Inter-rater reliability of the Structured Interview for DSM-IV Personality (SIDP-IV) in an adolescent outpatient population

Final thesis submitted for the degree of cand. psychol.

Institute of Psychology

University of Oslo

Autumn 2020

Candidate number 46

Eirik Wixøe Svela
Acknowledgments

First of all, I would like to thank Randi Ulberg for allowing me to take part in her research. Empirical research into psychoanalytic psychotherapy is essential in ensuring that patients are given the freedom to choose among different alternatives. Your wit, warmth, wisdom have been much appreciated.

Second, I would like to thank Line Indrevoll Stänicke, who has showed great sensitivity and deftness in helping me along at just the right junctures, not unlike like the carefully staged approach recommended by Hutsebaut and colleagues (2019).

Third, I would like to extend my most heartfelt gratitude to Hans Ole Korsgaard, without whose frankly heroic efforts this thesis almost certainly would not have materialised at all¹. I have appreciated our work together.

Finally, I would like to thank Victoria for safeguarding my overall sanity, well-being, and adult attachment needs throughout this journey. I love you.

¹ p < 0.0005.
Summary

Author: Eirik Wixøe Svela

Title: Inter-rater reliability of the Structured Interview for DSM-IV Personality (SIDP-IV) in an adolescent outpatient population

Primary advisor: Hans Ole Korsgaard, MD, PhD, The Nic Waal Institute

Secondary advisor: Line Indrevoll Stänicke, cand. psychol., PhD, University of Oslo

This thesis investigates the inter-rater reliability of the Structured Interview for DSM-IV Personality (SIDP-IV), a semi-structured diagnostic interview for assessing DSM-IV personality disorders, in an outpatient population of adolescents between the ages of 16 and 18.

The 25 participants in the study were interviewed and evaluated for the presence of personality disorders by experienced clinicians, and audio recordings were made from these sessions. Based on these recordings I blindly re-coded the interviews. This data was used to calculate the inter-rater reliability in the form of the intra-class correlation coefficient (ICC) and Cohen’s kappa coefficient. These results obtained indicated high inter-rater reliability for the instrument, with an ICC coefficient for criterion scores of 0.952 and a kappa for dichotomous diagnostic assessment of 0.896.

The data analysed was collected and made available by the research group behind the First Experimental Study of Transference In Teenagers (FEST-IT), who have graciously provided supervision and support throughout my work.
Introduction

Personality disorders constitute an important diagnostic group in adult psychiatric healthcare, with an overall prevalence estimated at 12.16% of the general population (Volkert, Gablonski, & Rabung, 2018). Studies have generally found that the relative proportion of patients suffering from personality disorders increases as a function of the level or intensity of care, the highest proportion being found in inpatient psychiatric care (Kongerslev, Chanen, & Simonsen, 2015; Kongerslev, Moran, Bo, & Simonsen, 2012; Oldham, Skodol, & Bender, 2009; Zimmerman, Chelminski, & Young, 2008). Relative to non-clinical samples and symptom disorders like depression or anxiety, personality disorders are found to be more associated with decreased subjective quality of life and role function (Torgersen, 2009), increased overall psychiatric and non-psychiatric morbidity (Castle, 2019), as well as increased suicidal and parasuicidal behaviour (Paris & Zweig-Frank, 2001; Videler, Hutsebaut, Schulkens, Sobczak, & Van Alphen, 2019); personality disorders also cause great cost to society in terms of healthcare service use and productivity loss (Feenstra et al., 2012; Wittchen et al., 2011).

Nevertheless, personality disorders have historically been considered notoriously difficult to treat, and as a consequence, some researchers claim that these patients have tended to receive inadequate and/or insufficient care due to unfounded notions of the viability of providing treatment for this population. Over the last few decades, researchers from a range of theoretical orientations have demonstrated that structured treatment programmes for personality disorders are effective both in absolute terms and compared to treatment as usual, some prominent examples being Bateman and Fonagy’s Mentalisation-Based Therapy (MBT) (Bateman, Campbell, Luyten, & Fonagy, 2018; Bateman & Fonagy, 2004b, 2015; Bateman, Gunderson, & Mulder, 2015), Kernberg’s Transference-Focused Psychotherapy (TFP) (Kernberg, 2016; Kernberg, Yeomans, Clarkin, & Levy, 2008) and Linehan’s Dialectical Behavioural Therapy (DBT) (Linehan, 1993; May, Richardi, & Barth, 2016; Teti, Boggiano, & Gagliesi, 2015; Wilks, Korslund, Harned, & Linehan, 2016). Concurrently, most developed-world countries have started to provide specialised in-patient and out-patient services to patients in this group, yielding promising results.

In child and adolescent psychiatric services, on the other hand, it seems that this diagnostic category is systematically underused (Korsgaard, 2017), some commonly cited reasons being a fear of stigmatising adolescents with complex mental health issues, as well as a concern that diagnosing personality problems in adolescents might be tantamount to pathologising normative developmental processes which will eventually resolve naturally (Laurensen,
Hutsebaut, Feenstra, Van Busschbach, & Luyten, 2013). However, research has found both of these concerns to be largely unfounded. In fact, adolescents who present with symptoms consistent with personality disorders are likely to receive such a diagnosis as young adults (Bernstein, Cohen, Skodol, Bezirganian, & Brook, 1996; De Clercq & De Fruyt, 2007; De Fruyt & De Clercq, 2014) and require treatment for their personality disorder, which seems to imply that necessary treatment is routinely deferred for no other apparent reason than a vague sense of caution (Kongerslev et al., 2015), while patients run the risk of developing more severe conditions and/or comorbid conditions (Bo et al., 2020). Studies have also found that personality diagnosis in adolescence seems to hold approximately the same degree of validity and reliability as does diagnosing these disorders in young adults (Chanen et al., 2004). This would seem to refute the concern that diagnosing personality disorders in adolescents pathologises normal processes (A. M. Chanen & Kaess, 2012; Kaess, Brunner, & Chanen, 2014; Kongerslev et al., 2015; Tromp & Koot, 2009; Widiger, De Clercq, & De Fruyt, 2009).

A corollary to this state of affairs is the fact that standard instruments and procedures used to diagnose personality disorders in adults, such as the Structured Interview for DSM-IV Personality (SIDP-IV) (Pfohl, Blum, & Zimmerman, 1997), the Structured Clinical Interview for DSM-IV Axis II Disorders (SCID-II) (First, Gibbon, Spitzer, Williams, & Benjamin, 1997), and the Millon Clinical Multiaxial Inventory III (MCMI-III) (Millon & Davis, 1997) have received relatively little attention in terms of validity and reliability studies conducted with adolescent samples, both in their original forms, in translation, and in working with various specific populations. This suggests that it is unclear how well-suited these instruments are in terms of assessing personality pathology in youth populations.

The purpose of this thesis is to examine the inter-rater reliability of the SIDP-IV in a sample of outpatient adolescent patients. Our results suggest that this instrument remains highly reliable in such a populations. Taken together with previous studies which establish the validity of this instrument for diagnosing personality disorders, especially when taking a dimensional approach, our results suggest that this instrument should be considered for routine clinical usage in working with adolescent populations. In keeping with recent findings which indicate that these disorders are underdiagnosed and undertreated in adolescent populations, we are hopeful that the present study can contribute to rectify this state of affairs. In order to place our findings within a broader clinical and intellectual context, and motivate our interest in the field of personality disorders, we will first present a brief survey of the field of personality disorders,
particularly as it pertains to children and adolescents, before we go on to present and discuss the data and analyses that were conducted.

**Personality and personality pathology**
The rationale for conducting a survey of general knowledge of personality theory and personality pathology in particular is that the fields are closely related. Notions of personality pathology – what it is, how it arises, and how it may be treated – in large part issue from general notions about what personality is, how it is shaped, and how people differ randomly and systematically in terms of personality organisation.

In addition, it would be impractical and possibly misleading to present the current state of knowledge in adolescent personality pathology without acknowledging that most theoretical and treatment approaches to personality disorders in adolescents are essentially adaptations of knowledge which was originally conceived in an adult setting.

**What is personality?**
A common definition of adult personality emphasises that personality refers to psychologically meaningful characteristics – often referred to as *traits* – that are relatively stable and enduring across situations and over time, within an individual. We might also add that notions of personality also explicitly or implicitly assume that there are relatively stable and enduring differences among individuals in terms of meaningful personality characteristics. Thus, a good theory of theory ought not only to account for what causes and enables a person to be, e.g., generally relatively confident, altruistic, or anxious, but also what might cause and enable people to differ in such qualities, possibly in non-random, systematic ways, e.g., such that most people who are generally rather confident might also tend to not be very anxious; a theory that makes this assumption should also ideally be able to account for why this is so.

As we will see, other theories reject the trait approach and suggest instead that observable personality behaviour is mediated by deeper, intermediate cognitive and affective mechanisms which account for personality stability over time and across situations, but also account for systematic diachronic and cross-situational variability. A related problem is the issue of the determinants or aetiology of personality: Is personality meaningfully shaped by life experiences, such as experiences in early childhood, unique life experiences, and an individual’s phenomenological, emic view of their life, or is it in fact mostly shaped by
biogenetic forces to which our views of our self and our own agency are simply epiphenomenal? The theories we will survey have responded differently to these issues, and while none of them can claim to be the last word on these issues, they provide context for some of the conceptual questions that are also highly in the field of personality pathology.

