Since self-regulation is often necessary in social contexts, it might also be interesting to investigate it more thoroughly with a focus on different relationships, like parent-child dyads, romantic partners, or teacher-student interactions. For example, one study has already investigated the association of relationship quality between parents and children and the self-regulation of the children (Moschko et al., 2023). Other studies have investigated self-regulated health behavior in dyads of romantic partners (e.g., Berli et al., 2021; Schwaninger et al., 2022; Szczuka et al., 2021).

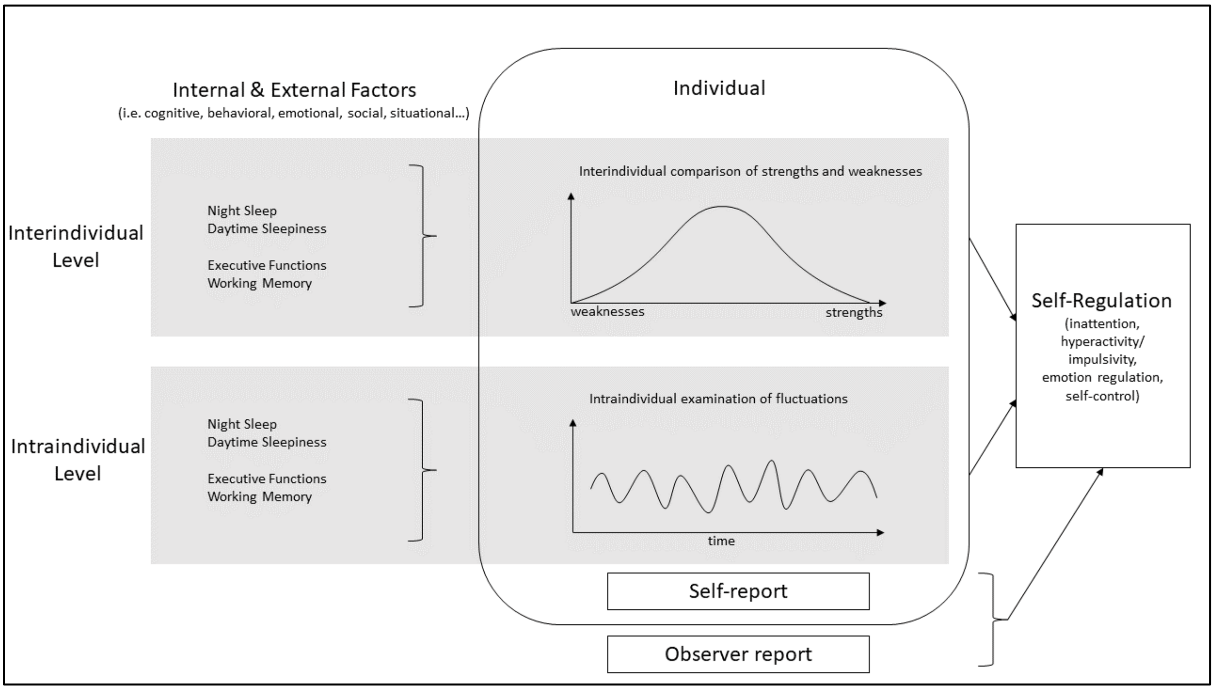
In our research project *Dyadic Investigation of Self-Regulation and Social Interaction in Intensive Longitudinal Research Design*, which is currently in preparation, we are trying to integrate many of these suggestions for future research. We plan to investigate self-regulation of both partners in a romantic dyad in their daily life via ambulatory assessment. Self-regulation will be measured three times per day for a period of 21 days using the dimensional questionnaire SWAN-DE-SB on a smartphone application. Additional variables which will be measured are life and relationship satisfaction. Although this research project is a promising first step towards a better understanding of self-regulation in humans, much more investigation is possible and necessary in the future until we can satisfactorily answer the question on how to best measure self-regulation.

7. Implications: How to Measure Self-Regulation

When we ask the question on how to measure self-regulation, there is definitely not one final answer (yet) to be formulated. In Figure 17, I present a model of possible considerations in the decision for a measurement approach. In manuscript one, self-regulation was measured on the inter- as well as on the intraindividual level. Assessment of internal factors like night sleep and daytime sleepiness were included into the measurements to inform about systematic associations between self-regulation and sleep on both levels. Manuscript two investigated the intraindividual fluctuations of self-regulation more thoroughly and examined the inter- and intraindividual associations with working-memory performance. Finally, manuscript three focused on the interindividual level and assessed strengths and weaknesses in inattention, hyperactivity/impulsivity, and emotion regulation dimensionally. Comparing self- and observer report in this manuscript allowed to investigate which additional information different perspectives on the self-regulation of a target person might convey.

Figure 17

Approaches in the measurement of self-regulation



As can already be seen by the different measurement approaches in the three manuscripts, the best way to measures self-regulation depends highly on the context of measurement, the reason of examination, resources in the assessment process, and on the sample which is investigated.

An important decision to make depending on the context and the resources is whether best to measure self-regulation and ADHD symptoms on the inter- or on the intraindividual level. Single measurements, which inform about interindividual differences, are much more economical than repeated measurements and might therefore be important first steps for the decision about, for example, the need for interventions or inclusion into empirical studies.

However, the investigation on the intraindividual level in a specific moment, best implemented trough daily measurements in real life of the individual with ambulatory assessment, provides important additional information about daily functioning and associated internal and external factors. In this dissertation, I have separated the discussions about the examination of self-regulation via ambulatory assessment and the examination of self-regulation with dimensional scales. However, combining both measurement methods – using dimensional scales in ambulatory assessment – might prevent ceiling effects in a general population and allows for more specific investigations of small changes in the self-regulation of cognition and behavior.

In clinical practice, measurements are usually used to diagnose a psychological disorder or to track treatment success. For the first objective, categorical scales might be used as most economical tools to inform a diagnosis. However, the second objective of tracking treatment success might require more fine-grained scales which also detect small changes and can depict strengths together with the weaknesses. Dimensional scales promise to be a considerable enhancement for this purpose. They are especially valuable since they allow for the assessment of strengths next to the weaknesses, which provides important information about resources in the treatment process. Empirical research contexts might also have individual goals which require different scales. For a group comparison of individuals with versus without ADHD, categorical scales provide the necessary cutoff scores. However, when investigating self-regulation within the general population, for example to examine the association with internal and external factors, dimensional scales contribute the necessary variance and the normal distribution that is needed for many statistical tests.

Including different reporters (i.e., self- and observer report) in the assessment process might provide important additional information on abilities of self-regulation. Since including several reporters often complicates the assessment process, it should be considered carefully, whether this additional information is necessary or whether one reporter brings enough insight and knowledge about different contexts so they can adequately inform about the target behaviors. Especially important in this consideration might be the age of the target person, since adolescents and adults are supposedly better equipped to inform about their self-regulation than children. But also symptom strength might be an indicator for the need of several reporters, implying that in a general population sample (where symptoms generally might be lower) there might be less need for several perspectives than in a clinical sample of individuals with strong ADHD symptoms.

Taken together, the best measurement of self-regulation might depend on the context and aim for which it is assessed, the resources which are available for the assessment and the demographical, internal, and external factors of the individuals which are assessed.

