

# **Nonsuicidal self-injury disorder: vurdering av interrater-reliabilitet og diagnosekriteriene ved bruk av CANDI (norsk versjon)**

*En pilotstudie*

Ingvild Haugland Øvretveit og Hanna Margrethe Kvarme



Hovedoppgave ved Psykologisk institutt,  
Det samfunnsvitenskapelige fakultet  
30 studiepoeng

UNIVERSITETET I OSLO

Vår 2020

## Sammendrag

Denne hovedoppgaven er forfattet av Ingvild Haugland Øvretveit og Hanna Margrethe Kvarme med tittel “Nonsuicidal self-injury disorder: vurdering av interrater-reliabilitet og diagnosekriteriene ved bruk av CANDI (norsk versjon) - en pilotstudie”. Vår veileder gjennom dette arbeidet har vært Annika Maria Désirée Melinder.

De foreslåtte diagnosekriteriene til Nonsuicidal self-injury disorder (NSSID) er i behov for videre studier, både hva gjelder reliabilitet og validitet. De ulike kriteriene har også vært diskutert med tanke på klinisk nytte. Det klinisk strukturerte intervjuet, Clinician-Administered Nonsuicidal self-injury Disorder Index (CANDI), tar for seg alle de foreslåtte kriteriene på NSSID i den femte versjonen av Diagnostic and Statistical Manual of Mental Disorders (DSM-5), og har tidligere vist seg å ha god interrater-reliabilitet, intern konsistens og konstruktvaliditet for en klinisk voksenpopulasjon med Non-suicidal self-injury (NSSI). Hovedformålet med denne pilotstudien var å undersøke interrater-reliabiliteten til den norske versjonen av CANDI på en klinisk ungdomspopulasjon med NSSI. Vi ønsket også undersøke den interne konsistensen, samt se resultatene på kriterienivå opp mot lignende studier for å diskutere kriterienes kliniske nytte. Data ble samlet inn av oppgavens forfattere og reliabilitetstesting ble utført på et utvalg av 19 ungdommer med NSSI, rekruttert gjennom “Villet egenskade”-studien ved Oslo Universitetssykehus. Ved å se på tidligere administrerte intervjuer, gjennomføre egne pilotintervjuer og undersøke selvskadingslitteraturen, ble det utarbeidet en skåringsmanual til bruk som veiledning for intervjuing og skåring av CANDI. Hva angår interrater-reliabilitet, viste resultatene kappakoeffisienter som strakk seg fra  $\kappa = .922 - 1.00$ , og intraclass coefficient mellom  $.861 - 1.00$  på item-nivå. Resultatene viste en absolutt enighet på kriterie- og diagnosenivå. Intern konsistens var adekvat med  $\alpha = .72$ . At resultatene ble oppnådd med begrenset trening og innenfor en rimelig tidsramme, styrker også gjennomførbarheten til intervjuet. Videre belyste resultatene flere trender som samsvarer med tidligere forskningsfunn: 1) E-kriteriet, som måler grad av ubehag av svekkelse, var det kriteriet som best skilte de individene med NSSID fra de med subterskel NSSI; 2) alle fylte B, C og D-kriteriet, som måler henholdsvis NSSIs funksjoner, følelser i forkant av NSSI og grad av overopptatthet, og NSSI som ikke er kulturelt akseptert; 3) antall funksjoner møtt og følelser i forkant korrelerer med grad av ubehag og svekkelse og kan på den måten kunne knyttes til alvorlighetsgrad av NSSI; 4) antall selvskadingsmetoder korrelerer med grad av overopptatthet, og kan kunne bidra til å skille mellom ulike subgrupper av NSSI; og 5) et frekvenskriterium på  $\geq 5$  episoder så i vårt utvalg ut til å være for liberalt. Oppsummert kan

det dermed se ut til at den nåværende organiseringen og vektleggingen av kriteriene er lite egnet til å fange opp heterogeniteten og alvorlighetsgraden blant ungdom med NSSI. Generaliserbarheten av resultatene begrenses av et lite og homogent utvalg, ikke fullstendig uavhengige ratere, samt mangel på andre mål på NSSI. Til tross for metodiske svakheter, fremstår CANDI som et viktig tilskudd i selvskadingslitteraturen, da det direkte tar for seg alle de foreslåtte DSM-5 kriteriene på NSSID, er anvendelig og viser god intern konsistens og inter-rater-reliabilitet.

## Forord

Arbeidet med denne hovedoppgaven har vært en lærerik opplevelse. Vi har fått muligheten til å fordype oss både i statistiske metoder og i selvskadingslitteraturen. Vi har også fått møte mange flotte ungdommer som har deltatt i dette prosjektet. Det er mange som har hjulpet oss på veien i arbeidet, og det er mange vi ønsker å rette en takk til.

Først og fremst takk til ungdommene! Takk for at dere turte og var villige til å dele av deres egne vonde erfaringer for å støtte forskningen. Vi ønsker dere alt vel!

Vi vil rette en stor takk til vår veileder Annika Maria Désirée Melinder. Takk, Annika, for all hjelp og støtte gjennom arbeidet, og for alle gode og konstruktive samtaler og tilbakemeldinger.

Vi vil gjerne takke Anne Trana Brager Larsen, som alltid har hatt en åpen dør. Takk for at du tok deg tid til å hjelpe oss, selv når du var meget opptatt med arbeidet med din egen doktorgrad. Takk også til Pål Zeiner, som har gitt oss viktige tilbakemeldinger og perspektiver, i tillegg til at han har tatt seg av alle søknader og kommunikasjon med REK.

Takk til BUP Oslo Nord og BUP Oslo Syd, for at vi har fått rekruttere deltakere fra deres enheter, og for at vi har fått benytte deres lokaler i forbindelse med intervjuer. Spesielt takk til Siri Toven på BUP Oslo Syd, som tok seg tid til å rekruttere og sette oss i kontakt med aktuelle deltakere.

Takk til familie og venner som har trøstet oss i motgang og alltid heiet på oss. Vi hadde ikke klart dette uten dere.

Ingvild Haugland Øvretveit og Hanna Margrethe Kvarme,  
Oslo, 05.05.20

# Innholdsfortegnelse

## 1. Innledning

1.1 Bakgrunn for oppgaven.....	8
1.1.1 Selvskadingsbegrepet.....	8
1.1.2 Prevalens.....	9
1.1.3 Selvskading i tenårene.....	9
1.1.4 Selvskadingens funksjoner.....	10
1.1.5 Diagnostiske utfordringer.....	11
1.1.6 The Clinician-Administrated Non-suicidal self-injury Disorder Index (CANDI).....	13
1.2 Formål med studien og problemstilling.....	14
1.3 Oppbygging og avgrensning av oppgaven.....	15

## 2. Teoretisk bakgrunn

2.1 Klinisk strukturerte intervjuer.....	15
2.1.1 Strukturerte intervjuer vs. semistrukturerte intervjuer.....	15
2.1.2 Ulike type skalaer.....	17
2.2 Metoder for å vurdere kvaliteten på et klinisk intervju.....	18
2.2.1 Reliabilitet og kilder til varians.....	18
2.2.2 Intern konsistens og interrater-reliabilitet.....	19
2.2.3 Reliabilitetskoeffisient.....	20
2.2.4 Intraclass correlation coefficient.....	20
2.2.5 Cohens kappa.....	21
2.2.6 Cronbachs alpha.....	22
2.2.7 Validitet.....	22
2.2.8 Korrelasjon.....	23

## 3. Materiale og metode

3.1 Utvikling av skåringsmanual og revidering av intervjuet.....	24
3.1.1 Endring av ordlyd og instruksjoner.....	25
3.1.2 Definerings av skåringer.....	27
3.1.3 Endring i skåringer.....	28
3.1.4 Markeringer.....	28
3.2 Bakgrunn for valg av deltakere.....	28