Regardless of one’s theoretical orientation, personality analysed at the trait level is today widely acknowledged to be shaped both by nature and nurture. Over the last decades a consensus has emerged that personality traits are meaningfully and significantly shaped by genetic variation, over and above what could be accounted for by shared environment influences, unique life experiences, and statistical error/random forces. This has been demonstrated using a range of research designs, including twin/adoption studies (Torgersen et al., 2000), genome-wide association studies (GWAS) and studies investigating single-nucleotide polymorphisms (SNPs) (Sanchez-Roige, Gray, Mackillop, Chen, & Palmer, 2018). Recent metastudies estimate the heritability coefficient for genetic influence on personality to be around 0.4 (Vukasović & Bratko, 2015).

Personality traits also appear to be shaped to some extent by shared family environment during early childhood, albeit apparently to a lesser extent than what has traditionally been assumed by some influential psychological theories (Bouchard & McGue, 2003); but see Matteson et al. (2013) for a contrasting view. Nevertheless, experiences in the family environment during childhood still does appear to account for some of the variation in influencing traits, especially if such experiences are negatively valenced or traumatic.

Some studies have also pointed out that there are effects due to a person’s so-called non-shared environment, another term for what we might call unique life experiences. We should note that current research designs make it difficult to differentiate the effect of a person’s non-shared environment from statistical error, due to the fact that this component of the variance is what ‘remains’ once genetic and shared environmental influences have been accounted for. Nevertheless, it seems plausible that extreme events, perhaps especially potentially traumatic ones, could influence a person’s personality traits in ways that are not accounted for by biological factors or the person’s upbringing.

In the following we will discuss two contrasting approaches to personality theory. The first theory we will discuss is the trait-theoretical or five-factor model, currently most associated with the work of Costa and McCrae (2003). Second, while arguably being out of vogue in mainstream academic psychology, we will discuss the psychoanalytic perspective, which
continues to influence some important approaches to normal and pathological personality and constitutes a perspective that is, in my opinion, closer to the ‘facts on the ground’ of clinical work with normal and disordered personality than trait-based approaches.

**Trait psychology/Five-Factor Model**

The trait model of personality psychology traces its historical roots back to the nascent personality psychology of the 1930s and 1940s and the influential work of writers like Cattell (1961) and Allport (1937). The model relies on two methodological assumptions, the statistical procedure of factor analysis, the aim of which is reduce complex data to higher-order, presumptively latent structures, and a form of natural language analysis that assumes that there is a direct correspondence between natural language terms and mental/psychological states, traits, and processes.

Thus, the trait approach to personality psychology contends that natural-language descriptive terms yield an exhaustive inventory of personality characteristics available to the human mind, and that by subjecting this data to a factor analysis, the superordinate structure of the human personality emerges. While it does not necessarily follow from these assumptions, trait approaches to personality psychology have generally tended to take what we might term a biologistic view of personality, according to which personality is materially, sometimes apparently almost exclusively, shaped by genetic/biological factors.

Which factors are deemed salient and how they interact vary as a function of what analyses are conducted and the data to be analysed. Over the years, several influential psychological writers have presented versions of a fundamentally trait-based approach, including Cattell and his sixteen-factor model and Eysenck’s three-factor model (Jackson, Furnham, Forde, & Cotter, 2000).

In recent years, the five-factor model (FFM) of Costa and McCrae has grown influential to the point of becoming effectively synonymous with the trait-based model in general. This model, commonly referred to as the NEOAC or OCEAN model, holds that the factors neuroticism (N), extroversion (E), openness to experience (O), agreeableness (A), and conscientiousness (C), are the basic personality dimensions. Neuroticism is the quality of easily experiencing negative affect; extroversion is the quality of being outgoing and gregarious; openness to experience is the quality of being generally open-minded, receptive to new ideas, and tolerant of ambiguity;
agreeableness is the quality of being generally empathetic towards others; and conscientiousness is the quality of being meticulous and mindful of duties and obligations.

Five-factor model researchers point to an extensive and growing body of research that establishes moderate to high reliability of five-factor personality configurations in individuals over short, intermediate, and – to a smaller extent – long periods of time (Hampson & Goldberg, 2006). A similarly expansive literature has identified correlations between five-factor personality dimensions and various psychologically salient life outcomes, such as longevity, divorce, employment status, and overall health service use (Lahey, 2009), which seems to suggest predictive validity for the model. Personality researchers have studied populations in a wide array of countries and cultures, and have generally demonstrated that analysis of linguistic data supports a five-factor solution in virtually every population studied so far (Allik & McCrae, 2002).

Importantly, in and of itself, the five-factor model makes no claims or assumptions about what shapes the personality, how it develops across the lifespan, what is normal and abnormal personality, or indeed whether such a distinction is meaningful on the theory’s assumptions. The five NEOAC factors are held to describe normative personality qualities, in which extreme high or low scores are only extreme in a strictly statistical sense, meaning that personality pathology cannot be inferred from individual NEOAC configurations or dimension scores. Furthermore, the five-factor model makes no assumptions about how the factors relate to each other, so that in principle, the factors may vary independently of each other, yielding a virtually infinite array of possible personality configurations.

Over the last decades, researchers within this tradition have increasingly turned their attention to the enterprise of extending this model into a general theory of personality, seeking to explain what personality is, what forces shape it, and how it develops across the lifespan. The Five-Factor Theory of personality (FFT) claims that an individual’s basic personality traits, i.e., their individual NEOAC levels and the configuration, are materially determined by strong endogenous predispositions, presumptively with neurobiological substrates that may one day be more precisely located, studied, and characterised. Moreover, the theory holds that endogenous processes serve to solidify the influence of these predispositions over the lifespan, implying that as individuals grows older and reaches adulthood, they will make decisions and conduct their life in ways that are manifestations of endogenous predispositions, e.g., choosing a profession or a partner, rather than general, universal life experiences, e.g., attending school, going through puberty. In other words, the theory holds that as individuals grows older, they
come to more fully realise and conform to a genetic blueprint that was already in place, rather than be influenced by life events and circumstances.

External to the individual, culture and the social environment (including the family environment) affect how the individual’s endogenous predispositions are expressed in the form of concrete ambitions, attitudes, habits, social roles, and relational attitudes and behaviours, which are what the theory calls characteristic adaptations. These are characteristic in the sense that they refer to something that is unique and defining to the individual in question, and adaptive in that they contribute to literal, relational, social, and psychological survival by furnishing the individual with ways of ‘making their personality work’ in their concrete social setting (McCrae & Costa, 1999).

Furthermore, a central aspect of an individual’s characteristic adaptations is their self-concept, which refers to cognitive and affective sets or schemas informing how the individual views him-/herself. The self-concept causes slight biases of perception which serve to adaptively reinforce the individual’s conception of him-/herself and to cement existing affective, cognitive, and behavioural patterns (Costa & McCrae, 2003).

Proponents of this theoretical framework point to the accrued evidence for the five-factor model as supportive of the broader theory. They point to cross-cultural data collected in support of the five-factor model as suggesting that a five-factor personality structure is a cultural universal that is most likely biologically determined. On this view, high moderate- and long-term ipsative and rank-order continuity of NEOAC personality factors, even in the face of significant life events, suggests a strong biological foundation that is largely irresponsible to external influence.

While acknowledging the accumulating evidence for diachronic stability, external/predictive validity of NEOAC personality traits, and data which suggest cross-cultural validity, detractors of the theory point to the fact that the procedures used to derive the factors – i.e., lexical analysis and factor analysis – and the measures used to assess the personality structure in population studies – i.e., instruments like the NEO-PI-3 – fundamentally rely on the notion that people generally have insight into their own personality makeup, and that they are able and willing to report this accurately.

Moreover, the theory seems to hold that the five-factor model is something of a brute fact of an individual’s genetically given neurobiological makeup. While not necessarily a conceptual problem in and of itself, the theory does not motivate why these particular dimensions should
be salient, or whether the dimensions covary in systematic ways, and if so, what the possible psychological or neurobiological mechanisms for such systematic covariance might be.

Finally, the traits under examination in models like the five-factor model are not directly observable, and thus the model would seem vulnerable to some lines of criticism that emphasise this aspect. One objection along these lines is on epistemological grounds, in other words, is it meaningful to infer the five NEOAC dimensions based on factor analysis and lexical analysis in the first place?

Another objection points to the fact that observable behaviour within an individual has been shown to have low cross- and inter-situation reliability (Mischel & Shoda, 1995), which means that the theory appeals to a latent, abstract entity to explain concrete behaviours, a conceptual move which would seem to introduce an intermediate level between the biogenetic foundation and concrete behaviours, all the while denying that this is in fact the case.

Some personality researchers have criticised the five-factor paradigm on another line of reasoning, which is that the five-factor theory does not capture the quality of psychological interiority, i.e., the emic aspect of being a person, with subjective experiences, opinions, and emotions (McAdams, 1995).