8. Conclusion

Self-regulation is an important capacity of humans and associated with many positive life outcomes. In my dissertation, I included three manuscripts investigating (new) approaches and challenges of measuring the self-regulation of individuals. Promising techniques are measurement in the daily life of participants via ambulatory assessment on technical devices like smartphones or tablets to investigate interindividual differences and intraindividual variance, investigating antecedents, correlates, and consequences of high or low self-regulation to better understand the underlying mechanisms of self-regulation inter- and intraindividually, the measurement with a dimensional focus on self-regulation within the general population to investigate strengths as well as weaknesses, and the inclusion of multiple observers to enhance the influence of context and insight. The insights gathered from the presented manuscripts open up new research hypotheses and stimulate future research to investigate each of the approaches in measuring self-regulation more thoroughly.

Literature

Abrams, J., Faraone, S. V., Woodworth, K. Y., Spencer, T. J., Biederman, I., & Biederman, J. (2018). Are Adult ADHD Patients Good Informants of Their Symptoms? A Qualitative Literature Review of Concordance between Clinician and Self-Report ADHD Symptoms. *Journal of Nervous and Mental Disease*, *206*(9), 739–743. https://doi.org/10.1097/NMD.0000000000000870

Alderson, R. M., Kasper, L. J., Hudec, K. L., & Patros, C. H. G. (2013). Attention-deficit/hyperactivity disorder (ADHD) and working memory in adults: A meta-analytic review. *Neuropsychology*, *27*(3), 287–302. https://doi.org/10.1037/a0032371

Alexander, L. A., & Liljequist, L. (2016). Determining the Accuracy of Self-Report Versus Informant-Report Using the Conners' Adult ADHD Rating Scale. *Journal of Attention Disorders*, *20*(4), 346–352. https://doi.org/10.1177/1087054713478652

Alexander, L. A., McKnight, P. E., Disabato, D. J., & Kashdan, T. B. (2017). When and How to use Multiple Informants to Improve Clinical Assessments. *Journal of Psychopathology and Behavioral Assessment*, *39*(4), 669–679. https://doi.org/10.1007/s10862-017-9607-9

American Psychiatric Association. (2013). DSM 5. In *American Journal of Psychiatry*. https://doi.org/10.1176/appi.books.9780890425596.744053

Arnett, A. B., Pennington, B. F., Friend, A., Willcutt, E. G., Byrne, B., Samuelsson, S., & Olson, R. K. (2013). The SWAN Captures Variance at the Negative and Positive Ends of the ADHD Symptom Dimension. *Journal of Attention Disorders*, *17*(2), 152–162. https://doi.org/10.1177/1087054711427399

Bailey, R., & Jones, S. M. (2019). An Integrated Model of Regulation for Applied Settings. *Clinical Child and Family Psychology Review*, *22*(1), 2–23. https://doi.org/10.1007/s10567-019-00288-y

Barkley, R. A. (1997). 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

Barkley, R. A. (2002). Major Life Activity and Health Outcomes Associated With Attention-Deficit/Hyperactivity Disorder. *Journal of Clinical Psychiatry*, *63*.

Barkley, R. A. (2011). Attention-Deficit/Hyperactivity Disorder, Self- Regulation, and Executive Functioning. In *Handbook of Self-Regulation - Research, Theory, and Applications* (Second, pp. 551–563). K.Vohs, R.Baumeister.

Barkley, R. A., Knouse, L. E., & Murphy, K. R. (2011). Correspondence and Disparity in the Self- and Other Ratings of Current and Childhood ADHD Symptoms and Impairment in Adults With ADHD. *Psychological Assessment*, *23*(2), 437–446. https://doi.org/10.1037/a0022172

Barkley, R. A., & Murphy, K. R. (2011). The Nature of Executive Function (EF) Deficits in Daily Life Activities in Adults with ADHD and Their Relationship to Performance on EF Tests. *Journal of Psychopathological Behavior Assessment*, 137–158. https://doi.org/10.1007/s10862-011-9217-x

Barta, W. D., Tennen, H., & Litt, M. D. (2012). Measurement Reactivity in Diary Research. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of Research Methods for Studying Daily Life* (pp. 108–123). The Guilford Press.

Bauer, I. M., & Baumeister, R. F. (2011). Self-Regulatory Strength. In K. D. Vohs & R. F. Baumeister (Eds.), *Handbook of Self-Regulation - Research, Theory, and Applications*. The Guilford Press.

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The Strength Model of Self-Control. *Current Directions in Psychological Science*, *16*(6), 351–355. https://doi.org/10.1111/j.1467-8721.2007.00534.x

Becker, D. R., McClelland, M. M., Loprinzi, P., & Trost, S. G. (2014). Physical Activity, Self-Regulation, and Early Academic Achievement in Preschool Children. *Early Education and Development*, *25*(1), 56–70. https://doi.org/10.1080/10409289.2013.780505

Becker, S. P. (2020). ADHD and sleep: recent advances and future directions. *Current Opinion in Psychology*, *34*(513), 50–56. https://doi.org/10.1016/j.copsyc.2019.09.006

Becker, S. P., Langberg, J. M., Eadeh, H., Isaacson, P. A., & Bourchtein, E. (2019). Sleep and daytime sleepiness in adolescents with and without ADHD: differences across ratings, daily diary, and actigraphy. *Journal of Child Psychology and Psychiatry*, jcpp.13061. https://doi.org/10.1111/jcpp.13061

Belendiuk, K. A., Clarke, T. L., Chronis, A. M., & Raggi, V. L. (2007). Assessing the concordance of measures used to diagnose adult ADHD. *Journal of Attention Disorders*, *10*(3), 276–287. https://doi.org/10.1177/1087054706289941

Berli, C., Schwaninger, P., & Scholz, U. (2021). “We Feel Good”: Daily Support Provision, Health Behavior, and Well-Being in Romantic Couples. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.622492

Bertrams, A., & Dickhäuser, O. (2009). Messung dispositioneller Selbstkontroll-Kapazität. *Diagnostica*, *55*(1), 2–10. https://doi.org/10.1026/0012-1924.55.1.2

Biederman, J., Mick, E., & Faraone, S. V. (2000). Age-Dependent Decline of Symptoms of Attention Deficit Hyperactivity Disorder : Impact of Remission Definition and Symptom Type. *American Journal of Psychiatry*, *157*, 816–818.

Bitto, H., Mörstedt, B., Faschina, S., & Stieglitz, R. D. (2017). ADHS bei Erwachsenen - Ein dimensionales oder kategoriales Konstrukt? *Zeitschrift Fur Psychiatrie, Psychologie Und Psychotherapie*, *65*(2), 121–131. https://doi.org/10.1016/B978-3-437-31585-5.01001-X

Blair, C., & Ursache, A. (2011). A bidirectional model of Executive Functions and Self-Regulation. In K. D. Vohs & R. F. Baumeister (Eds.), *Handbook of Self-Regulation - Research, Theory, and Applications* (pp. 300–320). The Guilford Press.

Blume, F., Buhr, L., Kuehnhausen, J., Köpke, R., Fallgatter, A. J., & Gawrilow, C. (2020). *Validation of a Self-Report Version of the German Strengths and Weaknesses of ADHD Symptoms and Normal Behavior Scale (SWAN-DE-SB). June 25.* https://doi.org/https://doi.org/10.31234/osf.io/x6bpd

Blume, F., Irmer, A., Dirk, J., & Schmiedek, F. (2022). Day-to-day variation in students' academic success: The role of self-regulation, working memory, and achievement goals. *Developmental Science*, *25*(6), 1–24. https://doi.org/https://doi.org/10.1111/desc.13301

Bodalski, E. A., Knouse, L. E., & Kovalev, D. (2019). Adult ADHD, Emotion Dysregulation, and Functional Outcomes: Examining the Role of Emotion Regulation Strategies. *Journal of Psychopathology and Behavioral Assessment*, *41*(1), 81–92. https://doi.org/10.1007/s10862-018-9695-1

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive Longitudinal Methods. An Introduction to Diary and Experience Sampling Research.* The Guilford Press.