3.3 Utvalget.....	29
3.3.1 Inklusjonskriterier.....	29
3.3.2 Eksklusjonskriterier.....	29
3.3.3 Endelig utvalg.....	29
3.4 Studiedesign og prosedyrer.....	29
3.4.1 Videofilming av intervjuene.....	30
3.4.2 Blinding.....	30
3.4.3 Randomisering av rekkefølge.....	30
3.5 Statistisk analyse og bearbeiding.....	30
3.5.1 Databehandling.....	30
3.6 Ethiske betraktninger.....	31
3.6.1 Har forskning på NSSI en iatrogen effekt? .....	31
3.6.2 REK-godkjenning og oppbevaring av sensitivt datamateriale.....	32
3.6.3 Informert samtykke.....	32
3.6.4 Informasjon til ansvarlig behandler.....	33
3.6.5 Pasientsikkerhet og forebygging av suicidal atferd.....	33

## 4. Resultater

4.1 Interrater-reliabilitet.....	34
4.1.1 A-kriteriet.....	34
4.1.2 D-kriteriet.....	35
4.1.3 B-kriteriet.....	35
4.1.4 C1-kriteriet.....	35
4.1.5 C2-kriteriet.....	35
4.1.6 C3-kriteriet og C-kriteriet totalt.....	36
4.1.7 E-kriteriet.....	37
4.1.8 F-kriteriet og nåværende NSSID.....	37
4.2 Intern konsistens.....	37
4.3 Korrelasjoner.....	37
4.3.1 Korrelasjoner med oppfylt diagnose.....	38
4.3.2 Korrelasjoner med E-kriteriet.....	38
4.3.3 Korrelasjoner med selvskadingsmetoder.....	38
4.4 En oversikt over kohorter.....	38

<b>5. Diskusjon</b>	
5.1 Resultatene.....	39
5.1.1 Interrater-reliabilitet og intern konsistens.....	39
5.1.2 Kriterier og klinisk nytte.....	42
5.2 Begrensninger.....	46
5.2.1 Utvalget.....	46
5.2.2 Raterne.....	47
5.2.3 Studiedesignet.....	48
<b>6. Konklusjon.....</b>	<b>50</b>
<b>7. Referanseliste.....</b>	<b>51</b>
<b>8. Appendiks</b>	
<i>Appendiks A - CANDI norsk versjon</i>	
<i>Appendiks B - CANDI norsk versjon (med endringer)</i>	
<i>Appendiks C - Skåringsmanual</i>	
<i>Appendiks D - Samtykkeskjemaer «Villet egenskade»</i>	
<i>Appendiks E - Informasjonsskriv om delprosjektet i «Villet egenskade»</i>	
<i>Appendiks F - Protokoll for pasientsikkerhet og selvmordsforebygging</i>	

# 1. Innledning

## 1.1 Bakgrunn for oppgaven

### 1.1.1 Selvskadingsbegrepet

Selvskading er et vidt begrep og blir brukt for å beskrive en rekke atferder og intensjoner som forsøkt henging, impulsiv selvforgiftning, og overflatisk kutting i respons til uholdbar spenning (Skegg, 2005, s. 1471). Det finnes ulike begreper på selvskading, som “deliberate self-harm”, “parasuicide” og “non suicidal self-injury”, og bruken av ulike begreper varierer på tvers av landegrenser (Skegg, 2005). Begrepet “deliberate self-harm” blir ofte brukt som et mer omfattende begrep på selvskadende atferd, da både selvskading med og uten suicidal intensjon uten dødelige utfall blir inkludert (Muehlenkamp, Claes, Havertape & Plener, 2012). Begrepet brukes hovedsakelig i Australia og innenfor europeiske land. I motsetning er det i Canada og USA mer vanlig å bruke begrepet nonsuicidal self-injury (NSSI), som eksplisitt utelukker enhver atferd med et hvilket som helst nivå av suicidal intensjon (Muehlenkamp et al., 2012). Nærmere bestemt defineres NSSI som «den tilsiktede, selvpåførte ødeleggelsen av kroppsvev uten suicidal intensjon og for grunner som ikke er sosialt eller kulturelt akseptert» (International Society for the Study of Self-injury, 2018, vår oversettelse). Vanlige metoder inkluderer kutting, brenning, slåing og kloring, hvor kutting fremtrer som den vanligste metoden (Klonksy, 2007). NSSI er også repetitiv og direkte, på den måten at det endelige utfallet av skaden inntreffer uten intervensjoner (Tørmoen, 2017, s. 11). Det vil si at atferder som blant annet alkoholmisbruk, sulting, risikoatferd og forgiftning ikke omfattes under NSSI-begrepet. Det blir også ofte gjort skiller blant individer med NSSI, for eksempel med hensyn til alvorlighetsgrad, som kan bli brukt som utgangspunkt for å dele NSSI opp i subgrupper (Lloyd-Richardson, Perrine, Dierker & Kelley, 2007; Muehlenkamp, Brausch & Washburn, 2017).

Videre utgjør selvskading en risikofaktor for senere selvmord (Cooper et al., 2005) og selvmordsforsøk (Nock, Joiner, Gordon, Lloyd-Richardson & Prinstein, 2006; Whitlock, Pietrusza & Purington, 2013), og det kan også noen ganger være vanskelig å skille disse atferdene da intensjonen om å dø i noen tilfeller kan være tvetydig (Gratz, 2003). Det har imidlertid blitt funnet empiriske forskjeller mellom ungdommer som driver med ulike typer selvskadende atferd med og uten suicidal innhold (Jacobson, Muehlenkamp, Miller & Turner, 2008; Muehlenkamp & Gutierrez, 2004), og det har blitt argumentert for at ikke-



suicidal og suicidal selvskading må differensieres på bakgrunn av intensjon, dødelighet, metoder, prevalens, frekvens og funksjoner (Muehlenkamp, 2005; Walsh, 2007).

Når det refereres til studier i denne oppgaven, vil i hovedsak NSSI-begrepet brukes når studien det refereres til eksplisitt bruker dette begrepet. "Selvskading" vil brukes i de resterende tilfellene.

### 1.1.2 Prevalens

Selv om det foreligger en internasjonal variasjon i selvskadingsbegrepet, peker Hawton, Saunders og O'connor (2012) på funn fra flere samfunnsbaserte studier (Leo & Heller, 2004; Madge et al., 2008; Hawton, Rodham, Evans & Weatherall, 2002; Hargus, Hawton & Rodham, 2009; Moran et al. 2012) som viser at rundt 10 % av ungdommer rapporterer å ha skadet seg selv. I kliniske utvalg er andelen betydelig høyere, hvor noen studier har funnet prevalensrater i overkant av 60 % (DiClemente, Ponton, & Hartley, 1991; Hauber, Boon & Vermeiren, 2019). Videre viser både samfunnsbaserte og kliniske studier at selvskading er mer vanlig blant tenåringsjenter enn hos tenåringsgutter (Leo & Heller, 2004; Madge et al., 2008; Hawton et al., 2002; Bresin & Schoenleber, 2015). Selvskading blir økende vanlig fra 12-årsalderen, særlig hos jenter, og mellom 12 og 15-årsalderen er forholdet jente:gutt rundt 5:1 eller 6:1 (Hawton et al., 2012). Det har blitt gjort få studier på selvskading hos barn under 12 år, men sykehusinnleggelse som følge av selvskading hos denne aldersgruppen er uvanlig (Hawton et al., 2012).

### 1.1.3 Selvskading i tenårene

Inngangen til tenårene innebærer altså en spesielt sårbar periode for å utvikle selvskadende atferd (Brown & Plener 2017). Hjernen er i utvikling, noe som også omfatter de eksekutive funksjonene (Blakemoore & Choudhury, 2006). Ulike svekkelser i eksekutive funksjoner har videre blitt koblet til ulike subgrupper av selvskadere (Fikke, Melinder & Landrø, 2011). Fikke et al. (2011) fant at ungdommer med større alvorlighetsgrad av NSSI (vurdert ut ifra vevsskade og antall metoder), hadde svekkelser i arbeidshukommelsen, noe som kan implisere at evnen til å distrahere seg fra eller å regulere negative humøritilstander er svekket. Videre ble det funnet svekkelser i inhibitorisk kontroll hos gruppen med mindre alvorlig NSSI, noe som kan tyde på en svekket evne til å undertrykke emosjonelle atferdsmessige uttrykk av negative emosjoner. Dette kan i sin tur resultere i en mer impulsiv form for

selvskading. Det ble ikke funnet svekkelser i inhibitorisk kontroll hos gruppen med større alvorlighetsgrad av NSSI, noe som kan tyde på en mer betenkt og overveiet form for selvskading (Fikke et al., 2011). Dette er også i tråd med tidligere funn om at ungdom med moderat/alvorlig NSSI mer sannsynlig planlegger selvskadingen i forkant (Lloyd-Richardson et al., 2007).