The psychoanalytic perspective
Psychoanalysis was always intended to be not only a mode of treatment and a programme of scientific research, but also a general theory of human personality (Thomä & Kächele, 1987). Despite this stated ambition, psychoanalysis today is hardly a unified body of theory and knowledge, and it would be difficult to do justice to this conceptual complexity within the confines of this thesis. For our purposes, we will discuss the psychoanalytic view of the personality from what is arguably a contemporary consensus standpoint, informed by object relations theory and ego psychology.

The most important assumption of psychoanalysis is also perhaps what sets it most apart from other approaches, namely the assumption of psychologically significant emotional and motivational processes that take place outside of normal consciousness. Psychoanalysis holds that some psychological content, because it is offensive to our conscious sensibilities, or exposes a conflict between contrasting wishes, needs, or desires, or conflicts with the demands of the social and physical reality within which the person is situated, is repressed through
unconscious so-called defence mechanisms (Vaillant, 1994). These mechanisms take a variety of forms, some of which achieve relative peace of mind at the cost of relatively great relational disruption, distortion of perception, or other forms of unease, whereas others seem to succeed more straightforwardly and are typically found in people who have attained a greater degree of psychological maturity. Unlike trait-based approaches, a psychoanalytic view of personality would posit that an individual’s habitual use of defence mechanisms – range, maturity, and tendency to resort to more decompensated functioning when faced with acute stressors – tell us something meaningful about the personality of that individual. In keeping with the general reasoning of Shoda and Michel (1995), this is a level of analysis below concrete, observable behaviour, which is also one of the primary reasons psychoanalytic theory has historically proven so controversial.

Another assumption of psychoanalytic theory is that interactions with the individual’s primary caregivers during early life contribute to determine a complex set of relational cognitions, expectancies, attitudes, affective states and dynamics, and concrete behaviours, which together make up a mental representation of what it means to be a person with a sense of self, what it means to relate to others, and what a relational situation is. In theory-internal language, caregivers during early life are referred to as primary objects, and the theory claims that the individual internalises so-called object relations, which are implicit and explicit mental representations of relationally framed thought, affect, and behaviour. According to this theory, it is especially during situations of so-called peak affect that the interplay between the individual’s needs and desires, the response of the objects, and the emotion associated with the situation, are internalised. Some aspects of object relations are thought to be more or less conscious and accessible to rational, deliberative cognitive processes, whereas others are unconscious and can only be inferred through surface behaviour and other indirect manifestations of latent cognitive-affective content (Kihlstrom, 1987).

A person’s object experiences interact with biogenetic, constitutional factors to influence the individual’s personality. Thus, as an example, a child with a timid, introverted temperament who grows up with parents who she perceives to be dependable, responsive to her needs, and who project a sense of resilience in the face of life’s adversities may grow up to expect that her needs will be met, and that her way of meeting the world – e.g., a little cautiously and unsure – is fundamentally acceptable and worthy of a place alongside other more confident personalities. Meanwhile, a child with a similar temperament whose primary caregivers – objects – are undependable, unpredictable in affect and behaviour, and who invalidate the child’s experience
as wrong or unwarranted, may come to feel that she cannot expect much from others, and has to rely primarily on herself. She may seek to regulate this experience through retreating into a world of fantasy, or by seeking to rebel against these conditions, or through other means in ways that are also determined partially by constitutional factors and partially through her internal object relationships, which interact transactionally.

According to psychoanalytic theorists, object relations are subject to a dialectical developmental process. During early childhood, children are in what Klein (1946) called the paranoid-schizoid position, which is a configuration of object relations in which representations of the object are unintegrated. The object, e.g., the mother, is not experienced as a human person, but rather as so-called part-objects, e.g., breasts, face, other parts of the body are experienced individually and the individual is unable to synthesise its experience of the other into a meaningful whole. The child’s representations of the world around him or her are subject to very primitive mechanisms of interpretation and cognitive control. When the child is fed, it feels fullness and satiation, and associates this state with the ‘good breast’, which comes to represent something that is protective, nourishing and fully good. On the other hand, when the child is hungry or otherwise frustrated, this is associated with the absence of this part-object representation, and so – according to this theory – it comes to associate this sensation with the so-called ‘bad breast’, an experience which feels malevolent, withholding, even persecutory. Over time, however, the child comes to realise that the two part-objects in fact represent an integrated whole, and reacts to this with sorrow and regret, entering into the so-called depressive position. Unlike Klein’s early formulations, modern-day object theorists emphasise how the human personality seems to oscillate between these positions throughout life. Under optimal conditions and given a good enough developmental trajectory, most people will occupy the depressive position most of the time, but anyone is subject to a temporary decompensation into paranoid-schizoid modes of affect, thought, and cognition under stress, or due to biogenetic vulnerabilities.

Taken together, then, one way of construing the personality according to psychoanalytic theory is as a phenotypic expression of a complex matrix of object relations, degrees of propensity to tolerate uncertainty and ambiguity or enter into splitting, projective modes of cognition, and of
densely patterned structures of defence, all of which interact materially with the individual’s biogenetic disposition\(^2\) and external life circumstances.

What is personality pathology?

Background/historical overview

The concept of personality disorder as it is currently understood essentially dates back to the 1980 publication of the third edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-III). The DSM-III is commonly acknowledged to be the first full realisation of many conceptual and clinical notions which today seem indispensable, including the diagnoses of borderline and narcissistic personality disorders as well as relatively precise criteria and thresholds for diagnosing personality disorders in general. Even so, this system is only semi-stable, as evidenced by the developments that were under deliberation for the DSM-5 and which eventually came to be implemented in the ICD-11, which perhaps points to an inherent difficulty in carving the concept of disordered personality at its joints (Gunderson, 1984) or identifying natural kinds.

In order to understand what is meant by the current diagnostic construct of personality disorder, it would be instructive to briefly turn our attention to the historical context in which it arose. Perhaps the first person to discuss a construct similar to our conception of personality disorders was Schneider (Schneider, 1923), who introduced a range of categories of so-called psychopathic personalities, by which he simply referred to disordered or pathological pathology. Schneider worked in an intellectual landscape dominated by adherents of a nascent form of biological psychiatry, led by figures like Eugen Bleuler and Emil Kraepelin (Berrios, 1993), and the followers of Sigmund Freud and the movement that would become psychoanalysis. Not unlike today’s situation (Ophir, 2015), the Kraepelinians and the Freudians had entered into a division of labour whereby severe psychopathology, such as psychotic illnesses and bipolar disorder, largely fell under the remit of biological psychiatry, whereas milder, so-called neurotic disorders were of primary interest to the Freudians.

Members of both camps seemed to assume the general validity of conceiving of mental disorders as analogous to somatic disorders, i.e., in the sense that the patient suffers from an abnormal condition which produces symptoms, and which is treatable in the sense that the

patient is fully restored to normal health. This makes the dual assumptions that a) the symptom does not in fact issue from qualities of the personality of the patient and that b) the underlying personality itself remains unimpaired despite the presence of a symptom disorder.

The Kraepelinians, while allowing that personality seemed to vary among individuals, probably due to genetic and biological factors, generally did not attend to this issue directly. Early psychoanalysts, on the other hand, did have a general formulation of normal personality development and would perhaps concede that aspects of personality structure might be relevant in terms of influencing symptom presentation and informing likely avenues of treatment/resistance to treatment. However, clinicians of this period mostly did not entertain the idea that personality per se might be the locus of pathology or the target of treatment.

It was not until the 1940s and 1950s that a new kind of patient, alongside the broadening application of psychoanalysis, necessitated a reformulation of psychoanalytic clinical theory, as it became clear that some patients who appeared to be normal, ‘neurotic’ patients experienced significant regression under the course of psychoanalytic treatment. Stern observed that this type of patient seemed to hover around a border between the neurotic and psychotic categories, and thus the notion of a nosological and conceptual borderline was conceived (Stern, 1938). These patients would sometimes become suicidal and exhibit parasuicidal behaviours, experience transient paranoid ideation and failure of reality testing, and generally did not seem to profit from standard psychoanalytic treatment. Psychoanalytic writers at the time proposed various terms for this heterogeneous group of patients, including ‘preschizophrenic’, ‘pseudoneurotic schizophrenics’ (Gunderson, 1984), reflecting a common assumption that this pattern of treatment response revealed an underlying susceptibility to psychosis and primary-process modes of thought (Kernberg, 1975), and that this condition might represent a prodromal stage in the development of schizophrenia. However, as Freud’s work on psychopathology and psychoanalytic method never discussed patients like the ‘pseudoneurotics’ who started to appear during this period, clinicians and theorists had to grapple with this issue for themselves, and so a period of relatively free and creative theorising ensued in which various writers put forth their views on how to define, understand, and treat the symptoms experienced by these patients.

In the early 1950s, amid this flurry of activity, the American Psychiatric Association published the first edition of the Diagnostic and Statistical Manual of Mental Disorders (1952), which featured personality disorder categories that were mainly derived from psychoanalytic theory (Crocq, 2013). The DSM-I specified twelve discrete personality disorders which were
subdivided into ‘personality pattern disturbances’, referring to pervasively disordered patterns of object relations and/or defense constellations, and ‘personality trait disturbances’, mirroring patterns of extreme or exaggerated individual traits, and four types of ‘sociopathic personality’, referring to various forms of antisocial behaviour patterns.

