



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Self-Other Concordance of Children and Adolescents
with their Caregivers in Foster Care on Internalizing and
Externalizing Behavior Problems“

verfasst von / submitted by

Jakob Florin Kreutter BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

Wien, 2020

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

UA 066 840

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Masterstudium Psychologie UG2002

Betreut von / Supervisor:

Univ.-Prof. Dr. Brigitte Lueger-Schuster

Table of Contents

Theoretical Background.....	4
Schema-project of the University of Vienna.....	5
Externalizing and Internalizing Behavior Problems.....	5
Externalizing Behavior Problems.....	5
Internalizing Behavior Problems.....	6
The Role of Sex and Age in Externalizing and Internalizing Behavior Outcomes.....	6
Gender Stereotypes and Biases.....	7
Cross-Informant Concordance.....	8
Research Questions and Hypotheses.....	10
Methods.....	12
Study Design.....	12
Participants.....	13
Assessment Instrument.....	14
Child Behavior Checklist (CBCL).....	14
Statistical Analysis.....	16
Missing Data Analysis.....	16
Analysis of outcomes within children.....	16
Analysis of outcomes between children.....	18
Analysis of outcomes between children and caregivers.....	18
Results.....	18
Social Structure.....	18
Prevalence of Internalizing and Externalizing Behavior Problems.....	21
Outcomes Within Children.....	21
Outcomes Between Children.....	22
Comparison of Self and Other Judgement.....	22
Age Differences in Cross-Informant Correspondence.....	22
Gender Differences in Cross-Informant Correspondence.....	25
Discussion.....	25
Remarks to the Social Structure.....	26
Summary of the Results.....	26
Elaboration on the Results.....	28

Prevalence.....	28
Age Differences.....	29
Age Differences in Cross-Informant Correspondence.....	32
Internalizing Problems.....	32
Externalizing Problems.....	33
Total Problems.....	34
Total Sample: Cross-Informant Correspondence.....	35
Gender Differences in Cross-Informant Correspondence.....	37
Limitations.....	40
Limitations Due to the Usage of Diagnostic Instruments.....	40
Limitations Due to the Nature of the Sample.....	41
Implications and Outlook.....	42
Implications for Future Research.....	42
Clinical Implications.....	43
Conclusion.....	44
References.....	45
List of Tables.....	58
List of Abbreviations.....	59
Appendix.....	60
Abstract (English).....	60
Abstract (German).....	61

Self-Other Concordance of Children and Adolescents with their Caregivers in Foster Care on Internalizing and Externalizing Behavior Problems

Externalizing and internalizing behavior problems are thought to be the most prominent mental health issues concerning children and adolescents. They often occur in children and adolescents with a traumatic history (Evans et al., 2008; Schmid et al., 2008). From here on the term *children* will include *children and adolescents* until adolescents alone are explicitly mentioned. Children in institutionalized care are a high-risk population for traumatic histories (Clausen et al., 1998; McAuley & Trew, 2000). Placing Children in foster care, means taking them out of their family systems, which is already a traumatic experience, but it is still often the safest method to leave abusive homes (McWey et al., 2010). Children in foster care are also more likely to have been exposed to more than one traumatic situation in their past and are thus more imperiled (McCann et al., 1996; Niemann & Häbeler, 2014; Oswald et al., 2010). This is also highlighted through the fact that children in foster care show 2.5–3 times more behavioral problems than children in the general population (Clausen et al., 1998).

The research project *Schematheoretische Weiterbildung und ihre Effekte auf in niederösterreichischen Landeseinrichtungen untergebrachte Kinder und Jugendliche* (Schema-Theoretical Training and its Effects on Foster Children and Adolescents in Lower-Austrian Foster Care Institutions) under guidance of Prof. Lueger-Schuster looks at children in institutionalized care in parts of Austria. This Master thesis takes place within this project but focuses on one screening instrument, namely the Child Behavior Checklist (CBCL) (Achenbach, 1991; German Version: Döpfner et al., 2014). This instrument is mainly used to screen for internalizing and externalizing behavior problems and also explore the resources of the child.

Children tend to show different posttraumatic symptoms than adults (Scheeringa et al., 2006). For example, ADHD-like symptoms, which are considered externalizing, are often caused by traumata (Luoni et al., 2018). Muniz et al. (2019) found that physical, sexual and emotional abuse, neglect, exposure to household violence and/or substance abuse, parental mental illness and parental incarceration all significantly predict either internalizing or externalizing, or both, outcomes in children. More knowledge about the prevalence and origins of said problems could help health care professionals, but also teachers and pedagogues recognize traumatic experiences in children earlier. Hence, these behavior

problems are important to investigate since they can be indicators of children suffering from past traumata.

This work will compare and analyze the cross-informant concordance of foster children and their caregivers in relation to externalizing and internalizing outcomes. Greater disagreement between children and others concerning children's behaviors was shown to lead to worsened therapeutic success and can even result in children becoming therapeutic non-responders (Humphreys et al., 2017). It also increases the likelihood for developing further psychopathologies (Humphreys et al., 2017, Romano et al., 2018). It is therefore important to analyze the concordance of judgement between foster children and their caregivers in order to improve assessment, care and treatment for this high-risk population.

Schema-project of the University of Vienna

The research project aims to evaluate the possible effects of a schema-theoretical training for the pedagogues, e.g. caregivers of the institutionalized children. As part of the study the pedagogues and the children have to fill out various screening instruments concerning the children. The children in foster placements in Niederösterreich get screened for various psychopathologies and other psychic properties. Emotion-regulation strategies, depressive and anxious symptoms as well as behavioral problems are of interest, but the main focus lies on the experience of traumatic events and the potential development of a post-traumatic stress disorder.

Externalizing and Internalizing Behavior Problems

Externalizing Behavior Problems

Achenbach (1966) was the first to conceptualize a framework for externalizing and internalizing behavior problems. Externalizing symptoms refer to aggressive, violent and delinquent behaviors, conduct problems, hyperactivity and problems with self-regulation (Achenbach, 1991; Bornstein et al., 2010; Campbell, 1995; Muniz et al., 2019). The younger children are the likelier it is that they execute aggressive and destructive behaviors as cognitive abilities and social skills are not yet developed properly (Fanti & Henrich, 2010). This means that externalizing behavior problems diminish over time in the general population and only persist in some children. Risk factors for these problems are low socioeconomic status, maternal depression, parental drug abuse, violence and other childhood traumata (Gilliom & Shaw, 2004; Moylan et al., 2010). Externalizing behaviors are negatively related to social competence, including conflicts with peers and parents (Bornstein et al., 2010). They lead to peer victimization and rejection and subsequently to more contact with deviant peers which helps to maintain the problems (Laird et al., 2001). These problems, especially

hyperactivity and inattention, were linked to poor academic performance (Shaw et al., 2012). Externalizing behaviors tended to entail substance abuse and other psychopathologies, including conduct disorder, antisocial personality disorder (Bornstein et al., 2010; Fanti & Henrich, 2010; Muniz et al., 2019) and mood and anxiety problems (Zaharakis et al., 2018). Aggressive and violent behaviors, as well as conduct problems were related to criminal behavior and offending (Zaharakis et al., 2018).

Internalizing Behavior Problems

Internalizing behaviors describe inward directed behaviors. These include depressive symptoms, anxiety, withdrawal, as well as somatic symptoms (Achenbach et al., 1991; Bornstein et al., 2010; Fanti & Henrich, 2010). In contrast to externalizing behaviors, internalizing problems tend to increase from infancy to adolescence (Achenbach et al., 1991). This might be due to cognitive development (Kovacs & Devlin, 1998) that leads to more self-reflection and more thorough analysis of negative happenings. Internalizing problems commonly prevent children to form friendships and good relations with others as isolation and withdrawal stand diametral to social contact (Oland & Shaw, 2005). Therefore, social competence is also not well developed in children with internalizing behaviors, since social competence, as Bornstein et al. (2010) define it, consists of self-regulatory and positive communication skills, social cognition and prosocial relationships with others.

The Role of Sex and Age in Externalizing and Internalizing Outcomes

Many children exhibit externalizing and internalizing outcomes simultaneously (Chen, 2010), especially when childhood risk factors are high (Fanti & Henrich, 2010), which is most often the case with children in foster care (Oswald et al., 2010). Research done in foster care institutions found nearly 37% of children had clinical scores on the internalizing scale of the CBCL and 49% of the sample fell in the clinical range on the externalizing scale (Greeson et al., 2011). Another study found almost 40% of children in this high-risk population to have externalizing problems in clinical range (Vanschoonlandt et al., 2013). The problems can co-occur from early childhood into late adolescence (Gillom & Shaw, 2004). Chen (2010) suggested that scoring high on internalizing problems is highly correlated with scoring high on externalizing problems and the same counted when scoring low on either of the problems. Yet, it seems like a diagnostic and therapeutic focus lies on externalizing rather than internalizing outcomes (Batstra et al., 2014; Milette-Winfrey & Mueller, 2018).

The likelihood for the occurrence of either internalizing or externalizing outcomes differs depending on the gender (Chen, 2010; Hoftun et al., 2012) and is moderated by age (Achenbach et al., 1991; Bornstein et al., 2010; Gillom & Shaw, 2004). As stated before,

internalizing behaviors increase as children mature (Achenbach et al., 1991). Some research suggested that there are similar rates for both genders in internalizing behaviors (Chen, 2010; Kitzman et al., 2003) but others found higher rates for girls than for boys (Achenbach et al., 1991; Moylan et al., 2010; Schmid et al., 2008). Yet others found that the increase in rates for girls begins with adolescence, especially in the area of somatization (Hoftun et al., 2012).

A great amount of research indicated higher rates in externalizing outcomes for boys than for girls (Achenbach et al., 1991; Evans et al., 2008; Merikangas et al., 2010). According to Chen (2010) the rates for externalizing problems on school entry are up to 10 times higher for boys than for girls. Achenbach and Rescorla (2000) suggest one reason for this multitude could be a difference in verbal ability. They showed that girls between 18–35 months have significantly more vocabulary than boys in the same age. This enables girls to better communicate own needs and frustrations which can then be met. The advanced language, and thus, social skills also help girls to form positive relationships earlier than boys (Chen, 2010). On the other hand, a significant amount of research suggests the situation is not as clear and easily dividable regarding the genders (Achenbach et al., 1995; Kitzman et al., 2003; Sternberg et al., 2006a; Wolfe et al., 2003). Rather, there is a complex interplay of age, gender and the kind of traumatic experience that lead to internalizing and externalizing outcomes. As the current state of research is ambiguous about the effects of gender and age on externalizing and internalizing behavioral problems it will be analyzed in this work. The findings from a sample of Austrian foster children will add to the current literature and help to increase the understanding of the interplay of age and gender concerning externalizing and internalizing behavior problems.

Gender Stereotypes and Biases

We have a lot of presuppositions and expectancies, i.e. stereotypes, when it comes to sex and gender (Ellemers, 2018). There are even traits characterized as typical masculine and feminine, for example being *assertive* is supposed to be a masculine trait and being *nurturing* is described as a feminine trait (Ellemers, 2018). Even though many of these assumptions are questioned recently and gender roles are being debated, these stereotypes keep playing an important role in everyday life. Stereotypes are a very valuable source of information and are often used in situations when one has to make decisions (Bodenhausen, 2005; Bodenhausen & Wyer, 1985), but they are also prone to errors (Bodenhausen, 2005; Koch et al., 2015; Rice, 1926). When it comes to gender, it is still unclear to which degree stereotypes reflect biological differences and to which degree they are manifestations of expectancies society has for the genders. For example, men are supposed to be more aggressive than women and

indeed they are more likely to commit a felony and act violently (Jang, 2007). It is still debatable if that is due to nurture or nature. Nonetheless, facts like these also strengthen those stereotypes and increase the likelihood for people to think of a male when thinking about aggressive behavior.

Gender stereotypes influence and bias our judgement for example in employment decision making (Koch et al., 2015) and in sports refereeing decisions (Souchon et al., 2004). Bruchmüller et al. (2012) found out that these rules of thumb are even used in clinical settings by therapists and clinicians, i.e. professionals, and are thus skewing the prevalence of disorders. In their study about ADHD diagnoses, they sent case vignettes to therapists for them to give a diagnosis. Three of four vignettes did not fulfill diagnostic criteria for ADHD and still 16.7% of practitioners indicated a diagnosis for these vignettes. It was twice as likely for a boy to be over-diagnosed than for a girl. As the male-to-female ratio in ADHD lies between 3:1 and 9:1 (Bruchmüller et al., 2012), therapists likely used the representativeness heuristic (Kahneman & Tversky, 1973), which means they based their judgements on intuition, base rate and stereotypes, while leaving vital information – the symptoms – out. This subsequently led to an overdiagnosis of ADHD in general and a greater rate of misdiagnosis for boys (Bruchmüller et al., 2012). It is important to note that two of the core symptoms of ADHD are hyperactivity and inattention and therefore externalizing behaviors. The aforementioned examples, crime and ADHD, are examples of externalizing problems and are, as shown, much more likely to be attributed to males. Thus, it is likely that the pedagogues, who care for the children in our sample, are also affected by these biases. This will be subject to investigation since a biased judgement from the caregivers can have bad consequences for the children. High discrepancies could lead to unnecessary or wrong treatment that negatively affects the children and helps to maintain the impairments (Humphreys et al., 2017; Romano et al., 2018).

Cross-Informant Concordance

Research concerning children is very often performed with information about the children given by a third party. This is sometimes due to the young age of children in question and on other occasions to get a more complete picture of the situation of the child. Historically, it was thought children cannot give valuable information, even when it is about themselves as they are either unable or reluctant to speak about their problems (Achenbach, 2006). This view has changed and nowadays it is generally accepted that the child's view is essential for a good understanding of the child's concerns. Data from multiple informants is necessary to determine the right assessment and following treatment (Achenbach et al., 1987;

Achenbach, 2006). Multi-observer reports differ quite a lot which can be attributed to the unique views and access to information each observer has (Achenbach et al., 1987; De Los Reyes et al., 2015; De Los Reyes & Ohannessian, 2016). The various informants tend to experience and interact with the child in specific situations. For example, a teacher observes the child only in a school setting, while one parent might mainly interact with the child at home while the other might additionally see how the child acts at the sports club or in other settings. This can lead to three quite different impressions. The agreement between persons that observe the child in the same context tends to be higher than the concordance between observers in different situations, meaning a parent pair has higher agreement than a parent with a teacher (De Los Reyes & Ohannessian, 2016). In their meta-analysis De Los Reyes et al. (2015) found the highest concordance to be between mother and father, which seems logical as they share many situations with the child and maybe reflect together on shared experiences. The parent-child correspondence was found to be low to moderate but still stronger than the teacher-child agreement. Correlational agreement between similar pairs (e.g. both parents or two teachers) was found to be up to Pearson's $r = .60$ (Achenbach et al., 1987). Correlations between parents and teachers only averaged at $r = .28$. Correlation between children and other informants was found to be as low as $r = .22$ (Achenbach et al., 1987). Similar cross-informant correlations have been repeatedly found and are now some of the most acknowledged findings in child pathology research (Achenbach, 2006, De Los Reyes & Kazdin, 2005; De Los Reyes et al., 2015). The mean cross-informant correlations with regard to child psychopathology are between Pearson's $r = .20 - .40$, which is low to moderate (De Los Reyes & Ohannessian, 2016). Even though these concordances are rather low, they still give valuable information and are essential for precise assessment. With these numbers one has to keep in mind that multiple-informant correlation is dissimilar to inter-rater reliability (Achenbach, 2006). People ought to have similar information and point of views when it comes to inter-rater reliability, which is not the case with cross-informants, therefore the two constructs are distinct.