Boonstra, A. M., Kooij, J. J. S., Oosterlaan, J., Sergeant, J. A., Buitelaar, J. K., & Van Someren, E. J. W. (2007). Hyperactive night and day? Actigraphy studies in adult ADHD: A baseline comparison and the effect of methylphenidate. *Sleep*, *30*(4), 433–442. https://doi.org/10.1093/sleep/30.4.433

Bridgett, D. J., Ganiban, J. M., Neiderhiser, J. M., Natsuaki, M. N., Shaw, D. S., Reiss, D., & Leve, L. D. (2018). Contributions of mothers' and fathers' parenting to children's self-regulation: Evidence from an adoption study. *Developmental Science*, *21*(6), 1–11. https://doi.org/10.1111/desc.12692

Brites, C., Salgado-Azoni, C. A., Ferreira, T. L., Lima, R. F., & Ciasca, S. M. (2015).

Development and applications of the SWAN rating scale for assessment of attention deficit hyperactivity disorder: a literature review. *Brazilian Journal of Medical and Biological Research*, *48*(11), 965–972. https://doi.org/10.1590/1414-431X20154528

Brose, A., Lövdén, M., & Schmiedek, F. (2014). Daily fluctuations in positive affect positively co-vary with working memory performance. *Emotion*, *14*(1), 1–6. https://doi.org/10.1037/a0035210

Brose, A., Schmiedek, F., Lövdén, M., & Lindenberger, U. (2012). Daily variability in working memory is coupled with negative affect: The role of attention and motivation. *Emotion*, *12*(3), 605–617. https://doi.org/10.1037/a0024436

Brown, T. E. (2006). Executive Functions and Attention Deficit Hyperactivity Disorder: Implications of two conflicting views. *International Journal of Disability, Development and Education*, *53*(1), 35–46. https://doi.org/10.1080/10349120500510024

Brown, T. E. (2008). ADD/ADHD and impaired executive function in clinical practice. *Current Psychiatry Reports*, *10*(5), 407–411. https://doi.org/10.1007/s11920-008-0065-7

Buhr, L., Moschko, T., Eppinger Ruiz de Zarate, A., Schwarz, U., Kühnhausen, J., & Gawrilow, C. (2022). The Association of Self-Reported ADHD Symptoms and Sleep in Daily Life of a General Population Sample of School Children: An Inter- and Intraindividual Perspective. *Brain Sciences*, *12*(4), 440. https://doi.org/10.3390/brainsci12040440

Campez, M., Raiker, J. S., Sarver, D. E., Friedman, L. M., Orban, S. A., & Rapport, M. D. (2020). Working Memory Capacity and ADHD Symptoms in Boys: Examining the Heterogeneity of Working Memory Functioning Using Latent Profile Analysis. *Journal of Psychopathology and Behavioral Assessment*, *42*(3), 450–463. https://doi.org/10.1007/s10862-019-09762-x

Carver, C. S., & Scheier, M. F. (2011). Self-Regulation of Action and Affect. In K. D. Vohs

& R. F. Baumeister (Eds.), *Handbook of Self-Regulation - Research, Theory, and Applications* (Second, pp. 3–20). The Guilford Press.

Caye, A., Swanson, J., Thapar, A., Sibley, M., Arseneault, L., Hechtman, L., Arnold, L. E., Niclasen, J., Moffitt, T., & Rohde, L. A. (2016). Life Span Studies of ADHD—Conceptual Challenges and Predictors of Persistence and Outcome. *Current Psychiatry Reports*, *18*(12). https://doi.org/10.1007/s11920-016-0750-x

Chan, G. F., Lai, K. Y., Luk, E. S., Hung, S.-F., & Leung, P. W.-L. (2014). Clinical utility of the Chinese Strengths and Weaknesses of ADHD-Symptoms and Normal-behaviors questionnaire (SWAN) when compared with DISC-IV. *Neuropsychiatric Disease and Treatment*, *10*, 1533–1542.

Chokroverty, S. (2017). Overview of Normal Sleep. In S. Chokroverty (Ed.), *Sleep Disorders Medicine*. Springer.

Christiansen, H., Hirsch, O., Albrecht, B., & Chavanon, M. L. (2019). Attention-Deficit/Hyperactivity Disorder (ADHD) and Emotion Regulation Over the Life Span. *Current Psychiatry Reports*, *21*(3), 16–18. https://doi.org/10.1007/s11920-019-1003-6

Christiansen, H., Hirsch, O., Philipsen, A., Oades, R. D., Matthies, S., Hebebrand, J., Ueckermann, J., Abdel-Hamid, M., Kraemer, M., Wiltfang, J., Graf, E., Colla, M., Sobanski, E., Alm, B., Rösler, M., Jacob, C., Jans, T., Huss, M., Schimmelmann, B. G., & Kis, B. (2013). German Validation of the Conners Adult ADHD Rating Scale-Self-Report: Confirmation of Factor Structure in a Large Sample of Participants With ADHD. *Journal of Attention Disorders*, *17*(8), 690–698. https://doi.org/10.1177/1087054711435680

Christiansen, H., Kis, B., Hirsch, O., Matthies, S., Hebebrand, J., Uekermann, J., Abdel-Hamid, M., Kraemer, M., Wiltfang, J., Graf, E., Colla, M., Sobanski, E., Alm, B., Rösler, M., Jacob, C., Jans, T., Huss, M., Schimmelmann, B. G., & Philipsen, A. (2012).

German validation of the Conners Adult ADHD Rating Scales (CAARS) II: Reliability, validity, diagnostic sensitivity and specificity. *European Psychiatry*, *27*(5), 321–328. https://doi.org/10.1016/j.eurpsy.2010.12.010

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*, *53*(5), 469–489. https://doi.org/10.1111/j.1469-7610.2011.02511.x

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Corbisiero, S., Mörstedt, B., Bitto, H., & Stieglitz, R.-D. (2017). Emotional Dysregulation in Adults With Attention-Deficit/Hyperactivity Disorder-Validity, Predictability, Severity, and Comorbidity. *Journal of Clinical Psychology*, *73*(1), 99–112. https://doi.org/10.1002/jclp.22317

Cronbach, L. J. (1960). *Essentials of Psychological Testing* (2nd ed.). Harper & Row.

Curran, P. J., & Bauer, D. J. (2011). Effects in longitudinal models of change. *Annual Review of Psychology*, *62*(1), 583–619. https://doi.org/10.1146/annurev.psych.093008.100356.