I tillegg til det ovennevnte har det også blitt vist til de fysiologiske endringene i puberteten som kan danne intense, ukjente og overveldende erfaringer, noe som igjen kan medføre en utfordring i emosjonsregulering (Suyemoto, 1998). Videre kjennetegnes også perioden av flere relasjonelle utfordringer, som psykologisk separasjon, etablering av grenser og representering av selvet til seg selv og andre (Stänicke, 2019, s. 9). På den måten kan selvskading både være en måte å regulere følelser på, samt et forsøk på å bygge et selv og vansker med å kommunisere i en relasjonell kontekst (Stänicke, 2019, s. 9).

#### 1.1.4 Selvskadingens funksjoner

Selvskading som en måte å regulere affekt på, er videre den funksjonen som har mottatt bredest empirisk støtte (Klonsky, 2007). Andre funksjoner som har blitt beskrevet er selvskading som en måte å straffe seg selv på, å stanse en følelse av depersonalisering eller dissosiasjon, å unngå å gjøre et selvmordsforsøk, å få hjelp eller å påvirke andre, å etablere en grense mellom seg selv og andre, og som en måte å generere oppstemthet på (Klonsky, 2007).

Samlet sett blir selvskadingens funksjoner ofte delt inn i intrapersonlige og interpersonlige funksjoner, hvor de intrapersonlige funksjonene refererer til funksjoner som er rettet mot å endre en persons indre tilstand, mens de interpersonlige funksjonene er rettet mot å påvirke det ytre miljøet (Turner, Chapman & Layden 2012). Nock og Prinstein (2004) demonstrerte dette nærmere i en modell de kalte firefaktormodellen (FFM). Modellen beskriver selvskadingens funksjoner som varierende langs to dikotome dimensjoner: automatiske (intrapersonlige) vs. sosiale (interpersonlige). Disse funksjonene kan være både positivt og negativt forsterket. Den automatiske-negative forsterkningen refererer til selvskading som en måte å oppnå reduksjon i spenning eller andre negativt affektive tilstander. I automatisk-positiv forsterkning, blir selvskading brukt for å skape en ønskelig fysiologisk tilstand. På denne måten kan selvskading også fungere som en måte å generere følelser på (Brown, Comtois & Linehan, 2002). På den andre siden refererer sosial forsterkning til selvskading

som en måte å moderere eller regulere det sosiale miljøet på, for eksempel ved å få oppmerksomhet. Sosial-negativ forsterkning refererer på sin side til selvskading som en måte å unnslipe ubehagelige sosiale interaksjoner, slik som å ende en krangel eller unnslipe krav (Nock & Prinstein, 2004).

Selv om affektreguleringsmodellen er den funksjonen som har mottatt bredest empirisk støtte, kan ulike funksjoner tjene ulike individer på ulike tidspunkt (Suyemoto, 1998). Suyemoto (1998) peker på at selvskading er et komplekst fenomen med en kontekstuell forankring, noe som utgjør en stor utfordring når det kommer til forståelsen av de intrapsykiske og interpersonlige funksjonene til atferden. Av samme grunn er det også vanskelig å differensiere en funksjon fra en annen. Selvskading kan tjene flere funksjoner samtidig, men et forsøk på å skille dem er nødvendig for å øke vår forståelse av atferden, samt for behandling av det enkelte individet (Suyemoto, 1998).

### 1.1.5 Diagnostiske utfordringer

På tross av økt enighet rundt selvskadingens karakteristikk og fenomenologiske trekk, har mangelen på enighet rundt terminologi og definisjon gjort klassifikasjonen av slik atferd vanskelig (Muehlenkamp, 2005). Selvskading er et vanlig, men til nå, også et dårlig forstått fenomen (Klonsky & Muehlenkamp, 2007).

Lenge har NSSI kun vært inkludert som et symptom på borderline personlighetsforstyrrelse (BPD), på tross av at atferden også har blitt assosiert med andre personlighetsforstyrrelser (Klonsky, Oltmanns & Turkheimer, 2003; Nock et al., 2006) og flere akse 1-symptomatologier (Andover, Pepper, Ryabchenko, Orrico & Gibb, 2005; Klonsky et al., 2003; Nock et al., 2006; Favazza & Conterio, 1989). Funnet om at NSSI forekommer på tvers av flere diagnoser, og på den måten utgjør et transdiagnostisk fenomen, heller enn et symptom på en enkelt lidelse (Bentley, Nock & Barlow, 2014; Wilkinson & Goodyer, 2011), har ført til et behov for en reklassifisering av NSSI (Muehlenkamp, 2005; Shaffer & Jacobson, 2009). Shaffer og Jacobson (2009) argumenterte for at å kun studere NSSI innen en BPD-kontekst eller som en manifestasjon av suicidalitet vil hemme forskning og behandling av NSSI. På motsatt side vil en standardisert definisjon kunne forbedre kommunikasjon mellom behandler og pasient, gi informasjon til behandling, samt fasilitere forskning på NSSIs natur, forløp og utfall (Wilkinson & Goodyer, 2011).

Etableringen av NSSI som egen diagnose er fortsatt på et tidlig stadium, og forslaget til kriteriene i Diagnostic and Statistical Manual of Mental Disorders (5th ed.; DSM-5; American Psychiatric Association, 2013) på Nonsuicidal self-injury disorder (NSSID) har gjennomgått flere revisjoner (Zetterqvist, 2015). Per nå er NSSID foreslått som en dikotom diagnose (Washburn, Potthoff, Juzwin & Styer, 2015), bestående av seks kriterier hvor alle må bli møtt for å få diagnosen. De nåværende kriteriene på lidelsen er (hentet fra DSM-5, vår oversettelse):

- A) I løpet av det siste året har individet, på fem eller flere dager, utført intensjonal selvpåført skade på overflaten av sin egen kropp av en sort som er sannsynlig å inducere blødning, blåmerker eller smerte (f.eks. kutting, brenning, stikking, slåing, overdreven gniing), med forventningen om at skaden bare vil føre til mindre eller moderat fysisk skade (d.v.s. ingen suicidal intensjon foreligger)
- B) Individet utfører den selvskadende atferden med en eller flere av følgende forventninger:
  - 1) Å oppnå lettelse fra en negativ følelse eller kognitiv tilstand
  - 2) Å løse en interpersonlig vanske
  - 3) Å inducere en positiv følellestilstand
- C) Den intensjonale selvskadingen er assosiert med minst en av de følgende:
  - 1) Interpersonlige vansker eller negative følelser eller tanker, slik som depresjon, angst, spenning, sinne, generalisert ubehag, eller selvkritikk, som inntreffer i perioden rett før den selvskadende handlingen
  - 2) Før handlingen utføres, forekommer en periode med overopptatthet med den intenderte atferden som er vanskelig å kontrollere
  - 3) Tanker på selvskading som inntreffer hyppig, selv når det ikke blir handlet på
- D) Atferden er ikke sosialt akseptert (f.eks. kroppspiercing, tatovering, del av et religiøst eller kulturelt ritual) og er ikke begrenset til hudplukking eller neglebiting.
- E) Atferden eller dens konsekvenser forårsaker klinisk signifikant ubehag eller forstyrrer interpersonlige, akademiske eller andre viktige funksjonsområder
- F) Atferden inntreffer ikke utelukkende under psykotiske episoder, delirium eller rusabstinenser. Hos individer med en nevroutviklingsforstyrrelse, er atferden ikke en del av repetitive stereotypier. Atferden er ikke bedre forklart av annen psykisk lidelse eller medisinsk tilstand (f.eks. psykotisk lidelse,

autismespektrumforstyrrelser, psykisk utviklingshemming, Lesch-Nyhan syndrom, stereotypisk bevegelseslidelse med selvskading, trichotillomania, ekskoriasjonlidelse (hudplukking)).