Interestingly, children in community samples recount more symptoms about themselves than their parents (van der Ende & Verhulst, 2005; Verhulst & van der Ende, 1992), whereas in clinical samples the opposite seems to hold true (Becker et al., 2004; Berg-Nielsen et al., 2003; Mokros et al., 1987). Another finding is that across all possible informants the agreement was found to be higher for externalizing than for internalizing behavior (Becker et al., 2004; De Los Reyes et al., 2015; Grills & Ollendick, 2002; Kolko & Kazdin, 1993). An explanation is that externalizing behaviors are, as the name suggests,

external and therefore better observable for everyone, whereas internalizing problems, like anxiety, are often hardly visible. This non-visibility could mean that these problems are detected only when they have manifested and could thus be harder to tackle in therapy. As stated before, it is noticeable that greater disagreement in ratings predicts worsened therapeutic progress, makes it more likely for patients to become therapeutic non-responders and increases the chance of developing more severe psychopathologies (Humphreys et al., 2017; Romano et al., 2018). Other research found greater self-other discrepancies to predict future emotional and behavioral difficulties for internalizing problems (Ferdinand et al., 2004).

Another aspect worth looking at is, if there is a greater correspondence between informants and younger children than with older children or adolescents (Achenbach et al., 1987; van der Ende & Verhulst, 2005; Verhulst & van der Ende, 1992). This would be in line with the reasoning that internalizing problems tend to occur as children mature (Kovacs & Devlin, 1998). This means the less observable problems increase with age and therefore the discrepancy is likely to increase. It is also suggested that younger children might be more restricted as to when and where they display behavior problems and therefore it is more easily observable (De Los Reyes & Kazdin, 2005), because the problems are always shown in similar situations. Other research could not replicate this finding of higher concordance with younger children (Choudhury et al., 2003; De Los Reyes et al., 2015; Kolko & Kazdin, 1993). Jensen et al. (1999) found a tendency, though insignificant, for younger children to indicate less problems than their parents. Yet other research found parent-child concordance to be higher for older than younger children (Grills & Ollendick, 2003; Lauth et al., 2010; Romano et al., 2018). Another interesting finding is that there seems to be less agreement for boys and others than for girls and others in the context of victimization (Ladd & Kochenderfer-Ladd, 2002). Gender roles could make it more difficult for males to admit to being victimized than for females. But since the roles of age and especially gender in the context of cross-informant concordance are still unclear it is important to look into this in more detail. In order to improve assessment and consequential treatment an important part of the present study is to investigate these coherences more thoroughly.

Research Questions and Hypotheses

The current research is a cross-informant analysis of externalizing and internalizing behavior outcomes measured with the CBCL. Part of the hypotheses will try to replicate findings as the existing literature is still ambiguous in certain parts. Other hypotheses are aimed at areas that have not been thoroughly investigated. Thereby, this research intends to

contribute to the current state of research and possibly extend the understanding of certain mechanisms.

Question 1: Are there differences in rates of internalizing outcomes compared to externalizing outcomes in foster children? If yes, are these differences moderated by gender? A unidirectional hypothesis cannot be derived from the current state of research as it is too ambiguous, thus the following hypothesis is of explorative nature.

Hypothesis 1 (H1): There are differences in rates of internalizing outcomes compared to externalizing outcomes in foster children.

Question 2: Do younger children (age 10–14) indicate more externalizing problems than adolescents? The literature suggests a decline of externalizing problems during maturation for most children (Fanti & Henrich, 2010). This decline assumes that typical child development can take place which is often not the case for foster children. Therefore, it is unclear if this finding holds for this high-risk population.

Hypothesis 2 (H2): Younger children (age 10–14) indicate more externalizing problems than older children and adolescents (age 15–19).

Question 3: Do younger children (age 10–14) indicate more externalizing than internalizing problems? Literature findings are consistent with the fact that internalizing problems tend to increase with age, while externalizing problems tend to do the opposite (Achenbach et al., 1991; Bornstein et al., 2010). The following hypothesis aims to contribute to the existing findings.

Hypothesis 3 (H3): Younger children indicate more externalizing than internalizing problems for themselves.

Question 4: Do caretakers have more cross-informant correspondence with younger or older children? There are contradictory findings in the literature that do not justify a directed hypothesis.

Hypothesis 4 (H4): Cross-informant correspondence between caregivers and younger children (age 10–14) is different than between caregivers and older children (age 15–19).

Question 5: Is there more agreement between cross-informants regarding externalizing or internalizing outcomes? It has been shown that externalizing problems have higher cross-informant agreement (De Los Reyes et al., 2015), thus this hypothesis is directed at replicating these findings.

Hypothesis 5 (H5): There is stronger cross-informant agreement on externalizing outcomes than on internalizing outcomes.

Question 6: Are caregivers prone to be affected by gender bias when indicating externalizing problems? A gender bias concerning externalizing behaviors has been found for males (Bruchmüller et al., 2012). A gender bias for externalizing behaviors in females has not been subject to investigation yet, therefore H6b will be openly explored.

Hypothesis 6a (H6a): Caregivers indicate more externalizing problems for boys than boys indicate for themselves.

Hypothesis 6b (H6b): Caregivers indicate different rates of externalizing problems for girls than girls indicate for themselves.

Question 7: Is the discrepancy between children and caregivers concerning internalizing problems greater for one gender? Girls learn to communicate their needs and frustrations earlier than boys and thus have those needs more often met (Achenbach & Rescorla, 2000). This could lead to increased sharing of problems by girls, as need fulfillment and/or decreased frustration is expected when sharing problems with others. Internalizing behaviors are hardly visible and thus communicating these increases the concordance between self and other ratings.

Hypotheses 7 (H7): The discrepancy concerning internalizing behaviors is greater for boys and caregivers than for girls and caregivers.

Methods

Study Design

This master thesis is embedded into a broader, ongoing field study done by a research group lead by Prof. Lueger-Schuster at the University of Vienna. The contracting authority and funding entity is the state of Niederösterreich. The “Schemastudie” has a correlative cross-section design. Measurements points are spaced in 1 to 2 years intervals. Various screening instruments were used to gain comprehensive data of the psychological welfare of youths living in foster care. The main interest lies in the assessment of trauma related symptoms, depressive symptoms and anxiety symptoms. This master thesis only uses the data obtained at the second measurement point, which was between March and September of 2019, and focuses on the CBCL (German Version: Döpfner et al., 2014) which assesses internalizing and externalizing behavior problems.

The participants were children and adolescents living in institutional care and their caregivers. The data gathering was done directly in the six different foster care homes (social pedagogue care centers) in Niederösterreich. The assessment was collected with paper and pencil questionnaires. The children self-assessed themselves and the caregivers got corresponding questionnaires to pass their judgement on the children. The testing of the

children was done while either psychologists or psychology students of the University of Vienna were present in order to assist the children if any problems arose and to support them. The complete test duration was 180 minutes, including breaks. One psychologist supervised two children at a time. In some cases children got individual supervision, for example when behavioral problems were known or emotional and cognitive difficulties were expected. It was also possible to split the testing into two 90 minutes sessions. The board of ethics of the University of Vienna approved the study beforehand.

At the beginning of the testing participants were made familiar with the aim of the study and the broader context. As the main source of data gathering was the children, it was of utmost importance to make their experience as comfortable as possible. Prior to the testing the voluntary nature of the participation, the anonymity of their data and the possibility to have their data erased until one month after the testing was explained to children. The participants had a sight-proof workspace and it was explained to them that they could make breaks whenever they wished to and ask questions at any time. The written consent was signed prior to filling out the questionnaires. In order to fill out the test battery as unobstructed as possible various instructions for the different questionnaires and scales were given beforehand. When psychologists noticed struggles or emotional difficulties they intervened. To regulate the strain on the children, a break was implemented every 30 minutes if the children did not require one earlier. In the end of the testing the children were asked about any problems that occurred and their emotional wellbeing. All of these actions were put into place to make the experience for the children as reasonable as possible.

Participants

The participants were children and adolescents in foster care and were between 10 and 19 years of age at the time. They were recruited from the six foster care homes in Niederösterreich and it was aimed that all children and adolescents participated provided that they wanted to and did not meet any exclusion criteria. The exclusion criteria were suicidality, acute psychotic symptoms, drug addictions, mental retardation, physical contradictions and lack of sufficient German proficiency. In total $N = 161$ youths participated of which, in a first step, $n = 36$ had to be excluded due to inappropriate handling of the tests or due to obvious lack of comprehension. In a second step $n = 6$ other children had to be excluded for this particular thesis since the age of the children was missing, which is vital information in this context. This leaves a total $n = 119$ of children and adolescents ($M_{\text{Age}} = 14.85$, $SD = 2.42$, range 10 to 19 years) for the data analysis. 78 participants were males (65.5%) and 41 were females (34.5%). The vast majority of the children were Austrian citizens (92.1%). The other-

perspective for all children was given by their caregivers, so that every case includes the values of the self and other perspective. The caregivers' professions were child/ youth care workers, social pedagogues and in some cases psychologists. The children were unequally distributed to the six different foster care homes.

Assessment instrument

Child Behavior Checklist (CBCL)

The CBCL, originally developed by Achenbach and Edelbrock (1983) and revised since then, is used in the German version (Döpfner et al., 2014). The CBCL-YSR (Youth Self Report) for the children and adolescents and the CBCL/6-18, which is the caregiver version, were used in this study. The two versions have only minor differences as will be elaborated further. The CBCL is an instrument to assess internalizing and externalizing behavior problems in children and adolescents. Internalizing behaviors reflect inwards directed behaviors, including depression, anxiety and somatic problems. Externalizing behaviors are outwards directed, such as aggression towards others, impulsivity and rule-breaking behaviors (e.g. Achenbach et al., 1987). It is a universally used and standardized instrument which is regarded as both valid and reliable (Achenbach & Rescorla, 2000; Döpfner et al., 2014). The caregiver version of the CBCL can be used to assess children and adolescents from age 6 to 18 years (Döpfner et al., 2014). It consists of 120 items that assess various emotional and behavioral problems the child has or had during the last six months. The items can be grouped in eight syndrome scales that have been established with factor analysis (Döpfner et al., 2014).

Three of these syndrome scales can clearly be attributed to the broadband *Internalizing Problems* scale. The *Anxious/Depressed* scale consists of 14 items asking about anxiety related symptoms (e.g. "Worries", "Fears going to school") and depressive ones (e.g. "Unhappy, sad, or depressed"). The nine item *Withdrawn/Depressed* scale reflects on isolating behaviors (e.g. "Would rather be alone than with others", "Withdrawn, doesn't get involved with others"). The *Somatic Complaints* syndrome scale also consist of nine items and asks about physiological problems (e.g. "Stomachaches", "Feels dizzy or lightheaded").

Two of the syndrome scales can be categorized into the *Externalizing Problems* scale. One is the 13-item *Rule Breaking/Delinquent Behavior* scale (e.g. "Steals at home", "Doesn't seem to feel guilty after misbehaving") denoting delinquent and disobedient behaviors. The other is the *Aggressive Behavior* scale consisting of 20 items reflecting on aggressive behaviors (e.g. "Physically attacks people").

Then there is the nine items *Social Problems* scale reflecting on interpersonal problems (e.g. “Doesn’t get along with other kids”), a seven item *Thought Problems* scale denoting problematic cognitions and repetitive behaviors (e.g. “Sees things that aren’t there”, “Repeats certain acts over and over, compulsions”) and the *Attention Problems* scale consisting of 11 items with statements concerning concentration difficulties and task management (e.g. “Inattentive or easily distracted”). These scales, as well as 33 items that are uncategorized, partly fit both, *Internalizing* and *Externalizing Problems*, broadband scales. The 33 other items include statements that denote problems not asked before (e.g. “Sleeps less than most kids”, “Thumb-sucking”).

The eight syndrome scales have sufficient internal consistency with Cronbach’s α between $\alpha = 0.65$ and 0.92 (De Caluwé et al., 2013). The CBCL also yields three broadband scales: *Externalizing Problems*, *Internalizing Problems*, and *Total Problems*. Cronbach’s α ranges from $\alpha = .85$ to $.94$ for the three broadband scales (Schmid et al., 2008). The items are scored on a 3-point-Likert scale ranging from 0 = *not true*, 1 = *somewhat or sometimes true*, to 2 = *very true or often true*. The data is reported as both raw and standardized T scores. These T scores are age and sex normed and are derived from the raw scores. They are normally distributed, with a mean of 50 and a standard deviation of 10. These T scores are built in order to improve comparison between age groups and different genders (Achenbach et al., 1987). Higher T scores, starting one *SD* above the mean, indicate substantial internalizing and externalizing behavior problems. T scores between ≥ 60 and ≤ 63 are considered to be *Borderline*, while T scores at or above the cutoff value ≥ 64 are considered to be in the clinical range (Döpfner et al., 2014).

118 of the 120 items are identical with the CBCL Youth Self Report (CBCL-YSR), which the children fill out. The two extra items are open questions about symptoms and problems that have not been listed. These two items are not included in the CBCL-YSR.

Both versions of the CBCL have a section that asks about activities and competencies of the child. In this section it can be indicated, for example, how many friends the child has and how often it meets the friends compared to how often other children in the same age meet their friends. This is an example of social competency. Academic competency is assessed by asking how well the child performs in various subjects. Activities are assessed by asking about sports the child does and hobbies it has. It is then asked how often and how good the child does the activity compared to others in the same age. This section is clustered into four competency scales (activities, social, academic, and total competency). The four competency

scales were only included for the CBCL version for the caregivers but not for the CBCL-YSR in the current study and thus are not a focus of this thesis.

Statistical Analysis

The statistical analysis was performed by using SPSS 26 (IBM Corp., 2019). The only exceptions were the correlation comparisons, which were calculated with an online calculator (Hemmerich, 2017) for Fisher's z transformation as SPSS has no default way to compare Pearson's r . Prior to analysis all cases that met the exclusion criteria and that had missing age information as well as the ones that showed obvious problems during testing were excluded. The final data set was first analyzed for missing data.