Dalton, D., & Ortegren, M. (2011). Gender Differences in Ethics Research: The Importance of Controlling for the Social Desirability Response Bias. *Journal of Business Ethics*, *103*(1), 73–93. https://doi.org/10.1007/s10551-011-0843-8

De Crescenzo, F., Licchelli, S., Ciabattini, M., Menghini, D., Armando, M., Alfieri, P., Mazzone, L., Pontrelli, G., Livadiotti, S., Foti, F., Quested, D., & Vicari, S. (2014). The use of actigraphy in the monitoring of sleep and activity in ADHD: A meta-analysis. *Sleep Medicine Reviews*, *26*, 9–20. https://doi.org/10.1016/j.smrv.2015.04.002

De Los Reyes, A., Thomas, S. A., Goodman, K. L., & Kundey, S. M. A. (2013). Principles Underlying the Use of Multiple Informants’ Reports. *Annual Review of Clinical*

Psychology, *9*(1), 123–149. https://doi.org/10.1146/annurev-clinpsy-050212-185617

Dewald, J. F., Meijer, A. M., Oort, F. J., Kerkhof, G. A., & Bögels, S. M. (2010). The influence of sleep quality, sleep duration and sleepiness on school performance in children and adolescents: A meta-analytic review. *Sleep Medicine Reviews*, *14*(3), 179–189. https://doi.org/10.1016/j.smrv.2009.10.004

Dirk, J., & Schmiedek, F. (2016). Fluctuations in elementary school children's working memory performance in the school context. *Journal of Educational Psychology*, *108*(5), 722–739. https://doi.org/10.1037/edu0000076

Dirk, J., & Schmiedek, F. (2017). Variability in Children's Working Memory Is Coupled With Perceived Disturbance: An Ambulatory Assessment Study in the School and Out-of-School Context. *Research in Human Development*, *14*(3), 200–218. https://doi.org/10.1080/15427609.2017.1340051

Döpfner, M., Hautmann, C., Görtz-Dorten, A., Klasen, F., & Ravens-Sieberer, U. (2015). Long-term course of ADHD symptoms from childhood to early adulthood in a community sample. *European Child & Adolescent Psychiatry*, *24*(6), 665–673. https://doi.org/10.1007/s00787-014-0634-8

Döpfner, M., Holtmann, M., Philipsen, A., Rothenberger, A., & Steinhausen, H.-C. (2020). Integrative ätiologische Modelle. In H.-C. Steinhausen, M. Döpfner, M. Holtmann, A. Philipsen, & A. Rothenberger (Eds.), *Handbuch ADHS: Grundlagen, Klinik, Therapie und Verlauf der Aufmerksamkeitsdefizit-Hyperaktivitätsstörung* (2nd ed., pp. 164–170). Kohlhammer.

Drechsler, R., Brem, S., Brandeis, D., Grünblatt, E., Berger, G., & Walitza, S. (2020). ADHD: Current Concepts and Treatments in Children and Adolescents. *Neuropediatrics*, *51*(05), 315–335. https://doi.org/10.1055/s-0040-1701658

Duckworth, A. L., Kirby, T. A., Gollwitzer, A., & Oettingen, G. (2013). From Fantasy to

Action: Mental contrasting with implementation intentions (MCII) improves academic performance in children. *Social Psychological and Personality Science*, *4*(6), 745–753. https://doi.org/10.1177/1948550613476307

Duckworth, A. L., & Tsukayama, E. (2015). Domain Specificity in Self-Control. In C. B. Miller, R. M. Furr, A. Knobel, & W. Fleeson (Eds.), *Character - New Directions from Philosophy, Psychology, and Theology* (pp. 393–412). Oxford University Press.

Ebner-Priemer, U. W., Kubiak, T., & Pawlik, K. (2009). Ambulatory Assessment. *European Psychologist*, *14*(2), 95–97. https://doi.org/10.1027/1016-9040.14.2.95

Ehm, J. H., Kerner Auch Koerner, J., Gawrilow, C., Hasselhorn, M., & Schmiedek, F. (2016). The association of ADHD symptoms and reading acquisition during elementary school years. *Developmental Psychology*, *52*(9), 1445–1456. https://doi.org/10.1037/dev0000186

Emser, T. S., Johnston, B. A., Steele, J. D., Kooij, S., Thorell, L., & Christiansen, H. (2018). Assessing ADHD symptoms in children and adults: Evaluating the role of objective measures. *Behavioral and Brain Functions*, *14*(1), 1–14. https://doi.org/10.1186/s12993-018-0143-x

Eppinger Ruiz de Zarate, A., & Gawrilow, C. (n.d.). *Physical Activity and Executive Functions in the daily life of schoolchildren*.

Erskine, H. E., Norman, R. E., Ferrari, A. J., Chan, G. C. K., Copeland, W. E., Whiteford, H. A., & Scott, J. G. (2016). Long-Term Outcomes of Attention-Deficit/Hyperactivity Disorder and Conduct Disorder: A Systematic Review and Meta-Analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*, *55*(10), 841–850. https://doi.org/10.1016/j.jaac.2016.06.016

Faraone, S. V. (2009). Using meta-analysis to compare the efficacy of medications for attention-deficit/hyperactivity disorder in youths. *Pharmacy and Therapeutics*, *34*(12), 678–684.

Faraone, S. V., & Buitelaar, J. (2010). Comparing the efficacy of stimulants for ADHD in children and adolescents using meta-analysis. *European Child and Adolescent Psychiatry*, *19*(4), 353–364. https://doi.org/10.1007/s00787-009-0054-3

Faraone, S. V, Biederman, J., & Mick, E. (2006). The age-dependent decline of attention deficit hyperactivity disorder : a meta-analysis of follow-up studies. *Psychological Medicine*, *36*, 159–165. https://doi.org/10.1017/S003329170500471X

Forgas, J. P., & Laham, S. M. (2017). Halo Effects. In R. F. Pohl (Ed.), *Cognitive Illusions: Intriguing Phenomena in Judgment, Thinking and Memory* (Second Edi, pp. 276–290). Routledge.

Frazier, T. W., Youngstrom, E. A., & Naugle, R. I. (2007). The Latent Structure of Attention-Deficit / Hyperactivity Disorder in a Clinic-Referred Sample. *Neuropsychology*, *21*(1), 45–64. https://doi.org/10.1037/0894-4105.21.1.45

Fricke-Oerkermann, L., Plück, J., Schredl, M., Heinz, K., Mitschke, A., Wiater, A., & Lehmkuhl, G. (2007). Prevalence and course of sleep problems in childhood. *Sleep*, *30*(10), 1371–1377. https://doi.org/10.1093/sleep/30.10.1371

Gaina, A., Sekine, M., Kanayama, H., Sengoku, K., Yamagami, T., & Kagamimori, S. (2005). Short-long sleep latency and associated factors in Japanese junior high school children. *Sleep and Biological Rhythms*, *3*, 162–165. https://doi.org/https://doi.org/10.1111/j.1479-8425.2005.00185.x

Galeano-Keiner, E. M., Neubauer, A. B., Irmer, A., & Schmiedek, F. (2022). Daily fluctuations in children’s working memory accuracy and precision: Variability at multiple time scales and links to daily sleep behavior and fluid intelligence. *Cognitive Development*, *64*(December), 101260. https://doi.org/10.1016/j.cogdev.2022.101260

Galeano Weber, E., Dirk, J., & Schmiedek, F. (2018). Variability in the Precision of

Children's Spatial Working Memory. *Journal of Intelligence*, *6*(1), 8. https://doi.org/10.3390/jintelligence6010008

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The Structure of Working Memory from 4 to 15 Years of Age. *Developmental Psychology*, *40*(2), 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Gawrilow, C., Kühnhausen, J., Schmid, J., & Stadler, G. (2014). Hyperactivity and Motoric Activity in ADHD: Characterization, Assessment, and Intervention. *Frontiers in Psychiatry*, *5*. https://doi.org/10.3389/fpsyt.2014.00171

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, *19*(1), 72–91. https://doi.org/10.1037/a0032138

Gooch, D., Maydew, H., Sears, C., & Norbury, C. F. (2017). Does a child's language ability affect the correspondence between parent and teacher ratings of ADHD symptoms? *BMC Psychiatry*, *17*(1), 129. https://doi.org/10.1186/s12888-017-1300-8

Gregory, A. M., Agnew-Blais, J. C., Matthews, T., Moffitt, T. E., & Arseneault, L. (2017). ADHD and Sleep Quality: Longitudinal Analyses From Childhood to Early Adulthood in a Twin Cohort. *Journal of Clinical Child and Adolescent Psychology*, *46*(2), 284–294. https://doi.org/10.1080/15374416.2016.1183499

Gregory, A. M., & Sadeh, A. (2012). Sleep, Emotional and Behavioural Difficulties in Children and Adolescents. *Sleep Medicine Reviews*, *16*(2), 129–136.