Table 1

*DSM-I personality disorders*

<table>
<thead>
<tr>
<th>Personality pattern disturbances</th>
<th>Inadequate personality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Schizoid personality</td>
</tr>
<tr>
<td></td>
<td>Cyclothymic personality</td>
</tr>
<tr>
<td></td>
<td>Paranoid personality</td>
</tr>
<tr>
<td>Personality trait disturbances</td>
<td>Emotionally unstable personality</td>
</tr>
<tr>
<td></td>
<td>Passive-aggressive personality</td>
</tr>
<tr>
<td></td>
<td>Compulsive personality</td>
</tr>
<tr>
<td></td>
<td>Personality trait disturbance, other</td>
</tr>
<tr>
<td>Sociopathic personality</td>
<td>Antisocial reaction</td>
</tr>
<tr>
<td></td>
<td>Dissocial reaction</td>
</tr>
<tr>
<td></td>
<td>Sexual deviation</td>
</tr>
<tr>
<td></td>
<td>Addiction</td>
</tr>
</tbody>
</table>

Table 2

*DSM-I diagnostic criteria for emotionally unstable personality*

‘In such cases the individual reacts with excitability and ineffectiveness when confronted by minor stress. His judgment may be undependable under stress, and his relationship to other people is continuously fraught with fluctuating emotional attitudes, because of strong and poorly controlled hostility, guilt, and anxiety. This term is synonymous with the former term “psychopathic personality with emotional instability.”’
Thus, contrary to widespread belief, the broad outlines of a system much like the ICD-10/DSM-5 diagnostic schedule was established as early as the 1950s; however, unlike the diagnostic systems found in the DSM-III onwards, the personality disorder categories in the DSM-I and DSM-II were given in descriptive, qualitative terms which allowed for considerable diagnostic discretion (Table 2), and indeed, research conducted in preparation for DSM-III demonstrated that these categories typically showed poor interrater reliability and were vulnerable to idiosyncratic usage (Spitzer & Fleiss, 1974).

In the 1960s and 1970s, groups of psychoanalytic theorists and clinicians initiated a more focused effort to define and characterise the concept of personality disorders. Perhaps the most prominent among these writers was Otto Kernberg (Kernberg, 1975, 1984), who introduced the notion of a spectrum of personality organisations, conceived of as relatively stable structures of personality at a level of abstraction deeper than surface phenomena like concrete symptoms, or individual patterns of defence mechanisms or object relations constellations.

Kernberg’s system of personality structure differentiates between neurotic personality organisation (NPO), characterised by intact reality testing, the use of mature defence mechanisms, and a robust differentiation between self and other and stable object relations; borderline personality organisation (BPO), characterised by mainly intact reality testing, the use of primitive defence mechanisms like splitting, denial, and projective identification, and unstable object relations, featuring poorly integrated representations of self and other, including binary processes of merging with the object and becoming painfully detached/isolated from it; and psychotic personality organisation (PPO), characterised by impaired reality testing, including varying degrees of actively psychotic processes, primitive defence mechanisms, much like in BPO, and unstable and very poorly represented object relations, including failure of object constancy, severe annihilation anxiety and uncertainty about the patient’s own existence.

Table 3

<table>
<thead>
<tr>
<th>Self-experience; self–other-differentiation</th>
<th>Defence structure</th>
<th>Reality testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neurotic personality organisation</td>
<td>Well-integrated object relationships and intact capacity for relationships here and now</td>
<td>Generally mature defences</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>Borderline personality organisation</td>
<td>Poorly integrated concept of self and ego functioning; poorly integrated object relationships producing instability in here-and-now relationships</td>
<td>Primitive defences such as splitting, denial, projective identification, manic defence</td>
</tr>
<tr>
<td>Psychotic personality organisation</td>
<td>Very poorly integrated concept of self, possibly involving failures of object constancy and failures of self-delineation; severely disturbed object relationships; symbiotic and/or autistic object-relating</td>
<td>Primitive defences such as splitting, denial, projective identification, manic defence</td>
</tr>
</tbody>
</table>

On Kernberg’s conception, most patients with severe personality disorders are organised at the borderline level, meaning that Kernberg’s notion of the BPO is a broader category than what is today called borderline personality disorder (BPD) and also includes what we today know as schizoid, paranoid, and narcissistic personality disorders. Less severe forms of personality pathology, such as the diagnostic entities today known as avoidant, obsessive-compulsive, and dependent personality disorders would – on Kernberg’s view – fall into the neurotic range, exhibiting use of mature defence mechanisms, robust reality testing, and well-integrated self
and object representations. On the other side of the spectrum, psychotic personality organisation includes overt psychotic functioning of the kind seen in schizophrenia, psychotic depression, or mania, as well as delusional disorder and some forms of severely decompensated personality disorders (Kernberg, 2019).

Working in parallel with Kernberg, Heinz Kohut (1971) challenged Kernberg’s system and delineated another form of closely related pathology, the narcissistic personality disorder. Like borderline personality organisation, these patients show disordered object relationships and the use of primitive defence operations like denial, manic defences, and projective identification; however, unlike borderline personality organisation, these patients appear to evince a higher degree of ego integration, allowing them a precarious but more unified sense of self and a higher degree of seemingly unimpeded functioning in some domains of life; these individuals are indeed sometimes quite effective intellectually, in superficial social contexts, and in professional roles. Kohut viewed pathological narcissism as an outcome of frustrated normal narcissistic processes which are part of the development of the so-called self-structure. Under normal circumstances, this development leads to a secure sense of self-esteem, but under suboptimal conditions, thought to occur when parents or caregivers are grossly inattentive to the child’s mirroring needs, meaning the child’s need to have its developmental accomplishments validated and ‘seen’ by the object, and to the child’s idealising needs, meaning the child’s need to look up to and – transiently – feel part of a protective, all-good object. These combined deprivations make it impossible for the child to experience and metabolise the requisite ‘optimal frustrations’ necessary to internalise what Kohut called ‘self-object’ functions, which in turn are necessary to have a sense of secure self-esteem and pursue narcissistic sublimations. According to Kohut, this leads to a condition – which has come to be known as narcissistic personality disorder – characterised by barren but superficially viable interpersonal relationships, shallow affect, and a very precarious underlying sense of self-esteem defended against by surface grandiosity.

While acknowledging the existence and importance of this symptom constellation and broadly agreeing as to the relevant etiological factors, Kernberg viewed narcissistic personality disorder as a subtype of the borderline personality organisation, the most significant difference being the presence of what he termed a ‘grandiose self’, constituted by introjected all-good self and object states, yielding a certain degree of self/ego coherence at the expense of a pervasively inadequate perception of self and others (Kernberg, 1975).
Kernberg’s contributions proved widely influential not only in psychoanalytic circles, but also in the broader world of psychiatry. At the same time, the concept of BPO was criticised for being abstract and difficult to operationalise in concrete terms that might be useful to the working clinician, and so a process began in which various writers offered their suggestions for how to define the novel ‘borderline syndrome’ (Grinker, 1978; Gunderson & Kolb, 1978), a process that was demarcated by the concurrent emergence of a broadly Kohutian concept of narcissistic personality disorder, and eventually culminated in the DSM-III definition of borderline personality disorder. Unlike the broader concept of borderline personality organisation, the DSM-III borderline syndrome is defined in terms of concrete, observable behaviour. This ensured the possibility of far greater interrater reliability, thus paving the way for research into the validity of the DSM-III diagnosis of borderline personality disorder. It also allowed clinicians of other theoretical persuasions to make the diagnosis without referring to theory-internal constructs like transference/countertransference dynamics.

Table 4

*DSM-III criteria for borderline personality disorder*

<table>
<thead>
<tr>
<th>The following are characteristic of the individual’s current and long-term functioning, are not limited to episodes of illness, and cause either significant impairment in social or occupational functioning or subjective distress.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. At least five of the following are required:</strong></td>
</tr>
<tr>
<td>(1) impulsivity or unpredictability in at least two areas that are potentially self-damaging, e.g., spending, sex, gambling, substance use, shoplifting, overeating, physically self-damaging acts</td>
</tr>
<tr>
<td>(2) a pattern of unstable and intense interpersonal relationships, e.g., marked shifts of attitude, idealization, devaluation, manipulation (consistently using others for one’s own ends)</td>
</tr>
<tr>
<td>(3) inappropriate, intense anger or lack of control of anger, e.g., frequent displays of temper, constant anger</td>
</tr>
<tr>
<td>(4) identity disturbance manifested by uncertainty about several issues relating to identity, such as self-image, gender identity, long-term goals or career choice,</td>
</tr>
</tbody>
</table>
friendship patterns, values, and loyalties, e.g., “Who am I?”, “I feel like I am my
sister when I am good”

(5) affective instability: marked shifts from normal mood to depression, irritability,
or anxiety, usually lasting a few hours and only rarely more than a few days, with
a return to normal mood

(6) intolerance of being alone, e.g., frantic efforts to avoid being alone, depressed
when alone

(7) physically self-damaging acts, e.g., suicidal gestures, self-mutilation, recurrent
accidents or physical fights

(8) chronic feelings of emptiness or boredom

B. If under 18, does not meet the criteria for Identity Disorder.

Table 5

**DSM-III criteria for narcissistic personality disorder**

The following are characteristic of the individual’s current and long-term functioning, are
not limited to episodes of illness, and cause either significant impairment in social or
occupational functioning or subjective distress:

A. Grandiose sense of self-importance or uniqueness, e.g., exaggeration of
achievements and talents, focus on the special nature of one’s problems.