I den femte versjonen av DSM ble NSSID inkludert i Seksjon III, som en tilstand som trenger videre forskning (American Psychiatric Association, 2013). NSSID ble ikke inkludert som en validert diagnose på grunn av mangelfull forskning, samt lav interrater-reliabilitet i DSM-5s feltstudier (Regier et al., 2013). Mye av forskningen har også belaget seg på selvrapporteringsskjema (Gratz, Dixon-Gordon, Chapman & Tull, 2015), og få studier har undersøkt alle DSM-5-kriteriene (Selby, Kranzler, Fehling & Panza, 2015). De foreslåtte DSM-5-kriteriene for NSSID kan dermed ikke betraktes som fullstendig validerte kriterier, men som et viktig startpunkt i forskning på lidelsen (Selby et al., 2015). Flere har også stilt spørsmålsteget til kriterienes kliniske nytte (Washburn et al., 2015; Zetterqvist, Perini, Mayo & Gustafsson, 2020; Brausch, Muehlenkamp & Washburn, 2016; Buelens et al., 2020), og det er et behov for standardiserte instrumenter som undersøker hvorvidt de nåværende kriteriene skildrer en NSSI-diagnose på en passende og meningsfull måte (Selby et al., 2015).

### 1.1.6 The Clinician-Administered Nonsuicidal self-injury Disorder Index

For å adressere de nåværende begrensningene og fasilitere videreutvikling av forskning på NSSID, utviklet Gratz et al. (2015) et klinisk, strukturert intervju for å vurdere det fulle settet av de diagnostiske kriteriene for NSSID i DSM-5; ”The Clinician-Administered Nonsuicidal self-injury Disorder Index” (CANDI).

Både CANDIs struktur og forskernes tilnærming til å utarbeide måleinstrumentet er basert på ”the Clinician-Administered Posttraumatic Stress Disorder Scale (CAPS) (Blake et al., 1995), et klinisk, strukturert intervju som blir betraktet som gullstandarden på posttraumatisk stresslidelse (Weathers, Keane & Davidson, 2001). I overensstemmelse med CAPS, ønsket Gratz et al. (2015) å utvikle et mål på NSSID som ville (1) samsvare med de nåværende diagnostiske kriteriene i DSM-5, (2) gi både dikotome og dimensjonelle data for hvert symptom og den overordnede lidelsen, (3) være tilgjengelig og brukbar for paraprofesjonelle, og (4) fremvise adekvat reliabilitet og validitet (Watson, 1990; Watson, Juba, Manifold, Kucala & Anderson, 1991).

CANDIs måleegenskaper ble undersøkt gjennom en studie på 107 unge voksne mellom 18 og 35 år. Måling av interrater-reliabilitet ble utført på mer enn 10% av utvalget (n=12) og viste en interrater-reliabilitet med en kappakoeffisient ( $\kappa$ ) på  $\geq .83$  for alle kriterier og på diagnosnivå (Gratz. et al, 2015). Mer spesifikt ble det rapportert en diagnostisk enighet på 100% for kriteriene A, B, C, D og F, og 92% enighet for kriterium E ( $\kappa=.83$ ). Enighet på diagnosnivå ble rapportert å være 92% ( $\kappa=.85$ ) (Gratz et al., 2015). Forskerne fant også støtte til CANDIs konstruktvaliditet (Gratz et al., 2015).

## 1.2 Formål med studien og problemstilling

I Norge har CANDI nylig blitt oversatt til norsk av en forskergruppe fra Nasjonalt senter for selvmord og selvmordsforebygging. Oversettelsen ble gjort i forbindelse med forskningsprosjektet ”Villet egenskade” igangsatt i 2016, hvor hensikten er å undersøke forekomster av og samvariasjon mellom villet egenskade-atferd og oppfylte kriterier for borderline personlighetsforstyrrelse blant ungdom i alderen 12-18 år som er i kontakt med psykisk helsevern for barn og ungdom (Universitetet i Oslo, 2018). Utvalget er på ca. 100 ungdommer som er rekruttert fra BUP Nord og BUP Syd i Oslo.

Den norske versjonen av CANDI er enda ikke reliabilitetstestet, slik at instrumentets måleegenskaper ikke er kjent, og derved ikke dets målesikkerhet. Et viktig moment i å forbedre instrumentets målesikkerhet, er å unngå subjektive fortolkninger av resultatene ved å innføre krav til interrater-reliabiliteten. Interrater-reliabilitet kan defineres som graden av enighet eller konsistens mellom to eller flere ratere med hensyn til et gitt mål (Cohen, Swerdlik & Sturman, 2013, s. 159). Med flere observatører som bruker det samme regelsettet ved skåring, økes instrumentets reproduserbarhet og sikrer høyere grad av objektivitet. Hovedformålet med denne hovedoppgaven er å bidra til at interrater-reliabiliteten av instrumentet blir undersøkt. I tillegg ønsker vi å undersøke CANDIs interne konsistens, samt diskutere våre funn opp mot lignende studier med tanke på kriterienes kliniske nytte.

Problemstillingen vår er således todelt; er interrater-reliabiliteten til den norske versjonen av CANDI tilfredsstillende og har NSSID-kriteriene god klinisk nytte?

Interrater-reliabilitet kan for denne oppgaven defineres som graden de to psykologistudentene skårer likt på item-nivå og kriterienivå i CANDI-intervjuet, ut ifra de ulike svarresponsene fra

ethvert intervjuobjekt med selvskadingsproblematikk som er trukket ut fra målgruppen. Den kliniske nytten til de foreslåtte NSSID-kriteriene defineres som deres evne til å reliabelt identifisere og avgrense en gruppe individer som lider og er svekket av sin NSSI (Zetterqvist, 2020, s. 3). Reliabilitetsdelen av denne definisjonen blir i denne oppgaven dekket av undersøkelse av interrater-reliabilitet og intern konsistens. Kriterienes kliniske nytte vil diskuteres basert på hvor mange som møter de ulike kriteriene og korrelasjoner innad i CANDI, sett opp mot tidligere forskningsfunn. Vi kommer ikke til å benytte oss av andre mål enn CANDI, og kan derfor ikke trekke slutninger angående kriterienes validitet.

### **1.3 Oppbygging og avgrensning av oppgaven**

I teoridelen av oppgaven redegjøres det først for utformingen av klinisk strukturerte intervjuer, samt metoder for å vurdere kvaliteten på et klinisk intervju. Herunder greies det ut om reliabilitet, validitet og korrelasjoner, samt mer spesifikk redegjørelse for de ulike reliabilitets- og korrelasjonskoeffisientene vi har benyttet oss av gjennom denne studien. Dette for å gi et bilde av det teoretiske bakteppet vi har hatt under arbeidet med CANDI, og i arbeidet med denne hovedoppgaven.

I metodedelen vil det først beskrives hvordan, og argumenteres for hvorfor, vi foretok noen endringer i ordlyd og instruksjoner i CANDI-intervjuet. Her beskriver vi også hvordan vi gikk frem i arbeidet med utviklingen av skåringsmanualen. Deretter redegjøres det for utvalget, studiedesignet og etiske betraktninger.

I resultatdelen presenterer vi hovedfunnene fra studien. Det vil gis en kort forklaring på hva hvert kriterium innebærer, og resultatene blir presentert i tekst og tabeller. Diskusjonsdelen vil først ta for seg reliabilitetsresultatene. Dernest vil de ulike kriteriene og korrelasjoner innad i instrumentet diskuteres og sees opp mot liknende forskning. Siste del av diskusjonsdelen vil ta for seg utfordringer knyttet til validitet og generaliserbarhet av resultatene. Oppgaven avsluttes med en konklusjon om funnene.