Gruber, R., Cassoff, J., Frenette, S., Wiebe, S., & Carrier, J. (2012). Impact of sleep extension and restriction on children's emotional lability and impulsivity. *Pediatrics*, *130*(5). https://doi.org/10.1542/peds.2012-0564

Hall, C. L., Valentine, A. Z., Groom, M. J., Walker, G. M., Sayal, K., Daley, D., & Hollis, C. (2016). The clinical utility of the continuous performance test and objective measures of activity for diagnosing and monitoring ADHD in children: a systematic review. *European Child and Adolescent Psychiatry*, *25*(7), 677–699. https://doi.org/10.1007/s00787-015-0798-x

Haslam, N., Williams, B., Prior, M., Haslam, R., & Graetz, B. (2006). The latent structure of attention-deficit / hyperactivity disorder : a taxometric analysis. *Australian and New Zealand Journal of Psychiatry*, *40*, 639–647.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Hilger, K., Sassenhagen, J., Kühnhausen, J., Reuter, M., Schwarz, U., Gawrilow, C., & Fiebach, C. J. (2020). Neurophysiological markers of ADHD symptoms in typically-developing children. *Scientific Reports*, *10*(1), 22460. https://doi.org/10.1038/s41598-020-80562-0

Hirsch, O., Chavanon, M. L., Riechmann, E., & Christiansen, H. (2018). Emotional dysregulation is a primary symptom in adult Attention-Deficit/Hyperactivity Disorder (ADHD). *Journal of Affective Disorders*, *232*, 41–47. https://doi.org/10.1016/j.jad.2018.02.007

Hiscock, H., Sciberras, E., Mensah, F., Gerner, B., Efron, D., Khano, S., & Oberklaid, F. (2015). Impact of a behavioural sleep intervention on symptoms and sleep in children with attention deficit hyperactivity disorder, and parental mental health: Randomised controlled trial. *BMJ (Online)*, *350*(January), 1–14. https://doi.org/10.1136/bmj.h68

Hofmann, W., Schmeichel, B. J., & Baddeley, A. D. (2012). Executive functions and self-regulation. *Trends in Cognitive Sciences*, *16*(3), 174–180. https://doi.org/10.1016/j.tics.2012.01.006

Hoorens, V. (1993). Self-enhancement and Superiority Biases in Social Comparison.

European Review of Social Psychology, *4*(1), 113–139. https://doi.org/10.1080/14792779343000040

Horn, R. (2009). *Raven's Standard Progressive Matrices*. Pearson.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hughes, D. M., Turkstra, L. S., & Wulfeck, B. B. (2009). Parent and self-ratings of executive function in adolescents with specific language impairment. *International Journal of Language & Communication Disorders*, *44*(6), 901–916. https://doi.org/10.1080/13682820802425693

Hunsley, J., & Mash, E. J. (2007). Evidence-based assessment. *Annual Review of Clinical Psychology*, *3*, 29–51. https://doi.org/10.1146/annurev.clinpsy.3.022806.091419

Hvolby, A. (2015). Associations of sleep disturbance with ADHD: implications for treatment. *ADHD Attention Deficit and Hyperactivity Disorders*, *7*(1), 1–18. https://doi.org/10.1007/s12402-014-0151-0

Ilkowska, M., & Engle, R. W. (2010a). Trait and State Differences in Working Memory Capacity. In A. Gruszka, G. Matthews, & B. Szymura (Eds.), *Handbook of Individual Differences in Cognition, Attention, Memory and Executive Control* (pp. 295–320). Springer. https://doi.org/10.1007/978-1-4419-1210-7

Ilkowska, M., & Engle, R. W. (2010b). Working Memory Capacity and Self-Regulation. In R. H. Hoyle (Ed.), *Handbook of Personality and Self-Regulation* (pp. 265–290). Wiley-Blackwell. https://doi.org/10.1002/9781444318111

Inzlicht, M., Werner, K. M., Briskin, J. L., & Roberts, B. W. (2021). Integrating Models of Self-Regulation. *Annual Review of Psychology*, *72*, 319–345. https://doi.org/10.1146/annurev-psych-061020-105721

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*, 1–11. https://doi.org/10.1186/1744-9081-5-15

Jungersen, C. M., & Lonigan, C. J. (2021). Do Parent and Teacher Ratings of ADHD Reflect the Same Constructs? A Measurement Invariance Analysis. *Journal of Psychopathology and Behavioral Assessment*, *43*, 778–792. https://doi.org/10.1007/s10862-021-09874-3

Kazak, A. E. (2018). Editorial: Journal article reporting standards. *American Psychologist, 73*(1), 1–2. https://doi.org/10.1037/amp0000263

Koch, E. D., Moukhtarian, T. R., Skirrow, C., Bozhilova, N., Asherson, P., & Ebner-Priemer, U. W. (2021). Using e-diaries to investigate ADHD – State-of-the-art and the promising feature of just-in-time-adaptive interventions. *Neuroscience and Biobehavioral Reviews*, *127*(June), 884–898. https://doi.org/10.1016/j.neubiorev.2021.06.002

Kofler, M. J., Alderson, R. M., Raiker, J. S., Bolden, J., Sarver, D. E., & Rapport, M. D. (2014). Working memory and intraindividual variability as neurocognitive indicators in ADHD: Examining competing model predictions. *Neuropsychology*, *28*(3), 459–471. https://doi.org/10.1037/neu0000050

Könen, T., Dirk, J., & Schmiedek, F. (2015). Cognitive benefits of last night’s sleep: Daily variations in children’s sleep behavior are related to working memory fluctuations. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *56*(2), 171–182. https://doi.org/10.1111/jcpp.12296

Kooij, J. J. S., Bijlenga, D., Salerno, L., Jaeschke, R., Bitter, I., Balázs, J., Thome, J., Dom, G., Kasper, S., Nunes Filipe, C., Stes, S., Mohr, P., Leppämäki, S., Casas, M., Bobes, J., Mccarthy, J. M., Richarte, V., Kjems Philipsen, A., Pehlivanidis, A., … Asherson, P. (2019). Updated European Consensus Statement on diagnosis and treatment of adult ADHD. *European Psychiatry*, *56*(1), 14–34.

https://doi.org/10.1016/j.eurpsy.2018.11.001

Kooij, J. J. S., Boonstra, A. M., Swinkels, S. H. N., Bekker, E. M., de Noord, I., & Buitelaar, J. K. (2008). Reliability, Validity, and Utility of Instruments for Self-Report and Informant Report Concerning Symptoms of ADHD in Adult Patients. *Journal of Attention Disorders*, *11*(4), 445–458. https://doi.org/10.1177/1087054707299367

Kostyrka-Allchorne, K., Wass, S. V., & Sonuga-Barke, E. J. S. (2020). Research Review: Do parent ratings of infant negative emotionality and self-regulation predict psychopathology in childhood and adolescence? A systematic review and meta-analysis of prospective longitudinal studies. *Journal of Child Psychology and Psychiatry*, *61*(4), 401–416. https://doi.org/10.1111/jcpp.13144

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity*, *47*(4), 2025–2047. https://doi.org/10.1007/s11135-011-9640-9