B. Preoccupation with fantasies of unlimited success, power, brilliance, beauty, or
ideal love.

C. Exhibitionism: the person requires constant attention and admiration.

D. Cool indifference or marked feelings of rage, inferiority, shame, humiliation, or
emptiness in response to criticism, indifference of others, or defeat.

E. At least two of the following characteristics of disturbances in interpersonal
relationships:

(1) entitlement: expectation of special favours without assuming reciprocal
responsibilities, e.g., surprise and anger that people will not do what is wanted

(2) interpersonal exploitativeness: taking advantage of others to indulge own desires
or for self-aggrandizement; disregard for the personal integrity and rights of
others
(3) relationships that characteristically alternate between the extremes of overidealisation and devaluation

(4) lack of empathy: inability to recognize how others feel, e.g., unable to appreciate the distress of someone who is seriously ill.

Arguably, however, this formalisation of a clearly definable clinical syndrome also gains practical utility at the expense of a convincing and unified understanding of the structure of personality pathology, as the broad domains of personality organisation or structure suggested by Kernberg were divided into more narrowly defined discrete ‘disorders’.

Alongside the narcissistic and borderline disorders, the DSM-III also specifies paranoid, schizoid, and schizotypal personality disorders, which are arguably more severe and closer to what Kernberg termed the psychotic personality organisation. On the other side of the spectrum, the DSM-III also defined avoidant and dependent personality disorders, which are often understood to be milder, less disabling conditions which perhaps have a less complex aetiology, possibly centred around disordered attachment, as well as the histrionic and obsessive-compulsive disorders, which may perhaps be construed as extremes of normal traits rather than pervasive disorders of character.

**Treatment**

Available specialised treatments for personality disorders (Storebo et al., 2020) at present fall into two broad categories: psychoanalytically informed treatments, such as Transference-Focused Psychotherapy (Kernberg, 2016) and Mentalisation-Based Therapy (Bateman & Fonagy, 2004a, 2015), and approaches informed by the broader cognitive-behavioural tradition, the primary exponent of this philosophy being Dialectical-Behavioural Therapy (Linehan, 1993; May et al., 2016).

**Personality disorders in adolescence**

Until the first decade of the 21st century, the topic of personality pathology in adolescence had not received much attention, and it was widely considered that a) the reliability of such diagnostic assessments would probably be low, b) the validity might be poor, as it was assumed that normative developmental tasks might easily be mistaken for personality pathology, and c)
diagnosing such disorders, even faced with seemingly incontrovertible signs of its presence, might risk stigmatising vulnerable adolescents at an especially sensitive time in their lives. However, with the publication of a seminal paper by Chanen and colleagues (2004) and a accruing body of subsequent work (A. M. Chanen, 2015; Kaess et al., 2014; Newton-Howes, Clark, & Chanen, 2015), a consensus appears to be emerging that personality diagnosis in adolescence is valid, reliable, and probably more common than has previously been assumed.

Normative development in adolescence

Nevertheless, this research also suggests that the unique developmental processes that take place during adolescence interact with normative developmental processes of the adolescent years, which impact the expression and quality of personality pathology, which we will examine in the following section.

Erikson was one of the first theorists to take a sustained interest in this developmental stage (REF 1968). In his hierarchy of developmental crises, the adolescent years are characterised by a crisis of identity vs. identity confusion. During this stage, the crisis to be overcome is the question of who the adolescent experiences him- or herself to be, as social roles, beliefs, values, and the adolescent’s sense of self are all subject to re-examination and rapid change. According to Erikson, this crisis is normally resolved and leads to a commitment to certain goals and ideals, as the adolescent achieves a sense of fidelity, referring to a capacity for attaining closeness with others and entering into relationships of various kinds. Individuals who fail to negotiate this crisis, on the other hand, may experience differing degrees of unease, aimlessness, and identity confusion.

According to Mahler (Mahler, Pine, & Bergman, 1975), the infant goes through a process of differentiation, in which he or she begins to orient to his or her surroundings rather than the mother exclusively. He or she then enters the phase of practising, in which the young child explores the environment, but returns to mother for comfort and reassurance; this seems to mirror the attachment theoretical notions of venturing out to explore, and returning, when needed, to a secure base (Blum, 2004; Waters & Cummings, 2000). The phase of rapprochement follows next, in which the child must contend with the vicissitudes and ambivalence of increasing autonomy, including processes of splitting, in which object representations are transiently viewed as all good or all bad (echoing Klein’s concept of the paranoid-schizoid position); and consolidation, in which the child successfully consolidates the
split representations into an integrated representation (echoing Klein’s depressive position), thus achieving object constancy, which enables the child to soothe him- or herself even when the mother is absent.

Several writers have pointed out how adolescence seems to recapitulate the same stages (Blos, 1979), and some have identified potential homologies with Erikson’s implied trajectory (Brandt, 1977). Indeed, in a stage which closely mirrors the rapprochement phase, we can all too easily picture a typical adolescent in a fit of despondency or hateful rage over an actual, imagined, or relational disappointment, or in a state of totalising admiration, elation or object-directed yearning; similarly, we can vividly picture the adolescent, having achieved a more integrated sense of the complexity of the other, quietly mourning the loss of aspects of his or her childhood self all the while consolidating a more mature, outward-directed, relationally oriented experience of him- or herself.

From the perspective of empirical observation and neuroscience, several factors are known which would seem likely to affect the presentation and course of personality disorders. As Bo and colleagues (Bo et al., 2020) point out, the limbic system of the adolescent matures at a more rapid pace than the prefrontal cortex, which on a psychological and behavioural level translates into relatively greater emotional reactivity and relatively lower capacity for rational judgment and sound decision-making. They show greater amygdala response when interpreting feeling expressions, possibly signalling greater attentiveness to social stimuli than adults. They also point to a tendency of adolescents to engage in what is referred to as _co-rumination_, in which problems and obstacles are shared among friends and subjected to a unique kind of arrested problem-solving which seems to have a role in deepening and consolidating peer relationships rather than positively attending to the concrete issue at hand. Bo and colleagues also highlight a tendency for adolescents to be relatively more responsive to risk-related suggestions, demands, and pressures from others in a group setting.

**Pathological personality development in adolescence**

As these theoretical perspectives seem to illustrate from different angles, the adolescent years carry with them the promise and threat of a more permanent kind of separation from the adolescent’s primary objects. From the point of view of attachment theory and psychoanalysis, such a transition inevitably awakens strong attachment and/or dependent needs, with strong attendant depressive affect, as well as the potential of avoidant, obsessive-compulsive, or
dependent defensive reactions, all of which could be construed as a normative pattern of development. However, it also seems plausible that unless these reactions are properly negotiated and worked through, they could ossify and become permanent. Hypomanic, narcissistic, borderline, schizoid, and paranoid patterns of pathology may plausibly be seen as representing more dramatic outcomes of such developmental failure, likely also including extreme forms of object frustration/trauma as well as a greater component of constitutional vulnerability in the form of biogenetic predisposing factors.

As mentioned above, research conducted in recent years (A. M. Chanen, 2015; Kaess et al., 2014) suggests that the diagnosis of borderline personality disorder is as valid and reliable in adolescents as in adults. These studies also show the potential that lies in early intervention, which could potentially help prevent and mitigate the consolidation of the disorder at an early stage. This could also contribute in preventing a deterioration, thus yielding a poorer prognosis, and severe comorbidities like major depression and PTSD.

Personality disorders are often found in adolescents, with prevalences between 41% and 64% in clinical samples (Kongerslev et al., 2015); personality disorders are often comorbid with each other. Even as patients sometimes fall short of the specific requirements for a diagnosis over time, studies have suggested that functional impairment remains significant across the lifespan. Comorbid symptom (Axis I) disorders are common, and is associated with a worse prognosis. Furthermore, the presence of personality disorders in adolescents is associated with significant utilisation of health resources, including specialised services.

Treatment
Randomised controlled trials have shown that targeted preventive and early-intervention approaches show promise for adolescents with borderline personality disorder (Kongerslev et al., 2015). As of this writing (Bo et al., 2020), no active treatments have been shown to be superior to other active treatments, nor to competent treatment-as-usual (TAU). While disappointing, such findings are in keeping with general results in psychotherapy research, and could point to the complexities of designing a study capable of demonstrating specific and superior effectiveness.

The treatment methods that have been examined include emotion regulating training (ERT), cognitive analytical therapy (CAT), integrative borderline personality disorder-oriented adolescent family therapy, mentalisation-based therapy (MBT), psychodynamic group therapy,
and dialectical behavioural therapy for adolescents (DBT-A). As Bo and colleagues note, all treatments were shown to be effective, but no more so than treatment as usual except in the case of MBT.