## **2. Teoretisk bakgrunn**

### **2.1 Klinisk strukturerte intervjuer**

#### **2.1.1 Strukturerte vs. semistrukturerte intervjuer**

Standardiserte, eller strukturerte, kliniske intervjuer består av et sett med spørsmål som skal stilles til en informant med eksplisitte retningslinjer for hvordan informantens responser skal skåres (Frick, Barry & Kamphaus, 2010, s. 254). Disse står i motsetning til de ustrukturerte intervjuene, hvor det er intervjueren selv som avgjør hvilke spørsmål som skal stilles og hva som er akseptable responser fra klienten. Selv om det ustrukturerte formatet gjør det mulig å skreddersy intervjuet til klientens behov, tenderer resultatene og tolkningen av slike intervjuer å være høyst idiosynkratiske til klinikerens som utfører intervjuet. Dette går utover intervjuets reliabilitet, og kan på den måten medføre problemer i forskning (Frick et al., 2010). På motsatt side fører standardiseringen av de kliniske spørsmålene, rekkefølgen på dem og den systematiske skåringen av klientens responser, til en uniformitet som muliggjør direkte sammenligning mellom psykologer, kliniske settinger og diagnostiske grupper (Rogers, 1995).

Selv om de standardiserte intervjuene alle gir noen grad av struktur og retningslinjer for standardisert administrering og skåring, er det betraktelig variasjon i mengden "fritt spillerom" som blir gitt til intervjueren på tvers av de ulike intervjuene (Frick et al., 2010). Mindre struktur gjør det mulig å skreddersy intervjuet mer til klientens responser, men krever en høyere grad av erfaring å administrere og har ofte lavere grad av reliabilitet (Young, O'Brien, Gutterman & Cohen, 1987). Disse mer fleksible standardiserte intervjuene kalles egentlig semistrukturerte intervju, men blir i litteraturen noen ganger bare omtalt under standardisert eller strukturerte intervju. Et semistrukturert intervju som er utformet med et koderammeverk bør inneholde mye struktur (Fylan, 2005, s. 66), og vi vil i denne oppgaven også omtale disse intervjuene under standardiserte/strukturerte intervju, da vi tar utgangspunkt i definisjonen nevnt innledningsvis (Frick et al., 2010). Et standardisert intervju i sin strengeste form kan imidlertid på mange måter sies å være et spørreskjema som blir administrert verbalt (Fylan, 2005, s. 66).

Structured Clinical Interview for the DSM, eller SCID (Spitzer, Williams, Gibbon & First, 1992), er et eksempel på et av de mer fleksible klinisk strukturerte intervjuene. SCID starter med lukkede "ja/nei" spørsmål, etterfulgt av en forespørsel om utdypning slik som "fortell meg mer om det" eller "hvordan var det for deg?". Spesifikke oppfølgingsspørsmål blir inkludert for å klargjøre responser og for å oppnå ytterligere informasjon for å gjøre skåringene. Ofte blir intervjueren også nødt til å legge til sine egne spørsmål for å få nok informasjon til å fylle ut skåringene. Rasjonalen for å tillate en slik klinisk bedømming i å



supplementere og modifisere spørsmålene i intervjuet er at validiteten på intervjuerens skåringer vil øke, selv om det kan gå på bekostning av interrater-reliabiliteten (Spitzer et al., 1992).

Mangelen på fullstendig struktur, kan imidlertid bli balansert av inklusjonen av hvert diagnostisk kriterium, tatt direkte fra DSM (Spitzer et al., 1992). En slik tilgang på diagnostiske retningslinjer kan redusere sjansen for at klinikere gjør ulike bedømmelser til samme data, da tvetydigheter reduseres (Ezpleta et al., 1997). Mesteparten av de vanlig brukte strukturerte intervjuene i dag er designet nettopp for å vurdere diagnostiske kriterier fra en av de nyere versjonene av DSM. Dette knytter vurderingen nært opp til dette klassifikasjonssystemet, og dersom målet til evalueringen er å sette eller klarere en DSM-diagnose, er slike strukturerte intervjuer et viktig vurderingsverktøy. Samtidig kan en slik tett oppkobling mot dette klassifikasjonssystemet være en svakhet når man skal vurdere lidelser med et svakt empirisk grunnlag (Frick et al., 2010).

### 2.1.2 Ulike typer skalaer

Strukturerte intervjuer innebærer en kvantifisering, eller en måling, av responser, noe som gjøres ved bruk av ulike type skalaer. Det finnes flere typer skalaer, og hvilken som benyttes avhenger av operasjonaliseringen av fenomenet som skal undersøkes, men også kvaliteter ved variablene i seg selv.

*Nominal* skala brukes for å markere variabler som ikke har noen kvantitativ verdi. Nominale skalaer er deskriptive kategorier, og de ulike alternativene er gjensidig utelukkende. Ja/nei-spørsmål faller innenfor denne kategorien. Etersom variablene ikke har noen reell kvantitativ verdi, gir det heller ikke mening å utføre matematiske operasjoner på nominale skaler. Antall variabler som faller innenfor hver kategori kan imidlertid telles opp, noe som kan gi informasjon om proporsjoner eller prosent (Cohen et al., 2013). *Ordinal* skala brukes når det er en gitt orden i rekkefølgen på variablene, men forskjellen mellom variablene er ukjent og ikke kvantifiserbar. I tillegg kan forskjellen mellom de ulike variablene være av ulik størrelse. Ordinale skalaer har ikke et absolutt nullpunkt. Derfor kan kun et begrenset antall matematiske operasjoner gjøres på denne typen skala (Cohen et al., 2013). I *intervallskalaer* er både rekkefølgen og forskjellen mellom variablene kjent. På grunn av dette vil det også gi mening å regne ut gjennomsnitt av målinger på en intervallskala. I likhet med ordinale

skalaer, har ikke intervallskalaer et sant nullpunkt, men nullverdi og minusverdier kan fremdeles gi mening (Cohen et al., 2013). I *ratioskalaer* er både rekkefølgen og forskjellen mellom variablene kjent, samtidig som skalaen har et sant nullpunkt. Dette gjør at alle matematiske operasjoner kan gjennomføres på en ratioskala. De to sistnevnte kalles ofte for kontinuerlige skalaer (Cohen et al., 2013).

## 2.2 Metoder for å vurdere kvaliteten på et klinisk intervju

### 2.2.1 Reliabilitet og kilder til varians

Å avgjøre hvor nøye, suksessfullt og effektivt det kliniske strukturerte intervjuet oppnår sine mål, i mange tilfeller å generere en passende diagnose, blir konseptualisert i form av reliabilitet og validitet (Young et al., 1987). Reliabilitet kan forklares som konsistens i målinger (Cohen et al., 2013, s. 146), eller i hvilken grad en måling er mulig å replikere (Koo & Li, 2016). Det antas at reelle forskjeller mellom intervjuobjekter til en viss grad vil være stabile, og derfor bør gi de samme konsistente skårene over ulike målinger (Cohen et al., 2013, s. 146). Dersom et intervju har lav reliabilitet, vil det inneholde mye målingsfeil. Det vil si at variansen i skårene i stor grad skyldes at instrumentet ikke gir konsistente målinger (Cohen et al., 2013).

Klassisk testteori baseres på en antakelse om at observerte skårer ( $X$ ) består av en kombinasjon av en sann skåre ( $T$ ) og målefeil ( $e$ ):  $X = T + e$ . Her reflekterer den sanne skåren reelle ulikheter mellom individene og konstruert som skal måles. Målefeil er variasjon i skårer som skyldes andre faktorer enn de vi ønsker å måle (Cohen et al., 2013).

Denne type målefeil kan deles inn i *systematisk* og *usystematisk* målefeil. Systematiske målefeil er feilkilder i målingene av en variabel, og den er typisk konstant eller proporsjonal i forhold til hva som er antatt å være den sanne verdien av variabelen som blir målt. Måleinstrumentet i seg selv kan ofte være en kilde til systematisk målefeil (Cohen et al. 2013, ss. 146-147). En slik systematisk målefeil i måleinstrumentet vil likevel ikke påvirke den relative posisjonen for de ulike målingene, fordi den samme systematiske målefeilen er til stede i alle målingene. Målefeilen vil på den måten ikke påvirke reliabiliteten i målingene (Cohen et al., 2013). Usystematiske målefeil skyldes derimot tilfeldige, uforutsigbare fluktueringer og inkonsistens i andre variabler i måleprosessen. Slike usystematiske målefeil kan variere fra en testsituasjon til en annen (Cohen et al. 2013, s. 146). Et stort utvalg kan

være med på å redusere effekten av usystematiske målefeil. Videre vil reliabilitetstesting kunne avdekke usystematiske feil, mens systematiske feil vil bli kamouflert som en del av “den sanne skåren”.