Missing Data Analysis

The missing data analysis was done for all $n = 119$ cases. The analysis was performed on the syndrome scale level as missing items within each scale led to the particular scale not being computed. Missing data for one scale further led to the corresponding broadband scale not being calculated since the broadband scales denoting externalizing and internalizing problems consist of the various syndrome scales. Having no value for either the *Externalizing Problems* or *Internalizing Problems* scale made it impossible to determine the *Total Problems* score as well. Thus, the number of valid cases for the *Total Problems* scale was the smallest for both the self and other perspective. It was refrained from imputing missing values as the risk to skew the data was too high. The missing data from the child and adolescent perspective as well as the missing data from their caregivers were analyzed and can be seen in Table 1. The data of the caregiver perspective has $n = 16$ missing cases on the *Total Problems* scale, which is 13.4% of the sample and also the maximum of missing data in this sample. The available data for some comparisons is therefore less than for others. The self-other comparisons could only be done for 97 cases, as these cases were the only ones that both caregivers and children had fully completed. Most missing data stems from the syndrome scale *Somatic Complaints* filled out by the caregivers.

Analysis of outcomes within children

The comparison of internalizing and externalizing outcomes within the children was done with One-Sample T-tests. Cases with missing values were excluded analysis by analysis. The standardized T scores of the CBCL were obtained from the descriptive statistics. The analysis was done with the data of $n = 116$ children as three children did not have complete data. Significant mean differences were tested on a 95% significance level.

Table 1*Cases of Missing Data for all Syndrome and Broadband Scales (N = 119)*

	Anxious/ Depressed	Withdrawn / Depressed	Somatic Complaints	Social Problems	Thought Problems	Attention Problems	Rule Breaking Behavior	Aggressive Behavior	Internal- izing Problems	External- izing Problems	Total Problems
Children	0 (0.0)	1 (0.8)	3 (2.5)	2 (1.7)	3 (2.5)	0 (0.0)	0 (0.0)	2 (1.7)	3 (2.5)	2 (1.7)	7 (5.9)
Caregivers	0 (0.0)	2 (1.7)	13 (10.9)	0 (0.0)	3 (2.5)	0 (0.0)	1 (0.8)	0 (0.0)	15 (12.6)	1 (0.8)	16 (13.4)

Note. Numbers in brackets () are the percentages the missing cases make of the total sample.

Analysis of outcomes between children

The comparison between the two different age groups (10–14 years and 15–19 years), as well as the two genders (there was no identified transgender or non-binary person in the sample) was done with Independent-Samples T tests. Bootstrapping was also performed for all tests.

Analysis of outcomes between children and caregivers

The comparison of outcomes between children and their caregivers was done with Paired-Samples T tests. As shown in Table 1 the data for the *Total Problems* scale was missing for $n = 16$ cases for the caregiver version. The children also had some missing data for some of the cases. These missing cases were not identical, which left $n = 97$ cases that had complete data from the children and caregiver perspective. Only these cases were included in this analysis. The Pearson's correlations between the scores of children and caregivers were analyzed with Fisher's z transformation (Fisher, 1921).

Results**Social Structure**

The majority of the children and adolescents went to special needs schools (33.6%), followed by lower secondary education schools (23.5%). Sixteen percent of the children were in apprenticeships at the time of testing, while two children went to an academic high school (1.7%) and one child was still in primary school (0.8%). The educational data was missing for 24.4% of the children. Most children still had contact to their parents (95.8%) and 79.8% of them saw either one or both parents at least once a week. Unfortunately, it was not obtained if the children saw both their parents or only one parent and if the parents were still living in one common household or if new partners moved in with a parent. More than 47% percent of the mothers and 39.5% of the fathers are employed and working, but 7.6% of the mothers indicated household chores as work, which is labor but technically counts towards unemployment. Another 7.6% of both mothers and fathers were unemployed. Only 3.4% of the fathers had a profession for which higher education is needed and none of the mothers had a job for which that is a prerequisite. One mother and one father were in jail at the time, while one father and three mothers were already deceased. The data was missing for 46.2% of the fathers and 29.4% of the mothers.

Table 2*Categorical Self-Assessments in Percentage for the Three Broadband Scales (N = 119)*

	Internalizing Problems		Externalizing Problems		Total Problems	
	Borderline ^a	Clinical ^b	Borderline ^a	Clinical ^b	Borderline ^a	Clinical ^b
Male (n = 78)	11.7	31.2	6.5	14.3	10.7	28.0
Combined	42.9		20.8		38.7	
Female (n = 41)	5.1	51.3	17.5	30.0	10.8	45.9
Combined	56.4		47.5		56.7	
All Children (n = 119)	9.5	37.9	10.3	19.7	10.7	33.9
Combined	47.4		30.0		44.6	

Note. ^a T-values between ≥ 60 and ≤ 63 are considered 'Borderline'. ^b T-values from ≥ 64 are considered 'Clinical'.

Table 3

Means & Differences of Self-Other Judgements for the Three Broadband Scales (N = 97)

	Internalizing Problems		Externalizing Problems		Total Problems	
	Self <i>M (SD)</i>	Caregiver <i>M (SD)</i>	Self <i>M (SD)</i>	Caregiver <i>M (SD)</i>	Self <i>M (SD)</i>	Caregiver <i>M (SD)</i>
Age 10–14 (<i>n</i> = 44)	61.11 (11.78)	62.41 (10.59)	54.23 (10.24)	62.36 (14.88)	59.0 (10.46)	66.16 (11.79)
Difference	1.3 (14.01)		8.14** (15.86)		7.16** (13.77)	
Age 15–19 (<i>n</i> = 53)	59.15 (12.55)	63.58 (9.15)	55.75 (9.88)	59.08 (9.95)	57.51 (11.35)	63.11 (9.01)
Difference	4.43* (14.02)		3.32* (10.73)		5.6* (12.21)	
Male (<i>n</i> = 68)	58.0 (12.29)	62.43 (9.82)	53.53 (9.73)	60.1 (13.41)	56.22 (10.77)	63.88 (11.07)
Difference	4.43* (15.44)		6.57** (14.14)		7.66** (14.26)	
Female (<i>n</i> = 29)	64.83 (10.65)	64.52 (9.76)	58.66 (9.93)	61.66 (10.06)	62.79 (10.02)	65.93 (8.71)
Difference	-0.31 (9.38)		3.0 (11.5)		3.14* (8.25)	
All Children (<i>n</i> = 97)	60.04 (12.18)	63.05 (9.79)	55.06 (10.02)	60.57 (12.47)	58.19 (10.92)	64.49 (10.42)
Difference	3.01* (14.03)		5.51** (13.45)		6.31** (12.89)	

Note. Mean values are standardized T scores. Difference $M = \text{Caregiver}M - \text{Self}M$. Asteriks indicate significant mean differences.
 * $p \leq .05$. ** $p \leq .01$.

Prevalence of Internalizing and Externalizing Behavior Problems

There are two different results for the prevalence for both kinds of behavioral problems. The self-assessed prevalence given by the children and the prevalence judgement by their caregivers. Standardized T scores from the CBCL starting at ≥ 60 are considered borderline, while scores ≥ 64 fall into the clinical range. As the borderline scores start already one *SD* above the mean, the percentages of both borderline and clinical scores were combined to illustrate the vulnerability of this high-risk sample. The self-assessed prevalence, meaning all scores starting at the borderline level, for internalizing problems lies at 49.4%, while the caregiver judgement indicates a prevalence of 66.3%. This probably means that the true prevalence lies somewhere between those two scores and it is fairly clear that more than half of the sample have serious internalizing behavior problems. The prevalence for externalizing behavior problems also lies between the self-assessed 30% and 50.8% given by the caregivers. The prevalence for the total problems, a mix of internalizing and externalizing problems combined, is between 44.6% (self-assessed) and 65.1% (other-assessed). The prevalence for girls for internal, external and total problems is higher than for boys, as can be seen more detailed in Table 2 and through the group means in Table 3.

Outcomes Within Children

The percentage of children who rated themselves as borderline and within the clinical range concerning internal and external problems, leading to the *Total Problems* score, can be seen in Table 2. Considering almost half (44.6%) of the children in this study score at least in the Borderline range on the *Total Problems* scale it is clear that this is a very vulnerable group. The comparison of outcomes yielded that there is a significant difference in rates of internalizing outcomes compared to externalizing outcomes with all children indicating more internalizing problems than externalizing ones ($t(115) = 4.624, p < .001$). Therefore hypothesis 1, which was of explorative nature, finds support as there are differences in rates for foster children. This tendency for the children to indicate more internalizing than externalizing problems is similar across all age groups and without any significant gender difference. The caregiver scores regarding internalizing and externalizing behavior problems were also analyzed. The result of the caregivers mean comparison endorses the notion that children in this sample suffer more internalizing than externalizing behaviors ($t(103) = 2.109, p = .037$).

The fact that all children indicated more internalizing behaviors than externalizing behaviors for themselves also means that hypothesis 3 needs to be rejected. In the age group of 10–14 years old it was expected that children show more externalizing than internalizing

behaviors. Interestingly, the opposite seems to hold true, as they indicated significantly more internalizing behaviors ($t(53) = 3.997, p = .999$). The analysis was done again for the age group of 10–12 years old to see if it made a difference when only the youngest children in the sample were included but the result was in essence the same ($t(21) = 2.246, p = .982$) even though the trend was less strong.

Outcomes Between Children

The second research question was if younger children (10–14 years old) show more externalizing behavior problems than older children (15–19 years old). This hypothesis (H2) needs to be rejected as well ($t(115) = 1.374, p = .696$), as there is no significant difference between the two groups. In contrast to what was expected the older age group had slightly higher scores on the broadband *Externalizing Problems* scale and indicated significantly more on the *Rule-Breaking Behavior* scale ($t(117) = 2.338, p = .021$), one of the externalizing syndrome scales, for themselves than the younger age group. There were no other significant mean differences between the age groups on any of the other syndrome scales, including the internalizing syndrome scales. The analysis of the caregivers' judgements yielded non-congruent results. They reported more externalizing behavior outcomes for the younger age group ($M = 62.86, SD = 14.47$ vs. $M = 58.82, SD = 9.89$), although insignificant ($t(101) = -1.678, p = .096$). They indicated notably higher significant scores for the *Aggressive Behavior* scale ($t(101) = -2.96, p = .004$) for the younger children.

Comparison of Self and Other Judgement

The comparison of self-other concordance was done for all children and their caregivers. It was hypothesized (H5) that there is stronger cross-informant agreement on externalizing outcomes than on internalizing ones. The mean differences between children and caregivers were significant for the *Internalizing Problems* scale ($t(95) = 2.114, p = .037$) and highly significant for the *Externalizing Problems* scale ($t(95) = 4.032, p < .001$). Still, the raw correlation concerning externalizing behaviors ($r(95) = .301, p = .003$) was stronger than for internalizing behavior problems ($r(95) = .199, p = .05$). Yet, Fisher's z transformation yielded a non-significant difference between the two cross-informant correspondences ($z = 0.8263, p = .204$). Even though there might be a trend of higher concordance for externalizing problems the hypothesis that there is stronger cross-informant correspondence for externalizing behaviors does not hold and thus needs to be rejected.

Age Differences in Cross-Informant Correspondence

The cross-informant correspondence was analyzed regarding differences between the two age groups with their caregivers. For hypothesis 4 it was expected that there is a

difference in the self-other correspondence for different age groups, but as the existing literature is very indistinct, it could not be said in which direction a possible difference would go. As can be seen in Table 3 the comparison of the mean differences for the three broadband scales resulted in significant mean differences between caregivers and children for all scales in the older age group (15–19 years old). The younger age group (10–14 years old) only had significant mean differences with their caregivers on two of the broadband scales but not on the internalizing scale ($t(42) = 0.613, p = .543$). In order to assess the concordance between children and caregivers the correlations (Pearson's r) of their judgements have to be analyzed. The correlations itself are not eligible data for comparison and were transformed into standardized Z scores first. The comparison of the correlations was done with Fisher's z transformation (Fisher, 1921). The differences in concordance of internal problems between the younger group with their caregivers and the older group with their caregivers was non-significant ($z = 0.119, p = .905$). The analysis yielded, in essence, the same outcome for external behavior problems ($z = -0.8984, p = .369$) as well as for the *Total Problems* scale ($z = -0.3019, p = .763$). As all these comparative analyses were insignificant, the raw correlations can only be cautiously looked at as possible trends. The correlation of the *Internalizing Problems* scale of caregivers with younger children is higher ($r(42) = .219, p = .153$) than the correlation of the same scale in the older age group ($r(51) = .195, p = .162$). The older age group has more cross-correspondence with their caregivers on the *Externalizing Problems* ($r(51) = .414, p = .002$ vs. $r(42) = .246, p = .108$) and *Total Problems* ($r(51) = .298, p = .03$ vs $r(42) = .239, p = .118$) scale compared to the younger group. A more detailed comparison of the cross-informant correlation of all scales can be found in Table 4. As can be seen, the higher correlation of the *Internalizing Problems* scale between younger children and caregivers can mainly be attributed to the syndrome scale *Withdrawn/Depressive*. Otherwise, there is either higher correlation between older children and caregivers on most other scales or no mentionable differences between the age groups. Research question 4 was of explorative nature as no directed result could be expected. It was found that there were no significant group differences in cross-informant correspondence for the two age groups, thus hypothesis 4 has to be rejected. When looking at the raw correlations one can see small, yet insignificant trends in that the younger age group had higher cross-informant correspondence on internalizing outcomes, while the older age group had higher self-other concordance on externalizing outcomes.

Table 4*Correlations between Children and Caregivers (N = 97)*

	Anxious/ Depressed	Withdrawn/ Depressed	Somatic Complaints	Social Problems	Thought Problems	Attention Problems	Rule Breaking Behavior	Aggressive Behavior	Internal- izing Problems	External- izing Problems	Total Problems
Age 10–14 (n = 44)	.078	.402**	.221	.263	.032	.226	.187	.189	.219	.246	.239
Age 15–19 (n = 53)	.392**	.07	.199	.429**	.352**	.168	.405**	.468**	.195	.414**	.298*
Male (n = 68)	.071	.152	.167	.197	.077	.20	.281*	.296*	.037	.286*	.147
Female (n = 29)	.499**	.336	.319	.686**	.474**	.349	.272	.354	.581**	.337	.62**
All children (N = 97)	.262*	.239*	.211*	.368**	.209*	.209*	.285**	.306**	.199*	.301**	.271**

Note. Correlations are presented as raw Pearson's *r*. Asteriks signify that children's and caregiver's correlations are significantly positively related.

* $p \leq .05$. ** $p \leq .01$.

Gender Differences in Cross-Informant Correspondence

Gender differences for the self-other correspondence are a key research aspect and were therefore analyzed. Hypothesis 6a finds support as caregivers do indicate significantly more externalizing problems for boys ($M = 60.1$, $SE = 1.61$) than boys indicate for themselves ($M = 53.53$, $SE = 1.21$). The one-sided paired T-test yielded a highly significant result ($t(67) = 3.834$, $p < .001$). The caregivers also indicated significantly more internalizing behavior problems for the boys than they did for themselves ($t(67) = 2.364$, $p = .021$) even though the difference was not as big as for externalizing outcomes. The mean scores and their differences can be found in Table 3.