Kuriyan, A. B., Pelham, W. E., Molina, B. S. G., Waschbusch, D. A., Gnagy, E. M., Sibley, M. H., Babinski, D. E., Walther, C., Cheong, J., Yu, J., & Kent, K. M. (2013). Young adult educational and vocational outcomes of children diagnosed with ADHD. *Journal of Abnormal Child Psychology*, *41*(1), 27–41. https://doi.org/10.1007/s10802-012-9658-z

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*(4), 389–433. https://doi.org/10.1016/S0160-2896(05)80012-1

Lai, K. Y. C., Leung, P. W. L., Luk, E. S. L., Wong, A. S. Y., Law, L. S. C., & Ho, K. K. Y. (2011). Validation of the Chinese Strengths and Weaknesses of ADHD-Symptoms and Normal-Behaviors Questionnaire in Hong Kong. *Journal of Attention Disorders*, *17*(3), 194–202. https://doi.org/10.1177/1087054711430711

Lakes, K. D., Swanson, J. M., & Riggs, M. (2012). The Reliability and Validity of the English and Spanish Strengths and Weaknesses of ADHD and Normal Behavior Rating Scales in a Preschool Sample : Continuum Measures of Hyperactivity and Inattention. *Journal of Attention Disorders*, *16*(6), 510–516. https://doi.org/10.1177/1087054711413550

Larsson, H., Anckarsater, H., Råstam, M., Chang, Z., & Lichtenstein, P. (2012). Childhood attention-deficit hyperactivity disorder as an extreme of a continuous trait: A quantitative genetic study of 8,500 twin pairs. *Journal of Child Psychology and Psychiatry*, *53*(1), 73–80. https://doi.org/10.1111/j.1469-7610.2011.02467.x

Lechuga, M. T., Pelegrina, S., Pelaez, J. L., Martin-Puga, M. E., & Justicia, M. J. (2016). Working memory updating as a predictor of academic attainment. *Educational Psychology*, *36*(4), 675–690. https://doi.org/https://doi.org/10.1080/01443410.2014.950193

Leiner, D. J. (2019). *SoSciSurvey* (3.1.06). https://www.soscisurvey.de

Leonard, J. A., Lydon-Staley, D. M., Sharp, S. D. S., Liu, H. Z., Park, A. T., Bassett, D. S., Duckworth, A. L., & Mackey, A. P. (2021). Daily fluctuations in young children’s persistence. *Child Development*, 1–15. https://doi.org/10.1111/cdev.13717

Leonhardt, A., Könen, T., Dirk, J., & Schmiedek, F. (2016). How differentiated do children experience affect? An investigation of the within- and between-person structure of children’s affect. *Psychological Assessment*, *28*(5), 575–585. https://doi.org/10.1037/pas0000195

Lidzba, K., Christiansen, H., & Drechsler, R. (2013). *Conners Skalen zu Aufmerksamkeit und Verhalten – 3 (Conners 3®). Deutschsprachige Adaption der Conners 3rd Edition® (Conners 3®) von C. Keith Conners.* Verlag Hans Huber, Hogrefe AG.

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

Ludwig, K., Haindl, A., Laufs, R., Rauch, W. A., Ludwig, K., Haindl, A., Laufs, R., & Rauch, W. A. (2016). Self-Regulation in Preschool Children ' s Everyday Life : Exploring Day-to-Day Variability and the Within- and Between-Person Structure. *Journal of Self-Regulation and Regulation*, *02*, 98–117. https://doi.org/10.11588/josar.2016.2.34357

m-Path. (2023). KU Leuven Research & Development. https://m-path.io/landing/

Malle, B. F. (2006). The actor-observer asymmetry in attribution: A (surprising) meta-analysis. *Psychological Bulletin*, *132*(6), 895–919. https://doi.org/10.1037/0033-2909.132.6.895

Marcus, D. K., & Barry, T. D. (2011). Does Attention-Deficit/Hyperactivity Disorder Have a Dimensional Latent Structure? A Taxometric Analysis. *Journal of Abnormal Psychology*, *120*(2), 427–442. https://doi.org/10.1037/a0021405.Does

Marx, I., Hacker, T., Yu, X., Cortese, S., & Sonuga-Barke, E. (2021). ADHD and the Choice of Small Immediate Over Larger Delayed Rewards: A Comparative Meta-Analysis of Performance on Simple Choice-Delay and Temporal Discounting Paradigms. *Journal of Attention Disorders*, *25*(2), 171–187. https://doi.org/10.1177/1087054718772138

Matthews, J. S., Ponitz, C. C., & Morrison, F. J. (2009). Early Gender Differences in Self-Regulation and Academic Achievement. *Journal of Educational Psychology*, *101*(3), 689–704. https://doi.org/10.1037/a0014240

Mayes, S. D., Calhoun, S. L., Bixler, E. O., Vgontzas, A. N., Mahr, F., Hillwig-Garcia, J., Elamir, B., Edhere-Ekezie, L., & Parvin, M. (2008). ADHD Subtypes and Comorbid Anxiety, Depression, and Oppositional-Defiant Disorder: Differences in Sleep Problems. *Journal of Pediatric Psychology*, *34*(3), 328–337.

https://doi.org/10.1093/jpepsy/jsn083

McClelland, M. M., Acock, A. C., Piccinin, A., Rhea, S. A., & Stallings, M. C. (2013). Relations between preschool attention span-persistence and age 25 educational outcomes. *Early Childhood Research Quarterly*, *28*(2), 314–324. https://doi.org/10.1016/j.ecresq.2012.07.008

McKinney, T. L., Euler, M. J., & Butner, J. E. (2020). It's about time: The role of temporal variability in improving assessment of executive functioning. *Clinical Neuropsychologist*, *34*(4), 619–642. https://doi.org/10.1080/13854046.2019.1704434

Mehl, M. R., & Conner, T. S. (2012). *Handbook of Research Methods for Studying Daily Life*.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14. https://doi.org/10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex "Frontal Lobe" Tasks: A Latent Variable Analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Molenaar, P. C. M., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science*, *18*(2), 112–117. https://doi.org/10.1111/j.1467-8721.2009.01619.x

Morrison, F. J., Kim, M. H., Connor, C. M., & Grammer, J. K. (2019). The Causal Impact of Schooling on Children's Development: Lessons for Developmental Science. *Current Directions in Psychological Science*, *28*(5), 441–449. https://doi.org/10.1177/0963721419855661

Moschko, T., Stadler, G., & Gawrilow, C. (2023). Fluctuations in children's self-regulation and parent-child interaction in everyday life: An ambulatory assessment study. *Journal of Social and Personal Relationships*, *40*(1), 254–276. https://doi.org/10.1177/02654075221116788

Muraven, M., Tice, D. M., & Baumeister, R. F. (1998). Self-Control as Limited Resource: Regulatory Depletion Patterns. *Journal of Personality and Social Psychology*, *74*(3), 774–789. https://doi.org/10.1037/0022-3514.74.3.774

Murphy, P., & Schachar, R. (2000). Use of self-ratings in the assessment of symptoms of attention deficit hyperactivity disorder in adults. *American Journal of Psychiatry*, *157*(7), 1156–1159. https://doi.org/10.1176/appi.ajp.157.7.1156

Muthén, L., & Muthén, B. (2021). *Mplus* (8.6).

Narad, M. E., Garner, A. A., Peugh, J. L., Tamm, L., Antonini, T. N., Kingery, K. M., Simon, J. O., & Epstein, J. N. (2015). Parent–teacher agreement on ADHD symptoms across development. *Psychological Assessment*, *27*(1), 239–248. https://doi.org/10.1037/a0037864

Nesselroade, J. R. (1991). The Warp and Woof of the Developmental Fabric. In R. M. Downs, L. S. Liben, & D. S. Palermo (Eds.), *Visions of Aesthetics, The Environment & Development: The Legacy of Joachim F. Wohlwill* (pp. 213–240). Lawrence Erlbaum Associates, Inc.