Når et instrument har høy reliabilitet, betyr det at sannsynligheten for målingsfeil i instrumentet er redusert til et minimum, og at resultatene til instrumentet i stor grad reflekterer intervjuobjektets sanne skåre. Det finnes flere typer reliabilitetsestimater, som måler ulike former for reliabilitet. Videre er reliabilitet en forutsetning for validitet (Young et al., 1987), som vil bli forklart ytterligere lenger ned i metoddelen.

### 2.2.2 Intern konsistens og interrater-reliabilitet

Ulike reliabilitetsestimater omfatter blant annet test-retest, parallell form, alternativ form og intern konsistens (Cohen et al., 2013). I denne oppgaven vil vi gå nærmere inn på intern konsistens og interrater-reliabilitet. Intern konsistens beskriver i hvilken grad alle itemene i et instrument måler det samme konstruktet, og er derfor koblet til testens reliabilitet. Dersom et underliggende konstrukt skal måles, er det nødvendig at alle spørsmålene eller itemene i instrumentet faktisk måler dette, og ikke noe annet. Intern konsistens gir et bilde av kilder til feilvarians i konstruksjonen av en test (Cohen, 2013). For å sikre validiteten, bør intern konsistens vurderes før et instrument blir benyttet i forskning eller i testing (Tavakol & Dennick, 2011).

Interrater-reliabilitet kan på sin side defineres som graden av enighet eller konsistens mellom to eller flere ratere med hensyn til et gitt mål (Cohen et al., 2013, s. 159). Den gir et mål på kilder til feilvarians i skåring og tolkning av testen (Cohen, 2013). Det gjøres ofte et skille mellom interrater-enighet og interrater-reliabilitet, hvor interrater-enighet representerer i hvilken grad de ulike raterne tenderer å gjøre nøyaktig de samme bedømmelsene til det vurderte subjektet. På motsatt side representerer interrater-reliabilitet i hvilken grad skåringene til ulike ratere er proporsjonale når de blir uttrykt som avvik fra sine gjennomsnitt. I praksis betyr dette at forholdet til et ratet individ til andre ratede individer er det samme, selv om de absolutte numrene som blir brukt for å uttrykke dette forholdet kan variere fra rater til rater (Tinsley & Weiss, 1975).

Dersom et intervju har vist høy interrater-reliabilitet, betyr dette at skårene kan samles inn av flere ulike ratere, som er reliable på det aktuelle skåringssystemet, på en systematisk og samsvarende måte (Cohen et al., 2013). De ulike raterne vil altså i stor grad oppnå den samme informasjonen fra den samme intervjustimulien (Young et al., 1987). Dette krever ofte at instrumentet inkluderer en manual for skåring, og/eller at raterne har fått opplæring i det spesifikke instrumentet.

### 2.2.3 Reliabilitetskoeffisient

En reliabilitetskoeffisient er en indeks som representerer ratioen mellom den sanne skåren i instrumentets varians og den totale variansen (Cohen et al., 2013, s. 145). Det finnes flere måter å regne ut en reliabilitetskoeffisient på, avhengig av hva slags kilde til feilmåling man er ute etter å måle. En koeffisient av interrater-reliabilitet gir informasjon om målingsfeil som er et resultat av skåringen på instrumentet (Cohen et al., 2013, s. 161). For å måle interrater-reliabilitet kan det brukes flere ulike statistiske prosedyrer. I denne oppgaven har vi benyttet oss av *Intraclass correlation coefficient* (ICC) og *Cohens kappa*.

### 2.2.4 Intraclass correlation coefficient

ICC måler forholdet mellom to variabler som måler det samme fenomenet (Field, 2009, s. 678). I mange tilfeller i statistikk tas det utgangspunkt i at data er organisert på ett enkelt nivå. I virkeligheten er det ikke helt slik - veldig mye data er egentlig organisert hierarkisk. For eksempel kan man måle prestasjon på en arbeidsplass. Anta disse arbeiderne deler et kontorlandskap, og det er flere kontorlandskap innen den samme bedriften. Da vil arbeiderne representere ett nivå, kontorlandskapet et annet, og bedriften et tredje nivå (Field, 2009). ICC er en analyse av varians, og den representerer den andelen av den totale variansen i utfallet som kan tilskrives til en klasse (Field, 2009, s. 729). I denne oppgaven vil ICC beregnes for å undersøke hvor mye av variansen i resultatene som kan tilskrives til raterne av CANDI.

ICC er et godt mål å bruke på ordinal, intervall og ratioskala (Hallgren, 2012). Det finnes flere ulike versjoner av ICC, og hvilken som bør benyttes avhenger av designet på studien. Shrout og Fleiss (1979) identifiserte seks ulike former for ICC. Senere utvidet McGraw og Wong (1996) til 10 ulike former for ICC, og det er deres terminologi som beregningene i Statistical Package for the Social Sciences (SPSS) er basert på. De ulike formene for ICC kan gi forskjellige resultater til tross for at de er benyttet på det samme datamaterialet. Dette er fordi

de ulike formene i seg selv bærer med seg forskjellige antakelser, og kan derfor lede til ulike fortolkninger av materialet. Det er følgelig nødvendig å markere hva slags type ICC som er blitt benyttet når ulike reliabilitetsberegninger skal utføres (Koo & Li, 2016).

I denne studien kommer vi til å benytte oss av ICC *two-way mixed effects, consistency, single rater/measurement* (Koo & Li, 2016). *Two-way mixed effects* blir benyttet fordi begge forfatterne av denne oppgaven kommer til å rate alle intervjuene, og vi er også de eneste raterne i denne studien. Vi er med andre ord ikke et tilfeldig utvalg ratere, trukket fra en større gruppe med mulige ratere (Shrout & Fleiss, 1979). *Consistency* representerer at vi søker konsistens i målingene våre og ikke absolutt enighet. *Single rater/measurement* representerer at vi beregner ICC for hver skåre, og således benytter en enkelt rater som grunnlag for våre målinger og ikke benytter gjennomsnittet av våre skårer som grunnlag for våre målinger, noe som ville gitt oss prosentandelen av vår enighet/uenighet på tvers av alle items (Shrout & Fleiss, 1979).

Beregning av ICC two-way mixed effects, consistency, single rater/measurement:

$$ICC = \frac{BMS - EMS}{BMS + (k - 1)EMS}$$

der BMS (*between targets*) =  $n - 1$ , og EMS (*residualvarians*) =  $(n - 1)(k - 1)$ , som begge er kilder til varians.  $n$  representerer utvalget og  $k$  representerer antall ratere (Shrout & Fleiss, 1979). Basert på et 95% konfidensintervall av ICC-estimatet, vil verdier mindre enn .50 indikere dårlig reliabilitet, verdier mellom .50 - .75 vil indikere moderat reliabilitet, verdier mellom .75 - .90 indikerer god reliabilitet, og verdier over .90 vil indikere meget god reliabilitet (Koo & Li, 2016). Konfidensintervallet henviser til at vi med 95% sannsynlighet kan si at den sanne reliabiliteten ligger innenfor et gitt spekter (Cohen et al., 2013, s. 177).

### 2.2.5 Cohens kappa

Cohens kappa kan brukes til beregning av interrater-reliabilitet når datamaterialet er organisert på en nominell eller ordinal skala. En forutsetning for denne beregningen er at kategoriene er gjensidig utelukkende. Beregning av Cohens kappa tar også høyde for at enighet oppnås ved tilfeldigheter (Cohen, 1960).

$$\text{Beregning av Cohens kappa: } \kappa = \frac{P_o - P_c}{1 - P_c}$$

Der  $p_o$  = proporsjonen av enheter der raterne er enige og  $p_e$  = proporsjonen av enheter hvor enighet er forventet ved en tilfeldighet (Cohen, 1960). Når oppnådd enighet tilsvarer enighet som forventes å oppnås ved tilfeldighet, vil  $\kappa = 0$ . Cohen selv foreslo følgende tolkning av enighet ved bruk av Kappa: .01 – .20 som ingen til litt enighet, .21 – .40 som noe enighet, .41 – .60 som moderat enighet, .61 – .80 adekvat enighet, og .81 – 1.00 som høy grad av enighet (McHugh, 2012). Likevel kan det være forskjeller i hvor mye uenighet man kan tolerere, basert på hva man undersøker, og mange tekster foreslår et minimum på .80 for akseptabel enighet (McHugh, 2012). Det er noen begrensninger ved å bruke Cohens kappa. Utrekningen tar for eksempel ikke høyde for grad av uenighet, og slik sett blir målene basert på et prinsipp om “alt eller ingenting”. Derfor kan det lønne seg å bruke vektet kappa, eller ICC dersom man har ordinale skalaer som strekker seg fra “ingen”, “litt”, “moderat”, “høy” og “ekstrem” (Hallgren, 2012), slik som CANDI-instrumentet har.