The research question concerning hypothesis 6b was of explorative nature and it was purely assumed that there is a difference in the self-other perspective regarding externalizing behaviors between caregivers and females. This hypothesis does not hold in our sample ($t(27) = 1.404$, $p = .171$). The trend, nonetheless, goes in the same direction as for the males, in that caregivers indicate higher rates of externalizing behaviors than the girls (see Table 3). Further analysis showed an interesting result that there is hardly any difference in the perception of internalizing behavior outcomes between caregivers ($M = 64.52$, $SE = 1.76$) and females ($M = 64.83$, $SE = 1.91$). The mean difference is highly non-significant ($t(27) = -0.178$, $p = .86$) and females indicate slightly higher rates of internalizing problems for themselves than their caregivers did. The cross-informant results regarding internalizing behavior outcomes for both genders are in line with the last research question. Hypothesis 7 states that the discrepancy concerning internalizing behaviors is greater for boys and their caregivers than for girls and their caregivers. Girls and their caregivers indicated almost identical means and have a relatively high correlation on internalizing outcomes ($r(27) = .581$, $p = .001$). Boys and caregivers indicated significant different scores with almost no correlation ($r(66) = .037$, $p = .763$), therefore it comes as no surprise that the one-sided Fisher's z transformation proves the difference between the two group interactions as being highly significant ($z = -2.702$, $p = .003$). Hypothesis 7 is supported in this sample.

Discussion

This master thesis aims at gaining new insights regarding externalizing and internalizing behavior problems in foster care children. A special focus laid on the cross-informant concordance between the children and their institutional caregivers concerning these behavioral problems, as low concordance of judgement positively correlates with less favorable therapeutic outcomes and increased psychopathology (Humphreys et al., 2017). It was assumed that gender and especially possible gender biases would play a vital role in the

caregiver assessment of the children, since it was shown that gender stereotypes can have a huge impact on the clinical assessment and subsequently allocation of diagnoses (Bruchmüller et al., 2012). This research hopes to shed some more light on underlying factors contributing to differences in self-other agreement.

Remarks to the Social Structure

The descriptive background data of the sample shows slightly more than a third of the children (33.6%) went to special needs schools, which is an indication that many of these children are somewhat, either cognitively and/ or emotionally, impaired, which in turn could likely be the result of traumatic experiences during childhood (Aas et al., 2012, van Os et al., 2017). Considering that a substantial number of children had to be excluded from the sample due to difficulties understanding and filling out the questionnaires, this explanation suggests itself even more. The fact that only two children went to an academic high school can be read as another sign that traumatic experiences have a negative effect on academic achievement. The educational data already clarify that this sample is not representative for the general population. The available data for the parents suggests that most children come from low socio-economic background since only a very small amount of the fathers (3.4%) and none of the mothers had work for which higher education was needed. The unemployment rate was, at least for the mothers, twice as high (15.2%) as for the general Austrian population in 2019 (7.4%: Arbeitsmarktservice [AMS], 2020), which supports this notion. These high numbers of unemployment can have multiple reasons, like low education and mental health problems, both often leading to poverty. Poverty is another risk factor for children to get into foster care (Meyer et al., 2010). Parental psychiatric disorders, especially substance abuse disorders, are one of the main reasons why children need to be taken out of their families and into institutionalized care (Meyer et al., 2010). Additionally, parental mental health problems predict maltreatment of children (Chaffin et al., 1996). This reasoning is in line with McAuley & Trew (2000) who define children in substitute care as a high-risk population for undergone traumatic experiences.

Summary of the Results

Almost half of the children scored at least in the borderline range of the *Total Problems* scale, marking the sample as being from a high-risk population. 47.4% of the children had at least a borderline score on the *Internalizing Problems* scale, while only 30% had a similar result on the *Externalizing Problems* scale. These rates are significantly different from each other and therefore hypothesis 1 holds. Children of all age groups and both genders indicated more internalizing behavior outcomes than externalizing ones.

This result means hypothesis 3 needs to be rejected as it was expected that younger children indicate more externalizing than internalizing problems for themselves. They indicated significantly more internalizing problems ($M = 61.11$, $SD = 11.78$) than externalizing ones ($M = 54.23$, $SD = 10.24$).

Hypothesis 2 states that younger children indicate more externalizing behavior outcomes than older children, which also did not happen ($p = .696$), thus this hypothesis needs to be rejected as well. Contrary to the expectation, the older age group had higher, yet insignificant, scores on the *External Problems* broadband scale.

Hypothesis 4 expected that there are differences in the cross-informant correspondence for different age groups. The self-other correspondence between caregivers and the younger age group and the one between the older age group and caregivers were neither significantly different from each other on the *Internalizing*, nor the *Externalizing*, nor the *Total Problems* broadband scale and thus hypothesis 4 needs to be rejected as well.

The cross-informant concordance was analyzed for all children in order to see if there is more agreement on internalizing or on externalizing behavior outcomes. Even though the trend went in the envisaged direction of more agreement for externalizing problems the analysis yielded a non-significant result. This means hypothesis 5 needs to be rejected for this sample.

Hypothesis 6a looked at cross-informant agreement concerning externalizing problems for males. Caregivers indicated highly significant more externalizing problems than boys did for themselves. Hypothesis 6a holds as this was the expected result. For research question 6b, on the other hand, a significant difference in the judgement of externalizing problems for females was expected without specifying if caregivers would indicate less or more externalizing behavior problems than the girls. Caregivers indicated more externalizing behavior outcomes for the girls than they indicated for themselves, comparable to the boys, but the trend was less strong and thus insignificant. Hypothesis 6b therefore needs to be rejected.

The self-other agreement concerning internalizing problems was significantly stronger for females with their caregivers than for males with their caregivers, which means that hypothesis 7 holds in this sample. Stronger correspondence for females was expected as the existing literature points in that direction (Achenbach & Rescorla, 2000; Ladd & Kochenderfer-Ladd, 2002)

Elaboration on the Results

The results need to be put in perspective and within a framework in order to give some possible explanations on their come abouts. Since it is quite interesting that more than half of the hypothesis needed to be rejected for this sample of a high-risk population, these findings will be extensively discussed.

Prevalence

To compare the prevalence data with similar research, the caregiver assessment of the prevalence has to be used as this was most often the source of prevalence in other studies. One study done in foster care institutions reported externalizing scores in the clinical range for 49% of the children in their sample but unfortunately did not report any scores in the borderline range (Greeson et al., 2011). Another study done in this population reported almost 40% of children to score at least in the borderline range of externalizing problems (Vanschoonlandt et al., 2013). The caregivers in this study indicated borderline plus clinical scores for 50.8% (38.1% clinical only) of the sample, which corroborates the existing findings in that a substantive proportion of this high-risk population suffers from extensive externalizing problems.

A somewhat different conclusion can be drawn for internalizing problems as the caregivers in this study indicated 66.3% of the children to fall into the combined borderline and clinical range, while 49% of the sample are in the clinical range. Greeson et al. (2011) found only 37% of children to be in the clinical range and others reported only 28% of their sample had clinically significant internalizing problems (Huffhines et al., 2020), which is quite a bit lower. It is interesting to see that almost two thirds of the children in our sample score at least in the borderline range for internalizing problems. Some studies tend to find more externalizing than internalizing problems (Bornstein et al., 2010, Greeson et al., 2011), while others found the opposite to be true (Sternberg et al., 2006a; Sternberg et al., 2006b). This study contributes to the latter by having higher internalizing scores throughout the sample.

Slightly more than 65% of the children, as seen by the caregivers in this study, score in the borderline plus clinical range (51.5% clinical only) on the *Total Problems* scale. Other studies found clinically significant behavior problems ranging from well over 40% (Burns et al., 2004; Clausen et al., 1998) to 65% (girls) and 74% (boys: Tarren-Sweeney, 2008). Similar to Tarren-Sweeney (2008) children in our study indicated particularly high scores for the syndrome scales *Attention Problems*, *Thought Problems* and *Rule Breaking Behavior*, additionally children had especially high scores on the *Withdrawn/Depressed* scale. These

results are in line with findings that between 67% (McCann et al., 1996) and 74% (Dölitzsch et al., 2014) of children in care settings have at least one psychiatric disorder, with conduct disorders, anxiety disorders, major depressive disorders and attention deficit disorders (in that order) being the most prominent (McCann et al., 1996). All of these disorders frequently occur in combination with PTSD, i.e. there is a high co-morbidity (McLeer et al., 1994; Saigh et al., 2002). PTSD symptomology in children and adolescents differs from adults (Scheeringa et al., 2006). For example, hyperactivity, impulsiveness and inattention are not only the core ADHD symptoms but also behavior outcomes very often displayed by children with a PTSD. Externalizing problems frequently stem from maltreatment (Herrenkohl et al., 1998; 2008; Sousa et al., 2011) and thus traumata (Luoni et al., 2018), which is why these results could signify that many children in our sample suffer from post-traumatic stress.

It is noticeable that the caregivers' judgement for every single syndrome scale, except *Somatic Complaints*, and each of the three broadband scales exceeded a mean T value of 60 for the whole sample, which is another indication of the vulnerability of this population. Another interesting finding is that there are more internalizing than externalizing behavior problems indicated by both children and caregivers. Some studies found externalizing problems to be more prominent (Greeson et al., 2011; Moylan et al., 2010), especially for males (Achenbach et al., 1991; Chen, 2010; Evans et al., 2008; Gauthier-Duchesne et al., 2017), while this sample, with a majority of it being male (65.5%), did not replicate these existing findings. One possible explanation for the results could be high rates of PTSD in our sample as there seems to be a strong connection between high internalizing scores on the CBCL with a PTSD diagnosis (Saigh et al., 2002; Wolfe et al., 1994). This would be in line with findings that children in foster care have increased levels of post-traumatic stress compared to peers who are considered at risk but not in foster care (Greeson et al., 2011; Rayburn et al., 2016). Wolfe et al. (1994) found significantly higher internalizing scores for sexually abused children with PTSD than for sexually abused children without PTSD, while there were no significant differences for both groups on the CBCL *Externalizing* scale. The hypothesis of scoring higher on internalizing problems when having a PTSD diagnosis compared to not having one, could be looked at in a follow-up study for this sample.

Age Differences

It was hypothesized (H3) that there will be age differences regarding the behavioral outcomes to the effect that younger children indicate more externalizing than internalizing behavior problems, which was not supported. Not only in the overall sample, but already in the younger age group more internalizing problems were indicated. The analysis of the

younger group (10–14 years) yielded significantly more internalizing than externalizing problems. This was even reconfirmed in the analysis of the youngest children (10–12 years) in our sample, even though the trend was less strong. This result seems to be in opposition with the presupposition that externalizing problems are more frequent the younger children are (Achenbach et al., 1991; Fanti & Henrich, 2010), although it has to be noted that their findings are based on even younger children (2–16 years). Internalizing problems on the other hand tend to increase as children mature (Achenbach et al., 1991; Kovacs & Devlin, 1998). This could be an important factor why there are more internalizing than externalizing behavior problems in our sample. The sample could already be too old to have low internalizing problem scores, since internalizing problems tend to increase with cognitive development (Kovacs & Devlin, 1998) and more self-referencing. Thus, the hypothesis of younger children having more externalizing than internalizing problems is probably only supported for even younger children. Most literature that suggests more externalizing than internalizing problems for children analyzed preschoolers up to the age of 12 years (Chen, 2010; Fanti & Henrich, 2010), which makes it likely that after this age the ratio starts to change. Support for this thesis comes from Bornstein et al. (2010) who examined a normative sample with three age groups. A steady progression could be seen where the youngest children (four years) had way more externalizing than internalizing problems compared to the middle (10 years) and oldest (14 years) age group. While the oldest group still had more externalizing than internalizing problems it had the smallest gap between the behavioral outcomes (Bornstein et al., 2010). Other research yielded similar results with the difference that in their oldest age group (10–14 years) the internalizing problems exceeded the externalizing ones already (Sternberg et al., 2006b). Our results prompt that the capabilities of mentalization and self-reflection in 10-year-olds can already be so evolved that children with traumatic histories may develop more internalizing problems than externalizing ones. Further research is needed to determine at which developmental stage internalizing problems match or even surpass the ratio of externalizing problems in high-risk populations. It would also be interesting to know if certain types of maltreatment lead to an earlier onset of internalizing problems than others. It still remains unclear why there are more internalizing than externalizing problems throughout this sample. Existing research suggests that it is still up to debate which maltreatments and traumata lead to either one of the behavioral problems, but rather that any maltreatment can predict both outcomes in children (Muniz et al., 2019; Tarren-Sweeney, 2008). An exception seems to be sexually traumatic experiences, which strongly predict more internalizing problems (Muniz et al., 2019). Still, the degree and

occurrence of all behavioral outcomes depend on a complex interplay of psychosocial events, age, gender, genotype, early neurological development and even pre-natal conditions (Rutter, 2000). In a meta-analysis by Teicher and Samson (2016) it became uncovered that certain forms of maltreatment lead to specific changes in the brain structure but that the degree of influence these traumatic experiences have depends on the age and gender of the child. It seems likely that these neurological alterations also have an effect on internalizing and externalizing behavioral outcomes, but further research is needed to strengthen our understanding of these interactions as it is still unclear how the neurological basis translates into behavioral outcomes.

The second research question aimed at age differences regarding externalizing behavior problems only. In line with the reported increase in internalizing problems with maturation (Achenbach et al., 1991; Bornstein et al., 2010; Kovacs & Devlin, 1998) and subsequent decline in externalizing outcomes (Achenbach et al., 1991; Fanti & Henrich, 2010) it was expected that the older age group (15–19) in this sample has less externalizing behavioral problems than the younger age group. This hypothesis needed to be rejected for this study, since the older group indicated even more, though insignificant, externalizing problems for themselves than the younger one. The older children scored significantly higher on the *Rule-Breaking Behavior* scale. This could have various reasons, one being that the older children are more reflective and thus find it easier to understand when they behaved delinquently or broke rules and therefore score more realistic. Another reason could be that older adolescents tend to test out boundaries more and just objectively break more rules than younger children. A third explanation could be that adolescent girls channel their externalizing problems more in delinquent and rule-breaking behaviors (Niemann & Häßler, 2014), while boys are not restrained via gender norms to express open aggressive behaviors. Girls did indeed score higher on the *Rule-Breaking Behavior* scale ($t(117) = -2.79, p = .006$) than boys. Thus, this could explain the significant discrepancy in rule-breaking behaviors between children and adolescents. Interestingly, the caregivers did not indicate any differences between the age groups concerning rule-breaking behavior but they scored significantly higher for the younger group on the *Aggressive Behavior* scale ($t(101) = -2.96, p = .004$). They also generally indicated more externalizing problems for the younger age group ($t(101) = -1.68, p = .096$). Even though this trend is not significant, it is remarkable that caregivers had an opposing assessment of the children. The caregivers' evaluation, in trend, supports the notion that externalizing behavior problems tend to decrease with age (Achenbach et al, 1991). They also indicated minimal stronger internalizing problems for the

older age group but this result was highly insignificant. Comparing the self- and other-assessed scores, the explanation of older adolescents being more mature, reflective and maybe also more honest with themselves than children and younger adolescents becomes more likely.