**Future initiatives**

The Norwegian national centre of competence for personality psychiatry (Norwegian *Nasjonal kompetansetjeneste for personlighetspsykiatri*) is currently working on designing guidelines for assessment, diagnosis, early intervention, and treatment of personality disorders in general. The centre is able to provide specialised training in mentalisation-based therapy (MBT), which is one of its research interests. The purview of this project is broad, and work is currently being undertaken to provide guidelines for diagnostic assessment of adolescents (Korsgaard, personal communication, 2020).

**The SIDP-IV – a semistructured diagnostic interview for assessing personality disorders**

The Structured Interview for DSM-IV Personality (SIDP-IV) is a semi-structured diagnostic interview for DSM-IV personality disorders. Its precursor, the Structured Interview for DSM-III Personality (SIDP), was published in 1983, and has seen broad use both clinically and in research settings (Battaglia, Cavallini, Macciardi, & Bellodi, 1997; Torgersen, Kringlen, & Cramer, 2001; Zimmerman & Coryell, 1989).

The SIDP-IV covers all DSM-IV personality disorder diagnoses, including the diagnoses of negativistic, self-defeating, and depressive personality disorders, which were included in appendix B of the DSM-IV with a recommendation for future research into their reliability and validity (Livesley, 1995). It also includes a brief interview on conduct disorder of childhood, which can yield important information as to the developmental history of patients with antisocial personality disorder.

Unlike some other semi-structured interviews for personality disorders, the SIDP-IV is structured around thematic areas, e.g., work style, close relationships, and emotions, which means that the items for each diagnostic category are not presented together. This arguably helps to create a more organic conversation. One could also speculate that this relative lack of transparency compared to an instrument like the SCID-II, which groups items together
according to their corresponding diagnosis, could be helpful in terms of detecting or preventing malingering or ‘faking bad’.

The SIDP-IV assesses the presence of symptoms indicative of personality disorder within the five years leading up to the interview. In working with adolescent samples, this is often changed to two years, which is also the case in this sample (Korsgaard, 2017).

Responses are given a score between 0 to 3 according to the guidelines set out below (Table 6).

Table 6

*Scoring guidelines for the SIDP-IV*

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Not present or limited to rare, isolated examples</td>
</tr>
<tr>
<td>1</td>
<td>Subthreshold – some evidence of the trait, but it is not sufficiently pervasive or severe to consider the criterion present</td>
</tr>
<tr>
<td>2</td>
<td>Present – criterion is clearly present for most of the last 5 years (i.e., present at least 50% of the time during the last 5 years)</td>
</tr>
<tr>
<td>3</td>
<td>Strongly present – criterion is associated with subjective distress or some impairment in social or occupational functioning or intimate relationships</td>
</tr>
</tbody>
</table>

Scores of 2 or 3 count towards the criteria for making an individual personality disorder diagnosis and are referred to as *criterion scores*. The number of criterion scores – i.e., questions judged to be at the ‘2’ or ‘3’ levels – are added up, and if the number of criterion scores for an individual diagnosis meets the threshold, e.g., five symptoms for borderline personality disorder or four for avoidant personality disorder, the diagnostic requirements for this diagnosis have been met. Thus, the SIDP-IV yields data at several levels, as outlined in Table 7.

Table 7
### Categories of data yielded by the SIDP-IV

<table>
<thead>
<tr>
<th>Data</th>
<th>Type (categorical/interval)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence/absence of discrete</td>
<td>Categorical</td>
<td>Indicates the presence or absence of an individual DSM-IV personality disorder; several may co-occur and suggests a higher disease burden</td>
</tr>
<tr>
<td>personality disorder</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Criterion score</td>
<td>Interval</td>
<td>Indicates presence of symptoms of the quality and severity commonly seen in personality disorders; a high amount of criterion scores may suggest personality pathology even in the absence of discrete personality disorders (Pagan, Oltmanns, Whitmore, &amp; Turkheimer, 2005). May be useful for dimensional classification of personality pathology</td>
</tr>
<tr>
<td>Sum scores</td>
<td>Interval</td>
<td>As this data type also includes the sum of ‘1’ scores, the sum scores of the full SIDP-IV or for the individual personality disorder categories may be useful in terms of assessing the extent of subthreshold personality issues</td>
</tr>
</tbody>
</table>

---
There have been some reliability studies of the SIDP-IV. Studying a non-clinical sample of 433 American air force recruits, Jane and colleagues (2006) found intraclass correlation coefficients for categorical diagnosis to vary between -0.01 (schizoid personality disorder) and 0.85 (avoidant personality disorder), with a mean at 0.5, indicating moderate overall reliability for this metric. Higher figures were obtained using criteria scores, ranging between 0.65 for schizotypal personality disorder and 0.9 for avoidant personality disorder, with a mean of 0.77.

While few studies have looked into the validity of the current version of the instrument, good levels of validity have been found for previous versions. For our purposes, the findings of Brent and colleagues are particularly relevant, as they looked into a sample of adolescents (Brent, Zelenak, Bukstein, & Brown, 1990).

**FEST-IT**

The data analysed in this thesis were collected from the First Experimental Study of Transference In Teenagers (FEST-IT), a research project funded by the University of Oslo and the Vestfold Hospital Trust and conducted by researchers at these institutions (Ulberg, Hersoug, & Hoglend, 2012). The aim of this project was to investigate the specific effects of transference work/interpretations in psychodynamic psychotherapy conducted with teenagers. The main hypothesis of the researchers was that psychodynamic psychotherapy making use of transference interventions would produce better outcomes than psychodynamic psychotherapy without transference interventions. The moderating effects of quality of object relations, presence of personality disorder, and gender would also be examined (Ulberg et al., 2012).

As discussed above, transference work has traditionally been considered a central element of psychoanalytic/psychodynamic psychotherapy. It is commonly thought that transference work, given enough time and within the framework of a safe and effective alliance, reactivates experiences and affect associated with the patient’s objects, rendering these aspects of the patient’s personality accessible to therapeutic work. Ideally, this enables the therapist and patient to produce a transmuting or corrective experience that can address some of the patient’s potentially maladaptive expectations and views of self and others (Thomä & Kächele, 1987).

In order to do so, the researchers have used what is called a dismantling design, a type of randomised controlled study (RCT) design which enables researchers to isolate the individual effect of one of several components in a complex procedure (Papa & Follette, 2015).
The general reasoning behind this design is that given a complex procedure $p$, consisting of the elements $a$, $b$, and $c$, we should be able to examine the specific effects or contributions of one element, for instance $a$, to the overall effect of $p$ by comparing a condition under which $a$, $b$, and $c$ are all present to a condition in which only elements $b$ and $c$ are present. This allows for analyses of individual effects of the element under observation, i.e., whether adding or removing $a$ changes the outcome or other process variables, as well as analyses of possible interactive effects, i.e., whether adding or removing $a$ changes the outcome of the aggregate procedure over and above what is accounted for by this variable individually. Furthermore, the dismantling design ensures that apart from the element under observation, other things are held equal to the greatest extent possible, an important desideratum in terms of internal and ecological validity (Borkovec & Castonguay, 1998). Moreover, unlike some designs which appear to compare an active condition to a pseudo-active condition that could not plausibly be effective on its own terms, this design provides a comparison between two bona fide active conditions (Shedler, 2018; Wampold, Minami, Baskin, & Callen Tierney, 2002).

Thus, in the FEST-IT project, patients were randomised to one of two conditions. Patients in both conditions received 45-minute sessions of psychodynamic psychotherapy once weekly for 28 weeks, with therapies lasting less than twelve sessions being defined as dropouts.

**Treatment**

The treatment followed a manual for short-term psychoanalytic psychotherapy for depressed adolescents which ‘combines aspects of [short-term psychoanalytic psychotherapy] that focus principally on techniques aimed at helping young people overcome developmental problems, as well as emphasizing the role of the interpretation of unconscious conflicts, attachment theory and the concepts of internal working models’ (Cregeen & Catty, 2017; Ness et al., 2018). Optionally, parallel work with parents was included if the adolescent consented to this. Furthermore, the use of antidepressant medication was permitted in cases of severe depression, as per conservative clinical norms (Grøholt, 2011).

Therapists in the transference condition were instructed to make use of transference-focused interventions, including ‘addressing transaction in the patient-therapist relationship’, ‘encouraging the patient to explore thoughts and feelings about the therapy and the therapist’, ‘encouraging the patient to discuss how he/she believes the therapist might feel or think about him/her’, ‘including the therapist in interpretive linking to dynamic elements (conflicts) and
direct manifestations of transference’, and ‘interpret and link repetitive interpersonal patterns to transaction between patient and therapist’ (Ulberg, Amlo, & Hoglend, 2014). They were encouraged to use such interventions between one and three times each session (Ulberg et al., 2012).

In the non-transference condition, the use of transference interventions was proscribed. Instead of focusing on the transference relationship, therapists were encouraged to use analogous interventions, including ‘address[ing] interpersonal transactions with other(s)’, ‘actively encourag[ing] the patient to explore thoughts and feelings about his/her relationship to other(s) including their style and behaviour’, ‘encourage the patient to discuss how other(s) might feel or think about him/her’, ‘make interpretive linking of dynamic elements (conflicts) in the patient’s relationships with other(s)’, and ‘attempt to explore interpersonal repetitive patterns with important other(s) and with parental figures’ (Ulberg et al., 2016).