### 2.2.6 Cronbachs alfa

De statistiske prosedyrene over gir mål på interrater-reliabilitet, mens når det kommer til intern konsistens til en test eller skala, brukes estimatet Cronbachs alfa. Utrekningen av Cronbachs alfa går ut på at dataene deles i to par på alle mulige vis, og deretter beregnes korrelasjonskoeffisienten for hver enkelt av disse parene. Det er altså to ting som beregnes: variansen innen et item, og kovariansen mellom hvert enkelt item og alle andre items i testen (Field, 2009).

$$\text{Beregningen for Cronbachs alfa: } \alpha = \frac{N^2 \times M(\text{COV})}{\text{SUM}\left(\frac{\text{VAR}}{\text{COV}}\right)}$$

Hvor  $N^2$  = kvadraten av alle itemene i testen,  $M(\text{COV})$  = gjennomsnittet av inter-item kovarians, og  $\text{SUM}(\text{VAR}/\text{COV})$  = summen av alle elementene i varians/covarians-matrisen. Alfaverdier strekker seg fra 0 til 1, og som regel vil en alfaverdi mellom .70 og .80 være akseptabel (Field, 2009). En høy alfa-koeffisient betyr imidlertid ikke nødvendigvis en høy grad av intern konsistens, ettersom alfa også blir påvirket av lengden på testen. Dersom lengden på instrumentet er for kort, vil verdien på alfa reduseres (Tavakol & Dennick, 2011). Det er viktig å notere seg at alfa er en egenskap hos skårene på et instrument på et bestemt utvalg av deltakere. Derfor bør forskere måle alfa hver gang instrumentet administreres, og ikke belage seg på tidligere publiserte alfa-estimer (Tavakol & Dennick, 2011).

### 2.2.7 Validitet

For at en test, et eksperiment eller måleverktøy skal vurderes som et godt verktøy, er det nødvendig med *både* høy reliabilitet og validitet. Reliabilitet er en forutsetning for validitet, på den måten at det beskriver i hvilken grad feilvariansen til et instrument begrenser dets validitet (Young et al., 1987). Validitet på sin side, henviser til et estimat av hvor godt et instrument måler det som det er tenkt til å måle. Med andre ord sier validiteten til et instrument noe om hvor nyttig instrumentet er for en spesiell hensikt (Cohen et al., 2013, s. 181).

Det er viktig å notere seg at et instrument aldri er universelt valid, det vil si valid til enhver tid for enhver populasjon. Når det sies at et instrument er valid, er dette innenfor rimelige forventede grenser (Cohen et al., 2013, s. 181). For eksempel vil et klinisk strukturert intervju som har vist seg å være valid for en voksenpopulasjon, ikke automatisk være valid for en ungdomspopulasjon.

Det finnes flere ulike typer mål på validitet, og noen typer validitet kan være mer relevante enn andre, avhengig av bruksområde og type instrument (Cohen et al., 2013, s. 191).

*Konstruktvaliditet* er et validitetsmål som oppnås ved å analysere (a) hvordan skårene på instrumentet relaterer seg til andre skårer og mål, og (b) hvordan skårene på instrumentet kan bli forstått innenfor et gitt teoretisk rammeverk, for å forstå det konstruktet eller trekket som instrumentet var designet for å måle (Cohen et al., 2013, s. 182). Konstruktvaliditet kan anses som et paraplybegrep for validitet, der mange validitetsbegrep faller inn under (Cohen et al. 2013).

### 2.2.8 Korrelasjon

I vurderingen av konstruktvaliditeten til et instrument er det mulig å benytte seg av korrelasjonsanalyser. På den måten kan skårer på instrumentet sammenlignes med andre mål som måler det samme underliggende konstruktet (O'Leary-Kelly & Vokurka, 1998). Vi har i denne studien valgt å bruke korrelasjonsanalyser innenfor det samme instrumentet for å undersøke om det foreligger ulike trender og sammenhenger blant itemene som blir vurdert. Dette kan være hensiktsmessig ettersom itemene CANDI er tenkt å måle ulike aspekter ved det samme fenomenet. Vi vil med denne metoden ikke få et direkte mål på konstruktvaliditet, men kan se eventuelle interne korrelasjoner opp mot eksisterende empiri. Dersom flere studier

indikerer at to variabler korrelerer, styrker dette også antakelsen om at det eksisterer en reell korrelasjon.

En *korrelasjonskoeffisient* er et tall som indikerer styrken på et forhold mellom to ulike variabler. En korrelasjon indikerer kun samvariasjon, og sier ingenting om kausalitet (Cohen et al, 2013). *Pearson r* ( $r$ ) er en korrelasjonskoeffisient som kan brukes når de to variablene som skal måles er kontinuerlige. Den tar hensyn til lineære forhold, som indikerer at om en variabel øker, vil den andre variabelen øke med én konstant sammen med den. *Spearman Rho* ( $\rho$ ) er en annen type korrelasjonskoeffisient som ofte brukes ved små utvalg, og spesielt når begge variablene som måles er organisert ordinalt. Denne formen for korrelasjon kan gi en høyere korrelasjonskoeffisient enn Pearson, dersom forholdet mellom variablene ikke er lineært (Cohen et al., 2013). *Signifikansnivå* ( $p$ ) gir et grunnlag for å kunne konkludere om korrelasjonen kan antas å være reell, altså statistisk signifikant, eller oppnådd ved en tilfeldighet. Det opereres vanligvis med at signifikansnivået må være under .05 eller .01, for å kunne konkludere om en korrelasjon er statistisk signifikant (Cohen et al. 2013).

Korrelasjonskoeffisienter uttrykkes som et tall mellom -1 og 1, og i følge Pallant (2010, s. 134) kan korrelasjonskoeffisienter mellom +/- .10 - .29 klassifiseres som svake, +/- .30 - .49 som moderate og +/- .50 - 1.00 som sterke. Det er dog stor variasjon i kategoriseringen av styrken på korrelasjonskoeffisienter (Akoglu, 2018), men vi velger å forholde oss til til Pallants kategorisering videre i denne oppgaven.

Selv om det vil bli foretatt korrelasjonsanalyser blant items innad i CANDI, er størsteparten av det metodiske arbeidet sentrert rundt interrater-reliabilitetstesting, og fremgangsmåten for denne vil gjennomgås i det følgende.

### **3. Materiale og metode**

#### **3.1 Utvikling av skåringsmanual og revidering av intervjuet**

Før oppstart av reliabilitetstestingen, så vi videoopptak av tre ulike klinikere som administrerte CANDI i forbindelse med hovedstudien. Vi oppdaget at instrumentet ble administrert på ganske ulike måter, med tanke på instruksjoner som ble gitt, oppfølgingsspørsmål, samt i hvor stor grad intervjuet ble administrert som et fullstendig intervju eller som delvis intervju, og delvis selvutfyllingskjema.