Age Differences in Cross-Informant Correspondence

Mean cross-informant correlations for emotional and behavioral problems between children and some kind of caregiver (teacher, caretaker, parent) have been found to be between $r = .20-.40$ (Achenbach et al., 1987; Achenbach, 2006; De Los Reyes et al., 2015; De Los Reyes, Kazdin, 2005). These correlations include both internalizing and externalizing behavior problems. The data on possible differences in self-other agreement mediated by age is quite ambiguous. There are some studies suggesting higher agreement when children are younger (Achenbach et al., 1987; De Los Reyes & Kazdin, 2005; van der Ende & Verhulst, 2005; Verhulst & van der Ende, 1992). Others found no difference regarding age (Choudhury et al., 2003; De Los Reyes et al., 2015; Kolko & Kazdin, 1993), while yet others found higher concordance between older children and caregivers (Grills & Ollendick, 2003; Lauth et al., 2010; Romano et al., 2018). An undirected hypothesis (H4) was formed accordingly, sole a difference between the two age groups was assumed.

Internalizing Problems. The comparison of the correspondence regarding internalizing behaviors yielded a highly non-significant result, meaning there was no difference in agreement between younger and older children with their caregivers. Caregivers indicated more internalizing problems than children across all ages. This is in line with previous works reporting clinically-referred children tend to score lower than their caregivers on internalizing problems (Becker et al., 2004; Berg-Nielsen et al., 2003; Mokros et al., 1987). Generally, the correlations for both groups with their caregivers were around Pearson $r = .20$, which is regarded as fairly low correspondence (De los Reyes & Ohannessian, 2016). The results are still within the expected range of cross-informant correlations. Especially when considering, that agreement has repeatedly been found to be weaker for internalizing problems (Becker et al., 2004; De Los Reyes et al., 2015; Kolko & Kazdin, 1993; Strijker et al., 2011). Still, it is interesting to see that younger children had a lot more correspondence with their caregivers regarding the *Withdrawn/Depressed* scale (young: $r(42) = .402$ vs. old: $r(51) = .07$), while the opposite was found for the *Anxious/Depressed* scale ($r(42) = .078$ vs. $r(51) = .392$). As both are clearly associated with the *Internalizing Behavior* scale, it becomes obvious why no difference for the broadband scale was found. Since such a clear division for these two internalizing syndrome scales regarding correspondence was not expected, there is

no ready-made explanation for it. One possible explanation could be that the interaction style between caregivers and the two age groups is different. Caregivers, for example, could feel more inclined towards intervening when a younger child withdraws and locks him/herself in the room than when an adolescent does it. Being withdrawn or at least withdrawn from adults and caregivers could also be seen as a common ‘symptom’ of adolescence and thus this behavior might be somewhat expected by the caregivers. If that hypothetical scenario holds, the younger children get confronted with their behavior through the intervention and therefore reflect it better, while the older children may not reflect as much on their withdrawal. It is also unknown to the caregivers what adolescents are doing during the time in which they are withdrawn, thus eventual leading to greater discrepancies in judgement.

Interaction and maturation could also play a role in agreement concerning anxious/depressed behaviors. Older children might be more open to admit being anxious and therefore seek help or support from their caregivers, which in turn helps the caregivers get a more precise assessment. A conversation about anxieties would also make them more salient for the adolescent and thus could create a more realistic judgement. Another aspect is that older adolescents are more likely to receive some kind of treatment for anxiety compared to younger ones (Merikangas et al., 2010), implying not only an increase in anxiety with age but also communicating these sufferings, which makes these problems more accessible to others. The younger children, on the other hand, just hit puberty and it could be that in this developmental stage they are unwilling to admit any anxieties to themselves, let alone others, thereby causing a greater discrepancy in the correspondence with others. All of these explanations are speculative and further research is needed to approve or disapprove these theses.

Externalizing Problems. Similar to the internalizing problems, caregivers indicated more externalizing problem behaviors than children across all ages. This endorses previous findings indicating higher other-assessed externalizing problems than self-assessed problems in clinical samples (Becker et al., 2004; Mokros et al, 1987; Sawyer et al., 1992). Although this is not a clinical sample per se, it is definitely a high-risk population, which could be considered clinical and is therefore better compared with a clinical population than with the general population. The trend, though insignificant, that the self-other correspondence is higher for externalizing than for internalizing behaviors across all age groups is in line with prior research (Becker et al., 2004; De Los Reyes et al., 2015; Grills & Ollendick, 2002). The self-other assessment regarding externalizing behavior outcomes did not differ significantly between the two age groups, even though the older age group correlated almost twice as

strong with their caregivers than the younger group. The two clearly attributable externalizing syndrome scales concerning rule breaking/delinquent behavior and aggressive behavior showed even greater correlative differences for the two age groups than the broadband *Externalizing Problems* scale. The correlations on the *Rule Breaking Behavior* scale were fairly low for the younger group ($r(42) = .187, p = .224$) and moderate for the older age group ($r(51) = .405, p < .01$), the latter being more than twice the size of the younger group. The same counts for the *Aggressive Behavior* scale, for which the discrepancy was even bigger ($r(42) = .189, p = .219$ vs. $r(51) = .468, p < .01$). Looking at the mean difference for externalizing behavior problems between caregivers and the younger age group ($M_d = 8.14, SD_d = 15.86$), it becomes evident that this is the greatest discrepancy between all caregiver-other pairs for one of the three broadband scales. Keeping in mind these mean values are already standardized scores, this difference is quite enormous. It also explains why there is so little correspondence between caregivers and the younger children, since there is too much divergence for the majority of younger children. Caregivers' judgements suggest that younger children display more externalizing problems than older children, which would be in line with existing research that these behavioral problems tend to decrease with age (Achenbach et al., 1991). On the other hand, younger children assessed less externalizing behavior outcomes for themselves compared to the self-assessment of the older children. Again, it seems plausible that the factor age and therefore cognitive maturation and ability to reflect play a vital role in the emergence of these results. The higher self-assessed scores by the older children could stem from a more objective judgement on the self, due to increased self-reflection. Whereas the lower scores of the younger children could be caused by not having these capabilities. It could be, for example, that the onset of puberty leads to so many changes and rapid developments within the child that it has no capacities to reflect that much on externalizing problems. Another possibility is that the social desirability is stronger in younger children compared to older children. At least for young adults it was found that they score higher on social desirability compared to older adults (Perlini & Lippe, 2006), so it could be that the same principle applies to children. This could mean that younger children answer, to a greater extent, in a socially desirable way, which camouflages the extend of externalizing problems.

Total Problems. Neither an age difference in self-other correspondence was found for internalizing nor for externalizing behavior outcomes. Consequently, there is also no significant difference for the *Total Problems* scale. Again, the older age group has slightly higher cross-informant agreement than the younger group. It should be noted that the older

age group also has higher correlations with their caregivers (compared to the younger group) for the majority of the syndrome scales. The difference in assessment for the syndrome scales in which the younger children and caregivers have higher agreement (*Withdrawn/Depressed*, *Social Problems*, *Attention Problems*) is only minimal, except for the *Withdrawn/Depressed* scale (see Table 3). The results of this current research add to the existing literature that have not found any significant differences in self-other agreement between age groups (Choudhury et al., 2003; De Los Reyes et al., 2015; Kolko & Kazdin, 1993). Cautiously one could discuss a possible trend, even though the analysis did not reveal any significant differences, that goes in the direction of greater cross-informant correspondence the older children get, which would be in line with the argumentation of cognitive maturation. Cognitive development leads to more self-referencing, better understanding of contexts and thus, in the case of adverse childhood experiences, more internalizing behavior outcomes (Kovacs & Devlin, 1998) but, as it seems, also to greater self-awareness and reflection.

Total Sample: Cross-Informant Correspondence

The overall cross-informant correspondence was expected to be low to moderate since self- other agreement in clinical settings tends to be rather weak (Achenbach et al., 1987; Achenbach, 2006; De Los Reyes & Kazdin, 2005; De Los Reyes et al., 2015; De Los Reyes & Ohannessian, 2016; Jensen et al., 1999). Very low correlations were found (e.g. younger group/caregivers on *Thought Problems*: $r(42) = .032, p = .836$) but also higher than expected correlations (e.g. females/caregivers on *Social Problems*: $r(27) = .686, p < .01$). The majority of the cross-informant concordances were in the expected range between Pearson $r = .20 - .40$ (De Los Reyes & Ohannessian, 2016). Externalizing behavior outcomes are more visible and thus better accessible for others, which is why higher agreement for these than for internalizing behaviors was found regularly (Becker et al., 2004; De Los Reyes et al., 2015; Grills & Ollendick, 2002; Kolko & Kazdin, 1993; Strijker et al., 2011). It was expected to replicate prior findings, which quite surprisingly did not happen. Looking at the raw correlations of the self-other agreement it seems like there is more correspondence for externalizing problems than for internalizing ones, although the analysis yielded a non-significant result. Since the result suggests no difference in concordances for the different problems, one can only cautiously speak about a possible trend that goes in the envisioned direction. Therefore, the question remains why the prior findings could not be replicated. There are some possible explanations for this non-significant difference. One reason could be that, as a result of the current zeitgeist, children speak more openly about their internalizing problems than they used to. This would lead to more understanding on the caregiver side and

also decrease the discrepancy in judgement. It is the author's opinion, that in the western world children became more vocal in recent years and nowadays, from a societal point of view, it is a lot more encouraged to speak about personal problems and hardships than it has been just some years ago. Thus, the younger generations could feel less reason to be ashamed of their experiences and struggles and communicate them more openly. The idea would be in line with the current findings that the mean difference between caregivers and children for internalizing problems ($M_d = 3.01$, $SD_d = 14.03$, $p = .037$) was on average smaller than for the externalizing problems ($M_d = 5.51$, $SD_d = 13.45$, $p < .001$). No research concerning a possible generational difference in communication of emotional states and/or past experiences has been found, which leaves this thesis a mere suggestion worth exploring.

The indication of more internalizing than externalizing problems across all ages and genders (see Table 3) could also have its roots in a more open communication of emotional problems. Following the logic of Achenbach and Rescorla (2000), who said verbal ability, and therefore communication of needs, decreases the occurrence of externalizing problems in girls compared to boys, the same could count for all children in the case of more frequent and open communication of emotional problems. Put differently, clearer communication of internalizing problems, like anxiety or depressive symptoms, could decrease the need to express experiences and sufferings in externalizing problems.

Since the internalizing problems are more prominent in the children's but also in the caregivers' assessment, another explanation for the insignificant difference in correspondences could be that the internalizing outcomes are foregrounded in the children's experiences, so the externalizing problems are somewhat marginalized, i.e. out of focus in the children's perception. The scope of their internalizing problems could occupy them in a way that it overshadows the extent of their externalizing ones and thus creates a correspondence for externalizing problems that is less strong than expected.

A third suggestion to explain the result is, there might be a gender difference for the agreement on internalizing behaviors. Females in this sample have stronger correspondence with their caregivers for internalizing than for externalizing problems, while boys and caregivers have hardly any correspondence. This could skew the overall internalizing problem agreement towards a non-significant difference. Further analysis, including the boys only, revealed a significant difference between internalizing and externalizing cross-informant correspondence ($z = 1.68$, $p = .047$, one-sided), with stronger agreement for externalizing problems. This finding supports the hypothesis of higher cross-informant concordance for externalizing problems at least for males. This discrepancy in self-other

agreement between the different types of problematic outcomes in boys but not in girls indicates that there might be a more general gender effect.

Gender Differences in Cross-Informant Correspondence

Based on existing research it was expected that boys from this sample indicate significantly less externalizing outcomes than their caregivers. Becker et al. (2004) found that caregivers in clinical samples indicate more problematic behaviors for the children than the children themselves do. As argued already, this sample should rather be regarded as a clinical than a community sample, since it is from a high-risk population of which the majority is likely to have experienced some kind of trauma (Clausen et al., 1998; McAuley & Trew, 2000) and consequential behavioral and emotional problems. The second assumption for this hypothesis was derived from Bruchmüller et al. (2012) who found healthcare professionals to be influenced by gender stereotypes in the diagnosis of ADHD. ADHD symptoms are mainly externalizing outcomes and thus it suggests itself that an overdiagnosis of externalizing problems for males is likely to occur in other contexts as well. Indeed, a highly significant discrepancy in self-other assessment for externalizing behaviors was found for the males in this sample, supporting the argumentation.

Caregivers also indicated significantly more internalizing behavior outcomes than boys did, which supports the understanding that children in clinical settings state less problems than their caregivers (Becker et al., 2004; Berg-Nielsen et al., 2003; Mokros et al., 1987). The discrepancy for externalizing problems was greater ($M_D = 6.57$, $p < .01$) than for internalizing ones ($M_D = 4.43$, $p < .05$), but still the interpretation of this finding is ambiguous. It remains unclear, if the overdiagnosis of externalizing problems is due to gender, gender stereotypes or other parameters, exemplarily reduced self-awareness of the boys, leading to the discrepancy.

The analysis of the female self-other correspondence regarding externalizing outcomes revealed no significant discrepancy, suggesting there might be a difference between the genders for self-other assessment. Girls indicated more externalizing problems for themselves compared to boys, which is in line with research suggesting girls in vulnerable populations are at higher risk for both externalizing and internalizing outcomes than boys (Handwerk et al., 2006; Sternberg et al., 1993). Caregivers indicated insignificantly higher rates of externalizing outcomes for the girls and these scores even surpassed the other-assessed scores for the boys, although the discrepancy between females and caregivers was less than half the size of the one between males and caregivers. This could mean several things. The significant difference for males but not for females could be the result of the

hypothesized gender stereotypical bias in that the caregivers indicate more externalizing problems for males than the boys do not exhibit, which would support Bruchmüller's et al. (2012) reasoning. This explanation does not need to be right for this sample because it has been shown that caregivers in clinical samples indicate more problems than children do (Becker et al., 2004; Mokros et al., 1987), which is also true for the girls in our sample. Therefore, the findings could solely be another proof of this effect. The discrepancy for females and caregivers was not as great as for boys and caregivers. This supports the original hypothesis, but their non-significance in discrepancy could also be the result of the quiet small sample size of the girls ($n = 29$) in this analysis, meaning the analysis could have just lacked the power to find an effect. Considering that Bruchmüller et al. (2012) demonstrated the gender bias on hypothetical patients' symptom sheets, i.e. not on real patients, it remains unclear if the caregivers, who work with the children on a daily basis, were affected by a gender bias and thus answered in a stereotypical way or if no such bias influenced them.