Neubeck, M., Johann, V. E., Karbach, J., & Könen, T. (2022). Age-differences in network models of self-regulation and executive control functions. *Developmental Science*, *e13276*, 1–13. https://doi.org/10.1111/desc.13276

Nigg, J. T. (2017). Annual Research Review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *58*(4), 361–383.

https://doi.org/10.1111/jcpp.12675

Nigg, J. T., Sibley, M. H., Thapar, A., & Karalunas, S. L. (2020). Development of ADHD: Etiology, Heterogeneity, and Early Life Course. *Annual Review of Developmental Psychology*, *2*(1), 559–583. https://doi.org/10.1146/annurev-devpsych-060320-093413

Owens, J., Gruber, R., Brown, T., Corkum, P., Cortese, S., O’Brien, L., Stein, M., & Weiss, M. (2012). Future Research Directions in Sleep and ADHD: Report of a Consensus Working Group. *Journal of Attention Disorders*, *17*(7), 550–564. https://doi.org/10.1177/1087054712457992

Paavonen, E. J., Räikkönen, K., Lahti, J., Komsi, N., Heinonen, K., Pesonen, A. K., Järvenpää, A. L., Strandberg, T., Kajantie, E., & Porkka-Heiskanen, T. (2009). Short sleep duration and behavioral symptoms of attention-deficit/ hyperactivity disorder in healthy 7- to 8-year-old children. *Pediatrics*, *123*(5), 857–864. https://doi.org/10.1542/peds.2008-2164

Pesonen, A. K., Räikkönen, K., Paavonen, E. J., Heinonen, K., Komsi, N., Lahti, J., Kajantie, E., Järvenpää, A. L., & Strandberg, T. (2010). Sleep duration and regularity are associated with behavioral problems in 8-year-old children. *International Journal of Behavioral Medicine*, *17*(4), 298–305. https://doi.org/10.1007/s12529-009-9065-1

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*, *56*(3), 345–365. https://doi.org/10.1111/jcpp.12381

Polanczyk, G. V., Willcutt, E. G., Salum, G. A., Kieling, C., & Rohde, L. A. (2014). ADHD prevalence estimates across three decades: An updated systematic review and meta-regression analysis. *International Journal of Epidemiology*, *43*(2), 434–442. https://doi.org/10.1093/ije/dyt261

Polderman, T. J. C., Derks, E. M., Hudziak, J. J., Verhulst, F. C., Posthuma, D., & Boomsma, D. I. (2007). Across the continuum of attention skills: A twin study of the SWAN ADHD rating scale. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *48*(11), 1080–1087. https://doi.org/10.1111/j.1469-7610.2007.01783.x

Purdie, N., Hattie, J., & Carroll, A. (2002). A review of the research on interventions for attention deficit hyperactivity disorder: What works best? *Review of Educational Research*, *72*(1), 61–99. https://doi.org/10.3102/00346543072001061

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

R Core Team. (2020). *R: A language and environment for statistical computing* (4.1.1.). R Foundation for Statistical Computing. https://www.r-project.org/

Raffaelli, M., Crockett, L. J., & Shen, Y. L. (2005). Developmental stability and change in self-regulation from childhood to adolescence. *Journal of Genetic Psychology*, *166*(1), 54–76. https://doi.org/10.3200/GNTP.166.1.54-76

Rauch, W. A., Gawrilow, C., Schermelleh-Engel, K., & Schmitt, K. (2014). Dispositionelle Selbstkontrollkapazität bei Kindern. *Diagnostica*, *60*(2), 61–72. https://doi.org/10.1026/0012-1924/a000099

Retz, W., Stieglitz, R.-D., Corbisiero, S., Retz-Junginger, P., & Rösler, M. (2012). Emotional dysregulation in adult ADHD: what is the empirical evidence? *Expert Review of Neurotherapeutics*, *12*(10), 1241–1251. https://doi.org/10.1586/ern.12.109

Reuter, M., Kühnhausen, J., Haas, P., & Gawrilow, C. (2020). Within-person relationship between daily time spent in natural spaces and affect as well as inattention in children's daily life. *Umweltpsychologie*, *24*(2), 176–199.

Reuter, M., Kühnhausen, J., Kleinhansl, M., Henn, L., & Gawrilow, C. (2021). Children's daily time in natural spaces during spring/summer time: Within-person restoration and the role of environmental attitude. *Umweltpsychologie*, *25*(2), 169–189. https://umps.de/php/artikeldetails.php?id=780

Riediger, M., Wrzus, C., Schmiedek, F., Wagner, G. G., & Lindenberger, U. (2011). Is Seeking Bad Mood Cognitively Demanding? Contra-Hedonic Orientation and Working-Memory Capacity in Everyday Life. *Emotion*, *11*(3), 656–665. https://doi.org/10.1037/a0022756

Riley, A. W. (2004). Evidence That School-Age Children Can Self-Report on Their Health. *Ambulatory Pediatrics*, *4*(4), 371–376. https://doi.org/10.1367/A03-178R.1

Robaey, P., Amre, D., Schachar, R., & Simard, L. (2007). French Version of the Strengths and Weaknesses of ADHD Symptoms and Normal Behaviors (SWAN-F) Questionnaire. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, *16*(2), 80–89.

Robins, R. W., Trzesniewski, K. H., Tracy, J. L., Gosling, S. D., & Potter, J. (2002). Global self-esteem across the life span. *Psychology and Aging*, *17*(3), 423–434. https://doi.org/10.1037/0882-7974.17.3.423

Robson, D. A., Allen, M. S., & Howard, S. J. (2020). Self-Regulation in Childhood as a Predictor of Future Outcomes: A Meta-Analytic Review. *Psychological Bulletin*. https://doi.org/10.1037/bul0000227

Rösler, M., Retz-Junginger, P., Retz, W., & Stieglitz, R. D. (2008). *HASE Homburger ADHS-Skalen für Erwachsene*. Hogrefe.

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

Schmidt, H., Daseking, M., Gawrilow, C., Karbach, J., & Kerner auch Koerner, J. (2022).

Self-regulation in preschool: Are executive function and effortful control overlapping constructs? *Developmental Science, July 2021*, 1–13. https://doi.org/10.1111/desc.13272

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2013). Keeping It Steady: Older Adults Perform More Consistently on Cognitive Tasks Than Younger Adults. *Psychological Science*, *24*(9), 1747–1754. https://doi.org/10.1177/0956797613479611

Schulz-Zhecheva, Y., Voelkle, M., Beauducel, A., Buch, N., Fleischhaker, C., Bender, S., Saville, C. W. N., Biscaldi, M., & Klein, C. (2017). ADHD Traits in German School-Aged Children: Validation of the German Strengths and Weaknesses of ADHS Symptoms and Normal Behavior (SWAN-DE) Scale. *Journal of Attention Disorders*, *1–10*. https://doi.org/10.1177/1087054716676365

Schumacher, A. M., Miller, A. L., Watamura, S. E., Kurth, S., Lassonde, J. M., & LeBourgeois, M. K. (2017). Sleep Moderates the Association Between Response Inhibition and Self-Regulation in Early Childhood. *Journal of Clinical Child and Adolescent Psychology*, *46*(2), 222–235. https://doi.org/10.1080/15374416.2016.1204921

Schwaninger, P., Lüscher, J., Berli, C., & Scholz, U. (2022). Daily support seeking as coping strategy in dual-smoker couples attempting to quit. *Psychology & Health*, *37*(7), 811–827. https://doi.org/10.1080/08870446.2021.1913157

Schwarz, U., & Gawrilow, C. (2019). Measuring and compensating for deficits of self-regulation in school children via ambulatory assessment. *Psychology in Russia: State of the Art*, *12*(4), 8–22. https://doi.org/10.11621/pir.2019.0401

Sciberras, E., Roos, L. E., & Efron, D. (2009). Review of prospective longitudinal studies of children with ADHD: Mental health, educational, and social outcomes. *Current Attention Disorders Reports*, *1*(4), 171–177. https://doi.org/10.1007/s12618-009-0024-1

Shaw, P., Stringaris, A., Nigg, J., & Leibenluft, E. (2014). Emotion Dysregulation in

Attention Deficit Hyperactivity Disorder. *American Journal of Psychiatry*, *171*(3), 276–293. https://doi.org/10.1176/appi.ajp.2013.13070966

Shrout, P. E., & Lane, S. P. (2012). Psychometrics. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of Research Methods for Studying Daily Life* (pp. 302–320). The Guilford Press.