The primary criterion for inclusion in the FEST-IT study was current presence of a unipolar depressive episode of any severity (Ness et al., 2018). Would-be participants with psychotic disorders and significant neurodevelopmental disorders, i.e., autism-spectrum disorders and generalised learning difficulties, were excluded. However, in order for the sample to be representative of patients in Norwegian outpatient child and adolescent psychiatry, patients with personality disorders and other kinds of comorbid pathology were not excluded; neither were patients with problematic substance use in the absence of clear dependence (Ulberg et al., 2012).

**Inclusion**

Patients eligible for inclusion in the study were assessed using a psychodynamic clinical interview, the Mini International Neuropsychiatric Interview (MINI) (Sheehan et al., 1998), and the Structured Interview of DSM-IV Personality (SIDP-IV), and were diagnostically assessed according to DSM-IV criteria. Patients who went on to be included were also administered the Psychodynamic Functioning Scales (PFS) (Ness et al., 2018), the Global Assessment of Functioning (GAF) (American Psychiatric Association, 1994), the Symptom Checklist-90 Revised (SCL-90-R), the Interpersonal Problems Circumplex Version (IIP-C), and the Beck Depression Inventory II (BDI-II); among these, the PFS was the primary outcome variable while the GAF was the secondary outcome variable (Ness, 2019).
The initial protocol for the project planned to include 100 participants. In the event, 69 participants were included, out of whom 21 participants dropped out, in line with what has commonly been found in studies on adults (Swift & Greenberg, 2012) and adolescents (De Haan, Boon, De Jong, Hoeve, & Vermeiren, 2013). 57 (82.6%) of the participants were female (Ulberg, personal communication, 2020).

The FEST study

The FEST-IT project shares some important aspects of its design with the First Experimental Study of Transference (FEST) project, which was conducted with adult participants for almost a decade, from 1993 to 2001. However, this study did not target a specific diagnostic group, and instead included 100 participants representative of the kinds of patients that are typically referred to a Norwegian outpatient psychiatric clinic, including patients with depression, anxiety, personality disorders, and interpersonal problems, while excluding patients with psychosis, bipolar disorder, organic mental disorder, substance use disorders, and long-term inability to work (P. Hoglend et al., 2006). As in the FEST-IT study, participants were randomised to a transference condition (n = 52) and a non-transference condition (n = 48). Participants were offered weekly 45-minute sessions for a maximum duration of one year.

The main hypothesis for the FEST study was that participants in the transference group would show better outcomes than participants in the non-transference group. No difference was detected (P. A. Hoglend et al., 2008). However, contrary to the stated expectations of the researchers, their results showed that ‘patients with a lifelong pattern of poor object relations profited more from [one] year of therapy with transference interpretations than from therapy without transference interpretations’, an effect which was found to persist over the four-year study period (P. A. Hoglend et al., 2008). The results also suggested that women respond better to transference interventions (Ulberg, Johansson, Marble, & Hoglend, 2009), which was even more pronounced in women with poor relational functioning (Ulberg, Marble, & Hoglend, 2009).
**Ethics**


The present study was individually approved by the The Central Norway Regional Ethics Health Committee, under the same filing code (REK: 2011/1424 FEST-IT).

**Methods**

**Properties of the sample**

For the purposes of this inter-rater reliability study, out of the total population of 67 participants, 25 were selected using a random algorithm (Ulberg, personal communication) to increase the likelihood of obtaining a sample that would be representative of the study population.

Of the 25 participants selected, nine were interviewed at the pre-treatment time-point (36%), nine were interviewed at the post-treatment time-point (36%), and seven were interviewed at the one-year follow-up time-point (28%).

All participants in this sample were female. As noted above, the proportion of female patients in the population at large was 82.6%, meaning that this sample was probably adequately representative of the population under examination.

**Procedure**

All participants included in the sample were screened for personality pathology using the SIDP-IV. The interview was conducted by experienced clinicians who were either specialists in clinical psychology (Norwegian *psykologspesialist*) or specialists in psychiatry. Furthermore, these clinicians had also received extensive training in psychodynamic/psychoanalytic psychotherapy.

For this thesis, the author, a graduate student in clinical psychology, listened to audio recordings of the SIDP-IV interviews and scored the interviews independently of the experienced clinicians.
In order to prepare for this, I received two sessions of training in the use of the instrument from the principal advisor, H. O. K., who also instructed the clinicians who performed the original SIDP-IV scorings. By scoring an example interview (which is not part of the present sample) interactively, I was able to observe a representative interview and receive feedback as I went along. At the conclusion of these sessions, it was considered that I had received sufficient training to validly and reliably score the SIDP-IV.

As noted in the section on the SIDP-IV, this instrument yields different kinds of data. In addition to categorical data on presence (or number, in the case of concurrent personality disorders) of symptoms corresponding to a personality disorder diagnosis, it also provides data on the amount of items judged at the ‘present’ or ‘strongly present’ level, which can be useful in creating a dimensional profile. Furthermore, including the number of items judged at the subthreshold level provides information on overall, low-grade personality issues.

Statistical analyses

This thesis examined the inter-rater reliability of the SIDP-IV. Inter-rater reliability is the extent to which two or more judges or observers agree about an assessment, such as whether a patient does or does not have a diagnosable personality disorder. However, different procedures are used depending on the nature of the assessment, e.g., categorical or continuous; the number of observers, e.g., two or more; and whether one is interested in absolute agreement or relative consistency. I decided to compute the intra-class correlation (ICC) coefficient and Cohen’s kappa coefficient, which are commonly used metrics of inter-rater reliability.

In analysing the data, I chose to compute a) ICC coefficients at the diagnostic level, b) ICC coefficients at the criterion score level, c) ICC coefficients at the continuous level, referring to the distribution of sum scores regardless of diagnostic status, d) Cohen’s kappa coefficient at the level of categorical diagnosis, referring to whether and how many diagnoses were made, and e) Cohen’s kappa at the level of presence or absence of any diagnosis, a dichotomous variable.

There are different versions of the ICC correlation computation depending on the variables cited above (Shrout & Fleiss, 1979). For this thesis, the same data were assessed by four experienced and one inexperienced judge (the author). One option would be to select a design which compared all five raters to each other, providing an average figure for the degree of reliability in the sample. However, because our sample is relatively small (n = 25), and because the judges
belonged to two clearly delineated categories, namely experienced clinicians and a relatively inexperienced graduate student, I decided to treat the four experienced clinicians as one aggregate rater, effectively reducing the design to a two-way comparison.

In the following analyses, I used IBM SPSS for Microsoft Windows, version 26, in which this ICC procedure is referred to as a two-way, random design, sometimes referred to as an ICC (2,2) model. Furthermore, I specified that I wanted the analyses to emphasise the degree of consistency rather than absolute agreement. In interpreting the results, I have used the average values rather than the single-rater values.

Cohen’s kappa coefficient is a reliability statistic which assumes a) that there are only two raters to be compared and b) that the assessed variable is categorical, e.g., presence or absence of personality disorder. Thus, having decided to treat the clinician group as equivalent to one individual, this statistic yields additional information about the degree of consistency of categorical judgments about presence or absence of diagnosis.

Results
Intra-class correlations
The intra-class correlation coefficients for the sample (n = 25) are presented in Table 8.

Table 8

<table>
<thead>
<tr>
<th></th>
<th>Categorical diagnosis</th>
<th>Criterion scores</th>
<th>Continuous scores (including ‘1’ scores)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intra-class coefficient</td>
<td>0.876</td>
<td>0.952</td>
<td>0.917</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>0.718-0.945</td>
<td>0.891-0.979</td>
<td>0.808-0.964</td>
</tr>
<tr>
<td>F-test value</td>
<td>8.036</td>
<td>20.902</td>
<td>12.062</td>
</tr>
</tbody>
</table>
Values of $df_1$ and $df_2$

<table>
<thead>
<tr>
<th></th>
<th>24 and 24</th>
<th>24 and 24</th>
<th>23 and 23$^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-test significant?</td>
<td>Yes, $p &lt; 0.0005$</td>
<td>Yes, $p &lt; 0.0005$</td>
<td>Yes, $p &lt; 0.0005$</td>
</tr>
</tbody>
</table>

**Kappa coefficients**

Cohen’s kappa coefficient was obtained for correspondence between number of diagnoses (0, 1, 2, or more) between the raters. This coefficient was also calculated for the dichotomous variable of diagnosis/no diagnosis. The results are presented in Table 9.

**Table 9**

*Kappa coefficients for categorical diagnosis and dichotomous diagnosis*

<table>
<thead>
<tr>
<th></th>
<th>Kappa coefficient</th>
<th>Asymptotic standard error</th>
<th>T-test statistic</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categorical diagnosis</td>
<td>0.623</td>
<td>0.134</td>
<td>4.376</td>
<td>Yes, $p &lt; 0.0005$</td>
</tr>
<tr>
<td>Dichotomous diagnosis</td>
<td>0.896</td>
<td>0.101</td>
<td>4.506</td>
<td>Yes, $p &lt; 0.0005$</td>
</tr>
</tbody>
</table>

**Discussion**

According to one common convention, ICC coefficients below 0.5 are indicative of poor reliability, coefficients between 0.5 and 0.75 indicate moderate reliability, coefficients between 0.75 and 0.9 indicate good reliability, and coefficients greater than 0.9 indicate excellent reliability (Koo & Li, 2016). Moreover, it is generally recommended that ICC coefficient findings be reported in terms of 95% confidence intervals, specifying the reliability at the lower and upper bounds of the variance, as the true score of the coefficient could theoretically fall anywhere in this range. As an example, a reported ICC coefficient of 0.8, with a lower bound of 0.7 and a higher bound of 0.9, may be reported as showing moderate to excellent reliability.