Comparing the genders, it is noticeable that the male group had greater discrepancy in correspondence with their caregivers for all three broadband scales. Adding the two age groups into comparison also shows that the male only group self-indicated the lowest scores of all groups on the *Internalizing*, *Externalizing* and *Total Problems* scale. This finding could be another indication that females in high-risk populations have more overall problems than males (Handwerk et al., 2006; Sternberg et al., 1993) but as other research did not find these differences (e.g. Achenbach et al., 1991; Niemann & Häbeler, 2014) another explanation suggests itself. The greater discrepancy between males and caregivers could have its roots in the males' self-assessment. Traumatized male youths might underrate their emotional and behavioral problems. This may be caused by more self-denial about their own problems and/or less self-awareness than females. A possible reason for this could be the influence of gender roles, which seem to be more prominent in low SES populations (DeVol et al., 2011), from which the majority of the children most likely stem. Boys might feel less inclined to admit to problems and struggles as these could be interpreted as weakness (Olliffe & Phillips, 2008). Being strong and healthy is a prominent view on masculinity, therefore it is, in that framework, logical to interpret diseases, injuries and mental health problems as not being masculine enough (Bartlett, 2005). Thus, implicit gender roles could hinder boys to acknowledge problems to themselves and others. Research examining childhood sexual abuse (CSA) found that boys have greater problems disclosing their experiences (Hébert et al., 2009), likely because of fear of homophobic stigma (Collin-Vézina et al., 2015). Other research has shown that the majority of men do not disclose emotional problems to health

care professionals (Bartlett, 2005; Oliffe & Phillips, 2008), even if they have suicidal ideation (Wide et al., 2011). This strongly suggests that male socialization prohibits men to tolerate their mental health problems and might even bias their perception to not notice certain behavior outcomes. The unwillingness or inability to disclose could mean that boys separate, i.e. dissociate their problems and consequently might not be as aware of their behaviors compared to girls. The current findings at least suggest this possibility.

The last hypothesis (H7) aimed at finding greater discrepancy in self-other judgement concerning internalizing problems for males than females, which occurred. One factor for this result was the significant difference in scores for boys and their caregivers, with caregivers indicating more problems than boys. This is in line with findings that clinically-referred children indicate less problems than their caregivers (Becker et al, 2004; Berg-Nielsen et al., 2003). The other factor was that there was hardly any difference between girls and caregivers, leading to a significant discrepancy in correspondence for the genders. Subsequently, the self-other agreement was much higher for girls than for boys. Higher correlation between self-reports and caregivers scores for girls have been found for all types of behaviors (Becker et al., 2004) and for the *Total Problems* scale of the CBCL only (Kolko & Kazdin, 1993).

In a community sample it was found that the discrepancy for internalizing problems was greater between girls and parents (Verhulst & van der Ende, 1992). The discrepancy found by Verhulst & van der Ende (1992) stems from higher scores indicated by the girls, which would be in line with research that found children from the general population tend to score higher than their parents (van der Ende & Verhulst, 2005) and results that females, especially at-risk, score higher on internalizing problems than males (Achenbach et al., 1991; Milette-Winfrey & Mueller, 2018; Moylan et al., 2010; Niemann & Häbeler, 2014; Schmid et al., 2008). The current analysis supports these latter findings as girls had significantly higher scores on internalizing outcomes than boys. The current correlation of scores is, in the context of self-other assessment of internalizing problems, extremely high ($r(29) = .581, p < .01$). The finding stands somewhat in opposition with previous research finding low to moderate agreement, rarely exceeding $r = .40$ (Achenbach, 2006; De Los Reyes & Kazdin, 2005; De Los Reyes & Ohannessian, 2016) and most often with stronger correlations for externalizing problems (Achenbach et al., 1987; De los Reyes et al., 2015), suggesting the agreement for internalizing problems is generally around $r = .20 - .30$. Furthermore, the current finding opposes the notion that children from clinical populations indicate less problems than others, since girls in this sample had slightly higher internalizing scores. It is

up to debate if these new results indicate a gender difference for self-other agreement or if the result is merely an artefact of the current study.

Noticeable are various findings concerning gender. Girls indicated more internalizing, externalizing and total problems than boys for themselves. Caregivers scored slightly higher for girls as well, prompting girls in this sample have more overall problems than boys. This contributes to existing findings for higher scores for girls in high-risk populations (Handwerk et al., 2006; Sternberg et al., 1993). Another finding worth examining is the fact that girls and caregivers had greater agreement on all, but one (*Rule Breaking Behavior*), syndrome scales and consequently on the three broadband scales. A more general implication could possibly be drawn from this, that there is greater caregiver-female agreement at least for girls in foster care. This is an important finding because greater disagreement leads to worse treatment outcomes and increases the chance for the development of more severe psychopathologies (Humphreys et al., 2017; Romano et al., 2018), indicating boys in this sample are at greater risk. As Achenbach and Rescorla (2000) showed girls have greater verbal ability in very early age (18-35 months), they learn to communicate effectively early in life, which helps them form good relationships earlier than boys (Chen, 2010). This experience could lead to increased sharing of problems, increasing the agreement of those with others. Another argument for this assertion is that females and males are socialized differently. While boys are encouraged to be *strong* and *independent* (Olliffe & Phillips, 2008; Wide et al., 2011), girls are socialized to be *nurturing* and *caring* (Ellemers, 2018; Wide et al., 2011), therefore also caring about others' and their own health, which among others is shown through the fact that the majority of people working in healthcare are women. "Self-reports of an individual's problems reflect the individual's perceptions, judgements and tolerance of his/her own behavior, feelings, thoughts and fantasies across different situations" (Verhulst & van der Ende, 1992, p. 1019) and the perception of the self, regarding (mental) health issues, as well as the tolerance towards own problems and distress might be higher for females than males, leading to higher agreement for females.

Limitations

Limitations Due to the Usage of Diagnostic Instruments

The CBCL was repeatedly shown to be a valid and reliable diagnostic instrument (Achenbach & Rescorla, 2000; Döpfner et al., 2014; Esser et al., 2017). It is used universally and is regarded as the standard tool to assess internalizing and externalizing behavior problems. For this analysis the standardized syndrome scores of the CBCL were used instead of the raw data as there was some missing data. The usage of standardized scores has the

downside of potentially reducing variability (Achenbach, 1991; Saigh et al., 2002). It is still unlikely that this procedure has influenced the results in a meaningful way.

Another limitation is the exclusive focus on the CBCL and not including other diagnostic instruments, especially concerning PTSD. The combined data of the CBCL and knowing if a child has a PTSD or not would have given more insights and could have contributed to the existing literature on that specific interaction (e.g. Saigh et al., 2002; Wolfe et al., 1994). It could have strengthened the understanding of the interplay of externalizing and internalizing behavior outcomes with PTSD and thus make it easier to assess PTSD by analyzing these problems.

Limitations due to the Nature of the Sample

Various limitations are given due to the nature of the sample, which could have an influence on the assessment of the true prevalence of behavioral problems for this specific population. One aspect is that due to ethical standards participation was voluntarily and several children did not want to participate. This self-selection shapes the sample in a certain way and it is not given that the children who did participate represent the true population of children in foster care. It could be hypothesized that the children who refused to partake are different, in certain aspects, to the ones that did. For example, it seems possible that children with stronger mental health problems and impairments rather not participated, which would skew the results in a significant way. Another explanation for the unwillingness to participate could be a weaker need to act socially desirable, by implication meaning that children who did might have scored in a more socially desirable way. Both explanations would lead to a lower degree of psychopathology but could also influence the self-other assessment.

The second limitation concerning the sample is the possibility that children trivialized their experiences and consequently their behavior problems. Trivialization seems likely as the majority of children from this high-risk population have suffered maltreatments (e.g. Oswald et al., 2010), and it is more comfortable to downplay experiences and sufferings than to engage in them. This again would have the effect of underrating the psychopathology.

A third limitation regarding the children could be low cognitive and intellectual abilities. Traumatic experiences often lead to worsened cognition (Aas et al., 2012) and therefore the ability to self-reflect and mentalize could be impaired in some children, leading to skewed self-assessments.

One limitation concerns the caregivers. It was not recorded how long caregivers already worked with each child. In few cases the names of the caregivers were missing and for those cases it was unknown if the caregivers who filled out the questionnaire were the key

carers. Both of these facts leave the possibility that the caregivers did not know the child well enough, which could lead to either over- or underreporting of symptoms.

Another very important limitation is the size of the sample. The sample included only $N = 119$ usable cases, which already is not a huge sample. Some of these cases still had missing data and the caregivers also often failed to indicate scores for all questions, leaving $n = 97$ cases for the analysis of self-other correspondence. Only $n = 29$ of these were females, which implicates too little test power. Eventual meaningful differences in scores, i.e. real effects, (e.g. girls/caregivers discrepancy for *Externalizing Problems*) did not reach significance and thus interpretation stayed ambiguous. Feasible effects might not have been found because of insufficient power. On the other hand, low test power increases the chance that found effects are based on measurement error. Low test power was especially an issue for the analysis of subsamples, mainly the female only since it was the smallest group. For the present study a high rate of type-II-errors is likely. An overall bigger sample and consequently greater power would have given the findings more validity.

Implications and Outlook

Implications for Future Research

The present study contributes to existing literature in the context of foster care, especially since no comparable study in the Austrian population has been done yet, to the knowledge of the Author. The sample size was fairly small, therefore further research including more institutionalized foster care homes in Austria would be beneficial to strengthen the current findings. This study is embedded into broader ongoing research and in a follow-up study explicit assessment of PTSD diagnosis could be done to analyze and strengthen the understanding of the interplay of PTSD with internalizing and externalizing behavior outcomes.

Further research could look more detailed into at which developmental point internalizing problems start to emerge and when they become prominent. Additionally, it would be interesting to see if certain types of maltreatments and traumata lead to an earlier onset of internalizing problems compared to other traumatic experiences. These findings could be helpful in clinical practice, since better tailored treatment could be offered. Similarly, neuropsychological research should explore if certain alterations of brain structures lead to either more internalizing or externalizing or both behavior outcomes. More knowledge in this domain could improve and fasten the diagnostic process in the long run.

Another question that arose is if there are age and gender differences in socially desirable answering behaviors which influence the self-assessment and subsequently self-

other correspondence. Differences should be known of and consequently kept in mind, which is why research is indicated. Furthermore, it is suggested that some classical diagnostic instruments do not capture male's mental health problems well enough (in the context of depression: Oliffe & Phillips, 2008; Wide et al., 2011). The current findings prompt this possibility for the CBCL as well, with boys scoring lower than girls on all domains. This could indicate fine-tuning of the CBCL for males and/or developing diagnostic tools that are specifically suited for males to capture the full scope of male psychopathology.

Lastly, the issue of generational differences in communication of emotional concerns and past sufferings has been raised. This is an interesting thought worth researching about but not as pressing as the others, since the hypothesis is that younger generations are more open to share their emotions and experiences, which would be beneficial for both diagnostic assessment and following treatment.

Clinical Implications

The present findings clearly indicate a higher prevalence of internalizing behavior outcomes. This is a very important finding, since it is suggested that the diagnostic and therapeutic focus lies on externalizing problems (Batstra et al., 2014; Milette-Winfrey & Mueller, 2018). It is important to give patients the best suited treatment possible to relieve the most suffering feasible. This is only achievable with the right focus as it was shown that therapy specifically targeting internalizing or externalizing problems had twice the effect size on the problems it intended to focus (Weisz et al, 1995). Present findings are a signal to health care professionals to keep the prominence of internalizing problems in mind.

Another aspect for health care professionals to look at is the likely possibility of differences in expression of psychopathology for the genders. It is unlikely that boys suffer less and have less psychopathology than girls, which is why therapists should take gender norms, that potentially constrain males in expressing their problems, into consideration. Otherwise, it is possible that many boys “fly under the radar” and do not get the treatment they need.

Looking at the overall high scores of externalizing and internalizing problems it is very plausible that many children suffer from various forms of maltreatment. As outlined earlier both behavior outcomes can be symptoms of PTSD in children and hence it is likely that the majority of children in this population are highly traumatized. The national network for trauma therapy (Österreichisches Netzwerk für Traumatherapie [ÖNT]) in Austria (ÖNT; 2020) lists only 27 therapists in Niederösterreich that are specialized in trauma therapy. This number is far too little given the amount of indicated traumata in this sample alone.

Considering the amount of people in the general population in need of a therapy because of traumatic experiences with and without a PTSD, plus the incoming refugees who, in many cases, suffered a substantial amount of traumatic experiences, it becomes clear that the health care system needs to adapt fast to cater for all the people in need. A focus on further trauma-therapeutic training for therapists and those who want to become therapists should be implemented.

Conclusion

The present study gives new insights into the magnitude of behavioral problems in foster care children in Austria. It was reaffirmed that children from this population are at high-risk and exhibit to a similar extent internalizing and externalizing behavior outcomes reported in studies from around the world (e.g. Greeson et al., 2011; Niemann & Häßler, 2014; Vanschoonlandt et al., 2013), which is an indication of having had multiple traumatic experiences (e.g. Evans et al., 2008; Oswald et al., 2010). Some gender related differences became uncovered as well. It was found that girls seem to be at greater risk for emotional and behavioral problems compared to boys but also there is greater concordance between girls' and caregivers' judgement regarding externalizing and internalizing problems. This signals an increased risk for boys to become therapeutic non-responders and to develop further psychopathologies (e.g. Romano et al., 2018). This research contributes to the existing but still ambiguous findings regarding children's and adolescents' mental health problems in institutionalized care. Both, agreement and disagreement with prior literature was found, indicating the need for further research in this area to improve assessment and dealings with children in care.