Shrout, P. E., Stadler, G., Lane, S. P., McClure, M. J., Jackson, G. L., Clavél, F. D., Iida, M., Gleason, M. E. J., Xu, J. H., & Bolger, N. (2018). Initial elevation bias in subjective reports. *Proceedings of the National Academy of Sciences*, *115*(1), E15–E23. https://doi.org/10.1073/pnas.1712277115

Sliwinski, M. J., Mogle, J. A., Hyun, J., Munoz, E., Smyth, J. M., & Lipton, R. B. (2018). Reliability and Validity of Ambulatory Cognitive Assessments. *Assessment*, *25*(1), 14–30. https://doi.org/10.1177/1073191116643164

Smith, B. H., Pelham, W. E., Gnagy, E., Molina, B., & Evans, S. (2000). The reliability, validity, and unique contributions of self-report by adolescents receiving treatment for attention-deficit/hyperactivity disorder. *Journal of Consulting and Clinical Psychology*, *68*(3), 489–499. https://doi.org/10.1037/0022-006X.68.3.489

Smithers, L. G., Sawyer, A. C. P., Chittleborough, C. R., Davies, N., Smith, G. D., & Lynch, J. (2018). A systematic review and meta-analysis of effects of early life non-cognitive skills on academic, psychosocial, cognitive and health outcomes. *Nat Hum Behav*, *2*(11), 867–880. https://doi.org/10.1038/s41562-018-0461-x.A

Sonuga-Barke, E. J. S. (2002). Psychological heterogeneity in AD/HD — a dual pathway model of behaviour and cognition. *Behavioural Brain Research*, *130*, 29–36.

Steyer, R., Schwenkmezger, P., Notz, P., & Eid, M. (1997). *Der Mehrdimensionale Befindlichkeitsfragebogen (MDBF)*. Hogrefe.

Swanson, & Alloway, T. P. (2012). Working memory, learning, and academic achievement.

In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook, Vol 1: Theories, constructs, and critical issues.* (Vol. 1, pp. 327–366). American Psychological Association. https://doi.org/10.1037/13273-012

Swanson, J. M., Schuck, S., Porter, M. M., Hartman, C. A., Sergeant, J. A., Clevenger, W., Wasdell, M., McCleary, R., Lakes, K., & Wigal, T. (2012). Categorical and Dimensional Definitions and Evaluations of Symptoms of ADHD: History of the SNAP and the SWAN Rating Scales. *The International Journal of Educational and Psychological Assessment*, *10*(1), 51–70.

Szczuka, Z., Kulis, E., Boberska, M., Banik, A., Kruk, M., Keller, J., Knoll, N., Scholz, U., Abraham, C., & Luszczynska, A. (2021). Can individual, dyadic, or collaborative planning reduce sedentary behavior? A randomized controlled trial. *Social Science & Medicine*, *287*, 114336. https://doi.org/10.1016/j.socscimed.2021.114336

Trull, T. J., & Ebner-Priemer, U. W. (2019). Ambulatory Assessment in Psychopathology Research: A Review of Recommended Reporting Guidelines and Current Practices. *Journal of Abnormal Psychology*.

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin*, *134*(3), 383–403. https://doi.org/10.1037/0033-2909.134.3.383

van der Ende, J., Verhulst, F. C., & Tiemeier, H. (2012). Agreement of informants on emotional and behavioral problems from childhood to adulthood. *Psychological Assessment*, *24*(2), 293–300. https://doi.org/10.1037/a0025500

Van Der Heijden, K. B., Smits, M. G., & Gunning, W. B. (2005). Sleep-related disorders in ADHD: A review. *Clinical Pediatrics*, *44*(3), 201–210. https://doi.org/10.1177/000992280504400303

Van der Meere, J. (2005). State Regulation and Attention Deficit Hyperactivity Disorder. In

D. Gozal & D. L. Molfese (Eds.), *Attention Deficit Hyperactivity Disorder. From Genes to Patients* (pp. 413–434). Humana Press.

Van Voorhees, E. E., Hardy, K. K., & Kollins, S. H. (2011). Reliability and Validity of Self- and Other-Ratings of Symptoms of ADHD in Adults. *Journal of Attention Disorders*, *15*(3), 224–234. https://doi.org/10.1177/1087054709356163

Vriend, J. L., Davidson, F. D., Corkum, P. V., Rusak, B., McLaughlin, E. N., & Chambers, C. T. (2012). Sleep Quantity and Quality in Relation to Daytime Functioning in Children. *Children's Health Care*, *41*(3), 204–222. https://doi.org/10.1080/02739615.2012.685039

Wang, L. P., & Maxwell, S. E. (2015). On Disaggregating Between-Person and Within-Person Effects With Longitudinal Data Using Multilevel Models. *Psychological Methods*, *20*(1), 63–83. https://doi.org/http://dx.doi.org/10.1037/met0000030

Ward, M. F., Wender, P. H., & Reimherr, F. W. (1993). The Wender Utah Rating Scale: An aid in the retrospective diagnosis of childhood attention deficit hyperactivity disorder. *American Journal of Psychiatry*, *150*(6), 885–890. http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L23156767%5Cnhttp://sfx.library.uu.nl/utrecht?sid=EMBASE&issn=0002953X&id=doi:&atitle=The+Wender+Utah+Rating+Scale%3A+An+aid+in+the+retrospective+diagnosis+of+childhood+attention+def

Whalen, C. K., Henker, B., Ishikawa, S. S., Jamner, L. D., Floro, J. N., Johnston, J. A., & Swindle, R. (2006). An Electronic Diary Study of Contextual Triggers and ADHD: Get Ready, Get Set, Get Mad. *Journal of the American Academy of Child and Adolescent Psychiatry*, *45*(2), 166–174. https://doi.org/10.1097/01.chi.0000189057.67902.10

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

World Health Organization [WHO]. (2018). *International classification of diseases for mortality and morbidity statistics (11th Revision)*. https://icd.who.int/browse11/l-m/en

Young, S., Bramham, J., Gray, K., & Rose, E. (2008). The Experience of Receiving a Diagnosis and Treatment of ADHD in Adulthood. A Qualitative Study of Clinically Referred Patients Using Interpretative Phenomenological Analysis. *Journal of Attention*, *11*(4), 493–503.

Zucker, M., Morris, M. K., Ingram, S. M., Morris, R. D., & Bakeman, R. (2002). Concordance of self- and informant ratings of adults' current and childhood attention-deficit/hyperactivity disorder symptoms. *Psychological Assessment*, *14*(4), 379–389. https://doi.org/10.1037/1040-3590.14.4.379