Thus, based on these criteria, the ICC coefficient for the categorical diagnosis, at 0.876 (95% CI 0.718–0.945), shows moderate to excellent reliability. The ICC coefficient for the criterion scores, at 0.952 (95% CI 0.891–0.979), shows good to excellent reliability, with most of the

---

$^3$ For the continuous test, one pair was left out as the experienced rater had not specified ratings for individual questions.
confidence interval falling within the excellent range. The ICC coefficient for continuous scores, at 0.917 (95% CI 0.808–0.964), shows good to excellent reliability.

A common convention for interpreting Cohen’s kappa coefficient holds that values between 0 and 0.2 indicate poor agreement, values between 0.21 and 0.4 indicate fair agreement, values between 0.41 and 0.6 indicate moderate agreement, values between 0.61 and 0.8 indicate good agreement, and values between 0.81 and 1 indicate very good agreement (Altman, 1999).

Based on these criteria, the kappa coefficient for number of discrete diagnoses, at 0.623, is good, whereas the kappa coefficient for the dichotomous question of diagnosis vs. no diagnosis, at 0.896, is very good.

**Strengths and limitations**

The most important strength of this data material lies in shedding light on the reliability of the SIDP-IV diagnostic interview in a youth population. To my knowledge, few other researchers have investigated this topic, and hopefully the findings of the present study will prove useful for researchers and clinicians working with adolescent patients.

The primary limitation of our research design is the fact that it only includes 25 participants. Some writers suggest that an inter-rater research design should have at least 30 participants, and ideally include at least three raters (Koo & Li, 2016). This potentially reduces the statistical power of our findings. In my view, however, the values of the coefficients obtained, which were all above 0.85 for the ICCs and above 0.6 for the kappa coefficients, hopefully offsets this potential weakness in the data. Indeed, recalculating the figures using a more conservative 99% confidence interval, leaving only a 1% probability that the true value of the ICC coefficient is in fact outside the confidence interval, produces results which are also broadly reliable, ranging from a minimum lower bound of 0.631 (moderate reliability), for categorical diagnosis, to a lower bound of 0.858 (good reliability) for criterion scores, as seen in Table 10.

**Table 10**

99% confidence intervals of the ICC coefficient
<table>
<thead>
<tr>
<th></th>
<th>Categorical diagnosis</th>
<th>Criterion scores</th>
<th>Continuous scores (including ‘1’ scores)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intra-class coefficient</td>
<td>0.876</td>
<td>0.952</td>
<td>0.917</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>0.631–0.958</td>
<td>0.858–0.984</td>
<td>0.748–0.973</td>
</tr>
<tr>
<td>F-test value</td>
<td>8.036</td>
<td>20.902</td>
<td>12.062</td>
</tr>
<tr>
<td>Values of $d/1$ and $d/2$</td>
<td>24 and 24</td>
<td>24 and 24</td>
<td>23 and 23</td>
</tr>
<tr>
<td>F-test significant?</td>
<td>Yes, $p &lt; 0.0005$</td>
<td>Yes, $p &lt; 0.0005$</td>
<td>Yes, $p &lt; 0.0005$</td>
</tr>
</tbody>
</table>

Another potential weakness is the decision to treat the four experienced raters as one individual for the purpose of the statistical analysis. This runs the risk of obscuring effects due to differences among these observers or potentially divergent individual assessments. In my opinion, this is compensated for by the advantages conferred by simplifying the data analysis. An alternative approach would be to perform a one-way, random effects analysis; however, such an approach is widely considered more appropriate for designs in which a larger pool of raters who are selected at random have assessed different subjects, creating a situation in which it is unfeasible to control for individual effects, an assumption which clearly does not hold for the present data material. Indeed, the data material does not appear to clearly meet the standard assumptions of any of the main ICC models, and therefore this decision hopefully presents a viable compromise, ensuring interpretability without compromising the essential integrity of the statistical logic.

A third potential shortcoming in my data material is the fact that the majority of the criterion scores and diagnoses made were within what is commonly referred to as the cluster C personality disorders, i.e., the avoidant, dependent, and obsessive-compulsive personality disorders. The SIDP-IV also includes the diagnostic categories of depressive, self-defeating, and negativistic disorders, which are conceptually similar in being disorders characterised by high baseline dysphoric affect and neuroticism (Phillips et al., 1998; Skodol, Oldham, Gallaher, & Bezirganian, 1994). With the exception of obsessive-compulsive personality disorder, most criterion score and positive diagnoses were found within this enlarged anxious-dysphoric category, with very few items endorsed in the antisocial, narcissistic, schizoid, schizotypal, and paranoid disorder categories.
One possible implication of this is a risk of artificial inflation of reliability values, the reason being that the likelihood of obtaining perfect correspondence between raters is significantly higher when the patient under observation clearly displays no pathology at all, as illustrated in table 11. Future studies recruiting a broader population and employing more sophisticated statistical methods will be able to elucidate the extent of this potential issue.

Table 11

Sample patient data illustrating how presence vs. absence of pathology differentially affects the reliability calculation

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Number of criterion scores as assessed by an experienced clinician</th>
<th>Number of criterion scores as assessed by the author</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avoidant personality disorder (cluster C)</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Depressive personality disorder (provisional anxious-dysphoric cluster)</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Dependent personality disorder (cluster C)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Schizoid personality disorder (cluster A)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Antisocial personality disorder (cluster B)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Narcissistic personality disorder (cluster B)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Another possible implication is that the reliability of cluster A and B personality disorders has not been fully illuminated due to the absence of positive cluster A and B pathology. In other words, while this data material suggests that cluster A and B diagnoses are reliably not made in the absence of such pathology, it tells us little about the reliability of these diagnostic categories when such pathology is present. Again, the only seeming way to elucidate this question more fully would be to conduct a new study using a more broadly recruited sample.
Finally, the data was extracted from a predominantly female population, and, as noted above, my sample was exclusively female. Strictly speaking, this implies that the findings are only properly generalisable to all-female populations. In my professional opinion, there were no particular aspects of the data which would suggest that the data would not – in theory – generalise well to a mixed-sex sample. However, more research will be needed to determine this with a greater degree of certainty.

Clinical implications

The results demonstrated good inter-rater reliability for the SIDP-IV diagnostic interview as a whole. In keeping with the findings of other researchers (Jane et al., 2006), the results proved most reliable when assessed dimensionally, i.e., at the level of criterion scores and continuous scores. While this suggests a relatively lower reliability of the diagnostic scores, which would seem to be the most useful metric in terms of practical clinical usage, this could potentially point to a problem inherent in the overall validity of diagnosing personality disorders individually, a topic which has recently seen much attention in the research literature (Tyrer, Mulder, Kim, & Crawford, 2019).

In any event, the results seem to suggest that this instrument, which was originally designed and normed based on adult samples, remains highly reliable in the context of an outpatient adolescent population. It also suggests that the instrument remains reliable when used by a relatively inexperienced graduate student, compared to the experienced clinicians who conducted the initial assessments. This would seem to indicate that it can safely be recommended for usage in routine clinical situations. Indeed, in a widely cited study by Torgersen and colleagues (2001), nurses and medical students with little previous experience in clinical psychiatry were successfully trained to use a previous version of the instrument. Taken together, these findings indicate that it is robustly reliable across a range of situations and clinical contexts.

Unlike countries like the United Kingdom (National Collaborating Centre for Mental Health, 2009) and Australia (National Health and Medical Research Council, 2012), Norway presently has no nation-wide recommended procedures for assessing personality pathology in adolescent patients, but work is currently being undertaken to rectify this situation (Korsgaard, personal communication). In presenting my findings, I hope to bring attention to this emerging area of
interest, and to highlight one potentially useful diagnostic tool for the consideration of policymakers.

**Conclusion**

In recent decades, the question of the utility and validity of diagnosing personality disorders in adolescence has received an increasing degree of attention within the broader field of personality disorders. A growing number of studies suggest that early identification and intervention could prove crucial to securing a positive outcome for this group. Indeed, it is possible that successful intervention in adolescence could help prevent or mitigate personality-related morbidity in adulthood, which is known to incur heavy human and economic costs, as well as forestalling the development of severe comorbid disorders. (Bo et al., 2020; Kongerslev et al., 2015; Korsgaard, 2017).

The findings indicate a high degree of inter-rater reliability between the author, a graduate student of clinical psychology, and the more experienced clinicians, with the ICC coefficient for criterion score reliability at 0.952 and the kappa for dichotomous diagnosis at 0.896. My findings suggest that the SIDP-IV could prove a useful tool for diagnosing personality disorders in adolescent populations, in research as well as clinical settings; it is therefore hoped that my research can prove useful in future efforts to improve the assessment, prevention, and treatment of personality disorders in adolescent populations.
References