References

- Aas, M., Navari, S., Gibbs, A., Mondelli, V., Fisher, H. L., Morgan, C., Morgan, K., MacCabe, J., Reichenberg, A., Zanelli, J., Fearon, P., Jones, P. B., Murray, R. M., Pariante, C. M., & Dazzan, P. (2012). Is there a link between childhood trauma, cognition, and amygdala and hippocampus volume in first-episode psychosis? *Schizophrenia Research*, 137(1–3), 73–79. <https://doi.org/10.1016/j.schres.2012.01.035>
- Achenbach, T. M. (1966). The classification of children's psychiatric symptoms: a factor-analytic study. *Psychological Monographs*, 80(7), 1–37. <https://doi.org/10.1037/h0093906>
- Achenbach, T. M. (1991). Manual for the Child Behavior Check- list/4–18. *Journal of Abnormal Child Psychology*.
- Achenbach, T. M. (2006). As others see us. *Journal of the American Medical Association*, 15(2), 94–98. <https://doi.org/10.1001/jama.1904.02500040043018>
- Achenbach, T. M., & Edelbrock, C. S. (1983). Manual for the child behavior checklist and revised child behavior profile.
- Achenbach, T. M., Howell, C. T., McConaughy, S. H., & Stanger, C. (1995). Six-year predictors of problems in a national sample of children and youth: I. Cross-informant syndromes. *Journal of the American Academy of Child and Adolescent Psychiatry*, 34(3), 336–347. <https://doi.org/10.1097/00004583-199503000-00020>
- Achenbach, T. M., Howell, C. T., Quay, H. C., Conners, C. K., & Bates, J. E. (1991). National survey of problems and competencies among four- to sixteen-year-olds: Parents' reports for normative and clinical samples. *Monographs of the Society for Research in Child Development*, 56(3). <https://doi.org/10.4135/9781452286143.n496>
- Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, 101(2), 213–232. <https://doi.org/10.1037/0033-2909.101.2.213>

- Achenbach, T. M., & Rescorla, L. A. (2000). Validity of ASEBA Preschool Scales. In *Manual for the ASEBA Preschool Forms and Profiles*.
- Arbeitsmarktservice (AMS) Österreich, Abt. Arbeitsmarktforschung und Berufsinformation. (2020). *Arbeitsmarktlage 2019*.
- Bartlett, F. (2005). Men's Mental Health: A silent crisis. *Visions Journal*, 2(5), 6–7.
- Batstra, L., Nieweg, E. H., & Hadders-Algra, M. (2014). Exploring five common assumptions on Attention Deficit Hyperactivity Disorder. *Acta Paediatrica, International Journal of Paediatrics*, 103(7), 696–700.
<https://doi.org/10.1111/apa.12642>
- Becker, A., Hagenberg, N., Roessner, V., Woerner, W., & Rothenberger, A. (2004). Evaluation of the self-reported SDQ in a clinical setting: Do self-reports tell us more than ratings by adult informants? *European Child and Adolescent Psychiatry, Supplement*, 13(2), 17–24. <https://doi.org/10.1007/s00787-004-2004-4>
- Berg-Nielsen, T. S., Vika, A., & Dahl, A. A. (2003). When adolescents disagree with their mother: CBCL-YSR discrepancies related to maternal depression and adolescent self-esteem. *Child: Care, Health and Development*, 29(3), 207–213.
<https://doi.org/10.1046/j.1365-2214.2003.00332.x>
- Bodenhausen, G. V. (2005). The role of stereotypes in decision-making processes. *Medical Decision Making*, 25(1), 112–118. <https://doi.org/10.1177/0272989X04273800>
- Bodenhausen, G. V., & Wyer, R. S. (1985). Effects of stereotypes on decision making and information-processing strategies. *Journal of Personality and Social Psychology*, 48(2), 267–282. <https://doi.org/10.1037/0022-3514.48.2.267>
- Bornstein, M. H., Hahn, C. S., & Haynes, O. M. (2010). Social competence, externalizing, and internalizing behavioral adjustment from early childhood through early adolescence: Developmental cascades. *Development and Psychopathology*, 22(4), 717–735.
<https://doi.org/10.1017/S0954579410000416>

- Bruchmüller, K., Margraf, J., & Schneider, S. (2012). Is ADHD diagnosed in accord with diagnostic criteria? Overdiagnosis and influence of client gender on diagnosis. *Journal of Consulting and Clinical Psychology, 80*(1), 128–138.
<https://doi.org/10.1037/a0026582>
- Burns, B. J., Phillips, S. D., Wagner, H. R., Barth, R. P., Kolko, D. J., Campbell, Y., & Landsverk, J. (2004). Mental health need and access to mental health services by youths involved with child welfare: A national survey. *Journal of the American Academy of Child and Adolescent Psychiatry, 43*(8), 960–970.
<https://doi.org/10.1097/01.chi.0000127590.95585.65>
- Campbell, S. B. (1995). Behavior problems in preschool children: A review of recent research. *Journal of Child Psychology and Psychiatry, 36*(1), 113–149.
<https://doi.org/10.1111/j.1469-7610.1995.tb01657.x>
- Chaffin, M., Kelleher, K., & Hollenberg, J. (1996). Onset of physical abuse and neglect: psychiatric, substance abuse, and social risk factors from prospective community data. *Child Abuse and Neglect, 20*(3), 191–203. [https://doi.org/10.1016/S0145-2134\(95\)00144-1](https://doi.org/10.1016/S0145-2134(95)00144-1)
- Chen, J. J. L. (2010). Gender differences in externalising problems among preschool children: Implications for early childhood educators. *Early Child Development and Care, 180*(4), 463–474. <https://doi.org/10.1080/03004430802041011>
- Choudhury, M. S., Pimentel, S. S., & Kendall, P. C. (2003). Childhood anxiety disorders: Parent-child (dis)agreement using a structured interview for the DSM-IV. *Journal of the American Academy of Child and Adolescent Psychiatry, 42*(8), 957–964.
<https://doi.org/10.1097/01.CHI.0000046898.27264.A2>
- Clausen, J. M., Landsverk, J., Ganger, W., Chadwick, D., & Litrownik, A. (1998). Mental health problems of children in foster care. *Journal of Child and Family Studies, 7*(3), 283–296. <https://doi.org/10.1023/A:1022989411119>
- Collin-Vézina, D., De La Sablonnière-Griffin, M., Palmer, A. M., & Milne, L. (2015). A preliminary mapping of individual, relational, and social factors that impede disclosure

- of childhood sexual abuse. *Child Abuse and Neglect*, 43, 123–134.
<https://doi.org/10.1016/j.chiabu.2015.03.010>
- De Caluwé, E., Decuyper, M., & De Clercq, B. (2013). The child behavior checklist dysregulation profile predicts adolescent DSM-5 pathological personality traits 4 years later. *European Child and Adolescent Psychiatry*, 22(7), 401–411.
<https://doi.org/10.1007/s00787-013-0379-9>
- De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, 141(4), 858–900.
<https://doi.org/10.1037/a0038498>
- De Reyes, A. L., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*, 131(4), 483–509.
<https://doi.org/10.1037/0033-2909.131.4.483>
- De Los Reyes, A., & Ohannessian, C. M. C. (2016). Introduction to the special issue: Discrepancies in adolescent–parent perceptions of the family and adolescent adjustment. *Journal of Youth and Adolescence*, 45(10), 1957–1972. <https://doi.org/10.1007/s10964-016-0533-z>
- DeVol, P., Payne, R-K., Smith, T. (2011). Bridges out of Poverty. Strategies for Professionals and Communities. OtB Learning Resources.
- Dölitzsch, C., Fegert, J. M., Künster, A., Kölch, M., Schmeck, K., & Schmid, M. (2014). Mehrfachdiagnosen bei schweizer Heimjugendlichen. *Kindheit und Entwicklung*, 23(3), 140–150. <https://doi.org/10.1026/0942-5403/a000140>
- Döpfner, M., Plück, J. & Kinnen, C. (2014). *CBCL/6-18R, TRF/6-18R, YSR/11-18R. Deutsche Schulalter-Formen der Child Behavior Checklist von Thomas M. Achenbach.* Hogrefe.
- Ellemers, N. (2018). Gender stereotypes. *The SAGE Encyclopedia of Psychology and Gender*. <https://doi.org/10.4135/9781483384269.n251>

- Evans, S. E., Davies, C., & DiLillo, D. (2008). Exposure to domestic violence: A meta-analysis of child and adolescent outcomes. *Aggression and Violent Behavior, 13*(2), 131–140. <https://doi.org/10.1016/j.avb.2008.02.005>
- Esser, G., Hänsch-Oelgart, S., & Schmitz, J. (2017). TBS-TK Rezension: „CBCL/6-18R, TRF/6-18R, YSR/11-18R. Deutsche Schulalter-Formen der Child Behavior Checklist von Thomas M. Achenbach“. *Psychologische Rundschau, 69*, 144–146. <https://doi.org/https://doi.org/10.1026/0033-3042/a000394>
- Fanti, K. A., & Henrich, C. C. (2010). Trajectories of pure and co-occurring internalizing and externalizing problems from age 2 to age 12: Findings from the national institute of child health and human development study of early child care. *Developmental Psychology, 46*(5), 1159–1175. <https://doi.org/10.1037/a0020659>
- Ferdinand, R. F., Ende, J. van der, & Verhulst, F. C. (2004). Parent-adolescent disagreement regarding psychopathology in adolescents from the general population as a risk factor for adverse outcome. *Journal of Abnormal Psychology, 113*, 198–206. <http://doi.org/10.1037/0021-843X.113.2.198>
- Fisher, R. A. (1921). On the probable error of a coefficient of correlation deduced from a small sample. *Metron, 1*, 3–32.
- Gauthier-Duchesne, A., Hébert, M., & Daspe, M. È. (2017). Gender as a predictor of posttraumatic stress symptoms and externalizing behavior problems in sexually abused children. *Child Abuse and Neglect, 64*, 79–88. <https://doi.org/10.1016/j.chiabu.2016.12.008>
- Gilliom, M., & Shaw, D. S. (2004). Codevelopment of externalizing and internalizing problems in early childhood. *Development and Psychopathology, 16*(2), 313–333. <https://doi.org/10.1017/S0954579404044530>
- Greeson, J. K. P., Briggs, E. C., Kisiel, C. L., Layne, C. M., Ake, G. S., Ko, S. J., Gerrity, E. T., Steinberg, A. M., Howard, M. L., Pynoos, R. S., & Fairbank, J. A. (2011). Complex trauma and mental health in children and adolescents placed in foster care: Findings from the national child traumatic stress network. In *child welfare*.

- Grills, A. E., & Ollendick, T. H. (2002). Issues in parent-child agreement: The case of structured diagnostic interviews. *Clinical Child and Family Psychology Review*, 5(1), 57–83. <https://doi.org/10.1023/A:1014573708569>
- Handwerk, M. L., Clopton, K., Huefner, J. C., Smith, G. L., Hoff, K. E., & Lucas, C. P. (2006). Gender differences in adolescents in residential treatment. *American Journal of Orthopsychiatry*, 76(3), 312–324. <https://doi.org/10.1037/0002-9432.76.3.312>
- Hébert, M., Tourigny, M., Cyr, M., McDuff, P., & Joly, J. (2009). Prevalence of childhood sexual abuse and timing of disclosure in a representative sample of adults from Quebec. *Canadian Journal of Psychiatry*, 54(9), 631–636. <https://doi.org/10.1177/070674370905400908>
- Hemmerich, W. (2017). StatistikGuru: Korrelationen statistisch vergleichen. Retrieved from <https://statistikguru.de/rechner/korrelationen-vergleichen.html>
- Herrenkohl, E. C., Herrenkohl, R. C., Egolf, B. P., & Russo, M. J. (1998). The relationship between early maltreatment and teenage parenthood. *Journal of Adolescence*, 21(3), 291–303. <https://doi.org/10.1006/jado.1998.0154>
- Herrenkohl, T. I., Sousa, C., Tajima, E. A., Herrenkohl, R. C., & Moylan, C. A. (2008). Intersection of Child Abuse and Children's Exposure to Domestic Violence. *Trauma, Violence, and Abuse*, 9(2), 84–99. <https://doi.org/10.1177/1524838008314797>
- Hoftun, G. Bø., Romundstad, P. R., & Rygg, M. (2012). Factors associated with adolescent chronic non-specific pain, chronic multisite pain, and chronic pain with high disability: The young-HUNT study 2008. *Journal of Pain*, 13(9), 874–883. <https://doi.org/10.1016/j.jpain.2012.06.001>
- Huffhines, L., Jackson, Y., & Stone, K. J. (2020). Internalizing, Externalizing Problems and Psychiatric Hospitalizations: Examination of Maltreatment Chronicity and Coping Style in Adolescents in Foster Care. *Journal of Child and Adolescent Trauma*. <https://doi.org/10.1007/s40653-020-00305-1>
- Humphreys, K. L., Weems, C. F., & Scheeringa, M. S. (2017). The Role of Anxiety Control and Treatment Implications of Informant Agreement on Child PTSD Symptoms.

Journal of Clinical Child and Adolescent Psychology, 46(6), 903–914.

<https://doi.org/10.1080/15374416.2015.1094739>

IBM Corp. (2019). IBM SPSS Statistics for Macintosh, Version 26.0. Armonk, NY: IBM Corp

Jang, S. J. (2007). Gender differences in strain, negative emotions, and coping behaviors: A general strain theory approach. *Justice quarterly*, 24(3).

<https://doi.org/10.1080/07418820701485486>

Jensen, P. S., Rubio-Stipec, M., Canino, G., Bird, H. R., Dulcan, M. K., Schwab-Stone, M. E., & Lahey, B. B. (1999). Parent and child contributions to diagnosis of mental disorder: Are both informants always necessary? *Journal of the American Academy of Child and Adolescent Psychiatry*, 38(12), 1569–1579. <https://doi.org/10.1097/00004583-199912000-00019>

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*. <https://doi.org/10.1037/h0034747>

Kitzmann, K. M., Gaylord, N. K., Holt, A. R., & Kenny, E. D. (2003). Child witnesses to domestic violence: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, 71(2), 339–352. <https://doi.org/10.1037/0022-006X.71.2.339>

Koch, A. J., D’Mello, S. D., & Sackett, P. R. (2015). A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *Journal of Applied Psychology*, 100(1), 128–161. <https://doi.org/10.1037/a0036734>

Kolko, D. J., & Kazdin, A. E. (1993). Emotional/behavioral problems in clinic and nonclinic children: Correspondence among child, parent and teacher reports. *Journal of Child Psychology and Psychiatry*, 34(6), 991–1006. <https://doi.org/10.1111/j.1469-7610.1993.tb01103.x>

Kovacs, M., & Devlin, B. (1998). Internalizing disorders in childhood. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 39(1), 47–63. <https://doi.org/10.1017/S0021963097001765>

- Ladd, G. W., & Kochenderfer-Ladd, B. (2002). Identifying victims of peer aggression from early to middle childhood: Analysis of cross-informant data for concordance, estimation of relational adjustment, prevalence of victimization, and characteristics of identified victims. *Psychological Assessment*, 14(1), 74–96. <https://doi.org/10.1037/1040-3590.14.1.74>
- Laird, R. D., Jordan, K. Y., Dodge, K. A., Pettit, G. S., & Bates, J. E. (2001). Peer rejection in childhood, involvement with antisocial peers in early adolescence, and the development of externalizing behavior problems. *Development and Psychopathology*, 13(2), 337–354. <https://doi.org/10.1017/S0954579401002085>
- Lauth, B., Arnkelsson, G. B., Magnússon, P., Skarphéðinsson, G. Á., Ferrari, P., & Pétursson, H. (2010). Parent-youth agreement on symptoms and diagnosis: Assessment with a diagnostic interview in an adolescent inpatient clinical population. *Journal of Physiology Paris*, 104(6), 315–322. <https://doi.org/10.1016/j.jphysparis.2010.09.004>
- Luoni, C., Agosti, M., Crugnola, S., Rossi, G., & Termine, C. (2018). Psychopathology, dissociation and somatic symptoms in adolescents who were exposed to traumatic experiences. *Frontiers in Psychology*, 9(12), 1–9. <https://doi.org/10.3389/fpsyg.2018.02390>
- McAuley, C., & Trew, K. (2000). Children's adjustment over time in foster care: Cross-informant agreement, stability and placement disruption. *British Journal of Social Work*, 30(1), 91–107. <https://doi.org/10.1093/bjsw/30.1.91>
- McCann, J. B., James, A., Wilson, S., & Dunn, G. (1996). Prevalence of psychiatric disorders in young people in the care system. *British Medical Journal*, 313(7071), 1529–1530. <https://doi.org/10.1136/bmj.313.7071.1529>
- McLeer, S. V., Callaghan, M., Henry, D., & Wallen, J. (1994). Psychiatric disorders in sexually abused children. *Journal of the American Academy of Child & Adolescent Psychiatry*, 35, 313–319.
- McWey, L. M., Cui, M., & Pazdera, A. L. (2010). Changes in externalizing and internalizing problems of adolescents in foster care. *Journal of Marriage and Family*, 72(5), 1128–1140. <https://doi.org/10.1111/j.1741-3737.2010.00754.x>

- Merikangas, K. R., He, J. P., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., Benjet, C., Georgiades, K., & Swendsen, J. (2010). Lifetime prevalence of mental disorders in U.S. adolescents: Results from the national comorbidity survey replication-adolescent supplement (NCS-A). *Journal of the American Academy of Child and Adolescent Psychiatry*, 49(10), 980–989. <https://doi.org/10.1016/j.jaac.2010.05.017>
- Meyer, A. S., McWey, L. M., McKendrick, W., & Henderson, T. L. (2010). Substance using parents, foster care, and termination of parental rights: The importance of risk factors for legal outcomes. *Children and Youth Services Review*, 32(5), 639–649. <https://doi.org/10.1016/j.chidyouth.2009.12.011>
- Milette-Winfrey, M., & Mueller, C. W. (2018). Treatment-as-usual therapy targets for comorbid youth disproportionately focus on externalizing problems. *Psychological Services*, 15(1), 65–77. <https://doi.org/10.1037/ser0000162>
- Mokros, H. B., Poznanski, E., Grossman, J. A., & Freeman, L. N. (1987). A comparison of child and parent ratings of depression for normal and clinically referred children. *Journal of Child Psychology and Psychiatry*, 28(4), 613–624. <https://doi.org/10.1111/j.1469-7610.1987.tb00227.x>
- Moylan, C. A., Herrenkohl, T. I., Sousa, C., Tajima, E. A., Herrenkohl, R. C., & Russo, M. J. (2010). The effects of child abuse and exposure to domestic violence on adolescent internalizing and externalizing behavior problems. *Journal of Family Violence*, 25(1), 53–63. <https://doi.org/10.1007/s10896-009-9269-9>
- Muniz, C. N., Fox, B., Miley, L. N., Delisi, M., Cigarran, G. P., & Birnbaum, A. (2019). The effects of adverse childhood experiences on internalizing versus externalizing outcomes. *Criminal Justice and Behavior*, 46(4), 568–589. <https://doi.org/10.1177/0093854819826213>
- Niemann, K., & Häßler, F. (2014). Seelische Gesundheit von Kindern und Jugendlichen in der stationären Jugendhilfe/Heimerziehung. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 63(7), 590–606. <https://doi.org/10.13109/prkk.2014.63.7.590>
- Österreichisches Netzwerk für Traumatherapie. (2020). *TherapeutInnen*. <https://oent.at/therapeutinnen/#toggle-id-2>

- Oland, A. A., & Shaw, D. S. (2005). Pure versus co-occurring externalizing and internalizing symptoms in children: The potential role of socio-developmental milestones. *Clinical Child and Family Psychology Review*, 8(4), 247–270. <https://doi.org/10.1007/s10567-005-8808-z>
- Oliffe, J. L., & Phillips, M. J. (2008). Men, depression and masculinities: A review and recommendations. *Journal of Men's Health*, 5(3), 194–202. <https://doi.org/10.1016/j.jomh.2008.03.016>
- Oswald, S. H., Heil, K., & Goldbeck, L. (2010). History of maltreatment and mental health problems in foster children: A review of the literature. *Journal of Pediatric Psychology*, 35(5), 462–472. <https://doi.org/10.1093/jpepsy/jsp114>
- Perlini, A. H., & Lippe, C. D. E. (2006). Resourcefulness and social desirability trait ascriptions of males: An evaluation of age and sex. *Social Behavior and Personality*, 34(7), 827–836.
- Rayburn, A. D., McWey, L. M., & Cui, M. (2016). The interrelationships between trauma and internalizing symptom trajectories among adolescents in foster care. *Children and Youth Services Review*, 61, 332–336. <https://doi.org/10.1016/j.childyouth.2016.01.006>
- Rice, S. (1926). “Stereotypes”: A source of error in judging human character. *Preventing School Failure*, 51(3), 49–51.
- Romano, E., Weegar, K., Babchishin, L., & Saini, M. (2018). Cross-informant agreement on mental health outcomes in children with maltreatment histories: A systematic review. *Psychology of Violence*, 8(1), 19–30. <https://doi.org/10.1037/vio0000086>
- Saigh, P. A., Yasik, A. E., Oberfield, R. A., Halamandaris, P. V., & McHugh, M. (2002). An analysis of the internalizing and externalizing behaviors of traumatized urban youth with and without PTSD. *Journal of Abnormal Psychology*, 111(3), 462–470. <https://doi.org/10.1037/0021-843X.111.3.462>
- Sawyer, M. G., Baghurst, P., & Mathias, J. (1992). Differences between informants' reports describing emotional and behavioural problems in community and clinic-referred

- children: a research note. *Journal of Child Psychology and Psychiatry*, 33(2), 441–449.
<https://doi.org/10.1111/j.1469-7610.1992.tb00878.x>
- Scheeringa, M. S., Wright, M. J., Hunt, J. P., & Zeanah, C. H. (2006). Factors affecting the diagnosis and prediction of PTSD symptomatology in children and adolescents. *American Journal of Psychiatry*, 163(4), 644–651.
<https://doi.org/10.1176/ajp.2006.163.4.644>
- Schmid, M., Goldbeck, L., Nuetzel, J., & Fegert, J. M. (2008). Prevalence of mental disorders among adolescents in German youth welfare institutions. *Child and Adolescent Psychiatry and Mental Health*, 2, 1–8. <https://doi.org/10.1186/1753-2000-2-2>
- Shaw, M., Hodgkins, P., Caci, H., Young, S., Kahle, J., Woods, A. G., & Arnold, L. E. (2012). A systematic review and analysis of long-term outcomes in attention deficit hyperactivity disorder: Effects of treatment and non-treatment. *BMC Medicine*, 10. <https://doi.org/10.1186/1741-7015-10-99>
- Souchon, N., Coulomb-Cabagno, G., Traclet, A., & Rascle, O. (2004). Referees' decision making in handball and transgressive behaviors: Influence of stereotypes about gender of players? *Sex Roles*, 51(7–8), 445–453.
<https://doi.org/10.1023/B:SERS.0000049233.28353.f0>
- Sousa, C., Herrenkohl, T. I., Moylan, C. A., Tajima, E. A., Klika, J. B., Herrenkohl, R. C., & Russo, M. J. (2011). Longitudinal study on the effects of child abuse and children's exposure to domestic violence, parent-child attachments, and antisocial behavior in adolescence. *Journal of Interpersonal Violence*, 26(1), 111–136.
<https://doi.org/10.1177/0886260510362883>
- Sternberg, K. J., Baradaran, L. P., Abbott, C. B., Lamb, M. E., & Guterman, E. (2006a). Type of violence, age, and gender differences in the effects of family violence on children's behavior problems: A mega-analysis. *Developmental Review*, 26(1), 89–112.
<https://doi.org/10.1016/j.dr.2005.12.001>
- Sternberg, K. J., Lamb, M. E., Greenbaum, C., Cicchetti, D., Dawud, S., Cortes, R. M., Krispin, O., & Lorey, F. (1993). Effects of Domestic Violence on Children's Behaviour Problems and Depression. *Developmental Psychology*, 29(1), pp. 44–52.

- Sternberg, K. J., Lamb, M. E., Guterman, E., & Abbott, C. B. (2006b). Effects of early and later family violence on children's behavior problems and depression: A longitudinal, multi-informant perspective. *Child Abuse and Neglect*, 30(3), 283–306.
<https://doi.org/10.1016/j.chiabu.2005.10.008>
- Strijker, J., Van Oijen, S., & Knot-Dickscheit, J. (2011). Assessment of problem behaviour by foster parents and their foster children. *Child and Family Social Work*, 16(1), 93–100. <https://doi.org/10.1111/j.1365-2206.2010.00717.x>
- Tarren-Sweeney, M. (2008). Retrospective and concurrent predictors of the mental health of children in care. *Children and Youth Services Review*, 30(1), 1–25.
<https://doi.org/10.1016/j.chilyouth.2007.05.014>
- Teicher, M. H., & Samson, J. A. (2016). Annual Research Review: Enduring neurobiological effects of childhood abuse and neglect. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 57(3), 241–266. <https://doi.org/10.1111/jcpp.12507>
- Van Der Ende, J., & Verhulst, F. C. (2005). Informant, gender and age differences in ratings of adolescent problem behaviour. *European Child and Adolescent Psychiatry*, 14(3), 117–126. <https://doi.org/10.1007/s00787-005-0438-y>
- van Os, J., Marsman, A., van Dam, D., Simons, C. J. P., Alizadeh, B. Z., Bartels-Velthuis, A. A., van Beveren, N. J., Bruggeman, R., Cahn, W., de Haan, L., Delespaul, P., Meijer, C. J., Myin-Germeys, I., Kahn, R. S., Schirmbeck, F., Simons, C. J. P., van Haren, N. E., van Os, J., & van Winkel, R. (2017). Evidence That the impact of childhood trauma on IQ is substantial in controls, moderate in siblings, and absent in patients with psychotic disorder. *Schizophrenia Bulletin*, 43(2), 316–324. <https://doi.org/10.1093/schbul/sbw177>
- Vanschoonlandt, F., Vanderfaeillie, J., Van Holen, F., De Maeyer, S., & Robberechts, M. (2013). Externalizing problems in young foster children: Prevalence rates, predictors and service use. *Children and Youth Services Review*, 35(4), 716–724.
<https://doi.org/10.1016/j.chilyouth.2013.01.015>
- Verhulst, F. C., & van der Ende, J. (1992). Agreement between parents' reports and adolescents' self-reports of problem behavior. *Journal of Child Psychology and Psychiatry*, 33(6), 1011–1023. <https://doi.org/10.1111/j.1469-7610.1992.tb00922.x>

- Weisz, J. R., Weiss, B., Han, S. S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin*, 117(3), 450–468. <https://doi.org/10.1037/0033-2909.117.3.450>
- Wide, J., Mok, H., McKenna, M., & Ogrodniczuk, J. S. (2011). Effect of gender socialization on the presentation of depression among men: A pilot study. *Canadian Family Physician*, 57(2).
- Wolfe, D. A., Crooks, C. V., Lee, V., McIntyre-Smith, A., & Jaffe, P. G. (2003). The effects of children's exposure to domestic violence: A meta-analysis and critique. *Clinical Child and Family Psychology Review*, 6(3), 171–187. <https://doi.org/10.1023/A:1024910416164>
- Wolfe, D. A., Sas, L., & Wekerle, C. (1994). Factors associated with the development of posttraumatic stress disorder among child victims of sexual abuse. *Child Abuse and Neglect*, 18(1), 37–50. [https://doi.org/10.1016/0145-2134\(94\)90094-9](https://doi.org/10.1016/0145-2134(94)90094-9)
- Zaharakis, N. M., Mason, M. J., Brown, A., Moore, M., Garcia, C., Foster, R., & Richards, S. (2018). Resiliency moderates the influence of somatization on externalizing problems. *Journal of Child and Family Studies*, 27(9), 2978–2989. <https://doi.org/10.1007/s10826-018-1141-1>

List of Tables

Table 1: Cases of Missing Data for all Syndrome and Broadband Scales.....	17
Table 2: Categorical Self-Assessments in Percentage for the Three Broadband Scales.....	19
Table 3: Means & Differences of Self-Other Judgements for the Three Broadband Scales...	20
Table 4: Correlations between Children and Caregivers.....	24

List of Abbreviations

ADHD	Attention deficit hyperactivity disorder
AMS	Arbeitsmarktservice Österreich
CBCL	Child Behavior Checklist
CSA	Childhood Sexual Abuse
ÖNP	Österreichisches Netzwerk für Traumatherapie
PTSD	Post-traumatic stress disorder

Appendix

Abstract (English)

Externalizing and internalizing behavior problems in children and adolescents in institutionalized foster care in Niederösterreich, Austria were analyzed. The self-other concordance with their caregivers regarding these behaviors was assessed. The sample were $N = 119$ children and adolescents ($M_{\text{Age}} = 14.85$, range 10–19 years). Caregiver-assessed scores of the three broadband scales of the Child Behavior Checklist (CBCL), the externalizing, internalizing and total problems scale, were each found to be in the clinical range for at least a third of the children. Children and caregivers both indicated more internalizing than externalizing problems. Correlations of self-other assessments were found to be mainly in the low to moderate range (r 's = .20–.30). Especially low concordance was found between males and caregivers for all types of problem behaviors. The greater discrepancies with their caregivers resulted from the boys indicating less problematic behaviors than the girls. Caregivers generally indicated a greater extent of problems than the children. These findings imply that boys might be more constrained to admit to their emotional and behavioral problems than girls. Increased therapeutic focus for this population should be on internalizing problems as they do not get as much attention as externalizing problems in the context of child psychopathology.

Abstract (German)

Externalisierende und internalisierende Verhaltensweisen von Kindern und Jugendlichen in Pflegeeinrichtungen in Niederösterreich wurden analysiert. Die Selbst-Fremdübereinstimmung hinsichtlich dieser Verhaltensweisen mit den Betreuer*innen wurden erfasst. Das Sample bestand aus $N = 119$ Kindern und Jugendlichen ($M_{\text{Alter}} = 14.85$, 10–19 Jahre). Die Werte der Betreuer*innen auf den drei Breitbandskalen der Child Behavior Checklist (CBCL) – internalisierend, externalisierend und Gesamtauffälligkeit – waren jeweils für mehr als ein Drittel der Kinder im klinisch auffälligen Bereich. Kinder und Betreuer*innen gaben mehr internalisierende als externalisierende Verhaltensprobleme an. Die Selbst-Fremdübereinstimmungen waren niedrig bis moderat (r 's = .20–.30). Die Übereinstimmung war besonders niedrig zwischen Jungen und deren Betreuer*innen für alle Problembereiche. Die größere Diskrepanz mit den Betreuer*innen lag daran, dass Jungen weniger Probleme angaben als Mädchen. Betreuer*innen indizierten generell schwerere Verhaltensproblematiken als die Kinder selbst. Diese Ergebnisse suggerieren, dass Jungen vielleicht eingeschränkter darin sind ihre emotionalen Probleme und Verhaltensweisen anzuerkennen und einzugestehen. Stärkerer therapeutischer Fokus sollte auf internalisierende Probleme gerichtet werden, welche im Kontext von Kinderpsychopathologie momentan nicht so viel Aufmerksamkeit bekommen.