For å kunne undersøke interrater-reliabiliteten til intervjuet, vurderte vi det nødvendig å øke strukturen til intervjuet gjennom å legge til tydelige instruksjoner til intervjuer, samt noen tilleggsspørsmål i intervjuet både for at ulike intervjuere skal presentere spørsmålene på samme måte, samt tydeliggjøre hva vi er ute etter for den som bli intervjuet. På samme måte er det også nødvendig at intervjuer har samme forståelse av hva de ulike svaralternativene innebærer. Når det kommer til dette intervjuet, så vi dette som særlig viktig ettersom flere av skåringskalaene er av typen ordinal skala. Den norske versjonen av CANDI har frem til nå blitt administrert på en slik måte at svaralternativene på de ordinale skalaene har blitt lest opp av intervjuer, og intervjuobjektet selv må definere seg på en skala fra 0-4, der en skåre på 0 representerer “ikke eksisterende” og en skåre på 4 representerer “ekstrem”. På den måten har intervjuet i stor grad blitt administrert som et muntlig spørreskjema, og å vurdere interrater-reliabilitet på en slik administreringsform ville vært lite meningsfullt. Vi valgte derfor å selv definere hva de ulike svaralternativene innebærer, og utarbeidet en skåringsmanual med veiledning for skåring av de ulike skalaene. Dette er også i samsvar med administreringen av CAPS (Blake et al., 1995). Operasjonaliseringen av de ulike svaralternativene er basert på videoer av tidligere intervjuer, egne pilotintervjuer, manualen til CAPS (Blake et al., 2000), GAF (Karterud, Pedersen, Løvdahl & Friis, 1998), selvutfyllingsskjemaet Impact of Non-Suicidal Self Injury Scale (INS) (Burke, Ammerman, Hamilton & Alloy, 2017), samt Anne Trana Brager-Larsens erfaring (personlig kommunikasjon, 02.10.19), som er prosjektkoordinator i “Villet egenskade”-prosjektet.

Skåringsmanualen (appendiks C), samt den reviderte versjonen av CANDI (appendiks B) finnes i sin helhet i appendix. Endringene i intervjuet er markert i grønt og vil gjennomgå i de følgende avsnittene.

### 3.1.1 Endringer i ordlyd og instruksjoner

I den opprinnelige engelske versjonen av intervjuet, er det på første side inkludert en liste over ulike selvskadingsmetoder som de i hovedstudien valgte å fjerne (se avsnitt 3.6 Etiske betraktninger). Dette medførte at det ble nødvendig å omformulere flere av spørsmålene på side 2 i CANDI, ettersom disse spørsmålene originalt henviste til denne listen.

Omformuleringene innebærer en endring som gjør at det er mulig å innhente tilnærmet den samme informasjonen, uten å henvise til listen.

I tillegg til dette har vi valgt å inkludere noen setninger med standard informasjon og instruksjoner som skal gis til klienten gjennom intervjuet. Den første av disse er helt i starten av intervjuet, og lyder som følger: “Noen ganger kommer jeg til å stille spørsmål som ligner veldig på hverandre. Det er fordi intervjuet er sånn, og det betyr ikke at du har svart feil den første gangen”. Årsaken er basert på forskning som viser at en del individer er mottakelige for suggesjon ved repeterte spørsmål. Det gjelder både voksne (Gudjonsson, 2003) og barn (Ceci & Bruck, 1999; Siegal, Waters & Dinwiddy, 1988). I en serie med eksperimenter viste blant annet Ceci & Bruck (1999) hvordan gjentatte spørsmål fikk barn til å endre sine opprinnelige svar. Forskere forklarer dette med at barnas endring av svarrespons kan skyldtes at de gjør seg slutninger om testlederens intensjoner og fordi de tror de tidligere svarte feil (Magnussen, 2017). Det er derfor helt sentralt at det tilføres initiale instruksjoner om grunner til at intervjueren eventuelt kommer til å stille samme spørsmål flere ganger (Melinder, 2011).

De to neste instruksjonene vi har inkludert er under *kriterium B* og *CI*, og er lagt til for å presisere for pasienten hva som menes når vi ber om svar oppgitt i prosent. Pasienten blir bedt om å estimere i prosent ulike forventninger til selvskadningen, samt ulike følelser i forkant av selvskadningen. Disse prosentandelene trenger imidlertid ikke gå opp i 100 %, da det kan foreligge flere årsaker til at man selvskader, på samme måte som flere følelser kan være tilstede samtidig. Da vi så gjennom tidligere videoklipp av administrering av intervjuet, oppdaget vi at det var viktig å presisere dette, da noen automatisk begynte å legge sammen prosent for å få tallene til å gå opp.

Flere av de andre setningene vi har valgt å inkludere, er for å presisere forskjellen mellom de ulike spørsmålene. Årsaken er at mange av spørsmålene ligner, samtidig som de har viktige forskjeller som vi vurderte det nødvendig å understreke.

Når det kommer til den delen av intervjuet som måler subjektivt ubehag og svekkelse (E-kriteriet), valgte vi å skrive forslag til åpne oppfølgingsspørsmål som “fortell mer”, “hvordan da” og “kan du utdype”. Brager-Larsen (personlig kommunikasjon, 02.10.19) forklarte at flere ungdommer umiddelbart ofte ikke oppgir å kjenne på ubehag og svekkelse, men at det med ytterligere oppfølgingsspørsmål kan komme frem mer ubehag og svekkelse enn det som opprinnelig blir rapportert. Dette er også i tråd med forskningslitteraturen som har pekt på at måten ubehag og svekkelse blir konseptualisert og målt på vil kunne ha betydning for hvor mange som møter kriteriet (Brausch, 2019). CANDI er et eksempel på et instrument som

måler ubehag og svekkelse ganske direkte; spørsmålene er altså formulert på en slik måte at det er personens egen oppfatning av ubehag og svekkelse som blir vurdert. Wilkinson og Goodyer (2011) har problematisert det faktum at NSSI ofte blir sett på som en løsning som reduserer ubehag, heller enn at det forårsaker det. Vi ønsket å vurdere E-kriteriet direkte, men vurderte at ytterligere åpne oppfølgingsspørsmål kunne være hjelpsomt både for intervjuer og intervjuobjekt, for å undersøke grad av svekkelse og ubehag nærmere.

### 3.1.2 Definerings av skåringer

Når det kom til skåringen av kriterium E, så vi det nødvendig å være ekstra nøye med å tydeliggjøre instruksjoner i skåringsmanualen gjennom forslag til svaralternativ og hvordan disse skulle skåres. Hovedårsaken til dette var at det i Gratz et al. (2015) sin studie viste seg at kriterium E var det kriteriet med lavest interrater-reliabilitet. Etersom det ikke er mulig å komme med svareksempler som vil kunne dekke alle mulige svarresponser, valgte vi å skrive generelt hva som definerte en skåre for “ikke eksisterende”, “mild”, “moderat”, “sterk” og “ekstrem”. Dette gjorde vi ikke bare for kriterium E, men gjennomgående for hele intervjuet.

Når det gjaldt skårene for E-kriteriet, valgte vi å skille en skåre på moderat (2), sterk (3) og ekstrem (4) med grad av funksjonstap. Videre definerte vi at en skåre på mild (1) innebærer at det er tilstedeværelse av endring, men at den ikke er særlig merkbar for individet selv og at det ikke foreligger noen form for funksjonstap. En skåre på ikke eksisterende (0) innebærer at individet ikke rapporterer noen som helst form for ubehag eller svekkelse.

Vi fulgte samme prinsipp for skårer på 0 og 1 når vi operasjonaliserte de andre ordinale skalaene i intervjuet. Under C2 hvor spørsmålene dreier seg om hvor påtrengende tankene er, skilte vi en skåre på 2, 3 og 4 på i hvor stor grad intervjuobjektet klarer å skyve tankene bort hvis de ønsker det (2), prøver å skyve tankene bort uten å klare det (3) eller vet at de ikke klarer å skyve tankene bort, og dermed ikke engang forsøker (4). Det samme skillet ble gjort under C3 hvor spørsmålene dreier seg om trang. Dette kriteriet måler et av nøkkeltrekkene ved NSSID (Gratz et al., 2015; Lengel & Mullins-Sweatt, 2013) som er overopptatthet med NSSI som er *vanskelig å kontrollere*. Å skille intensitetsskårene på i hvilken grad man klarer å styre oppmerksomheten bort fra tanker/trang om selvskading, var dermed et forsøk på fange opp dette kontrollaspektet.

Vi har også valgt å tydeliggjøre hvordan utsagn uten tydelige generelle mønstre, eller store variasjoner innen det siste året skal skåres. Når det gjelder førstnevnte, skal det etterspørres hvordan deltakeren opplever det hyppigst. Når det gjelder deltakere som rapporterer at selvskadningen har endret seg mye i løpet av det siste året, skal det tas utgangspunkt i de siste tre månedene. En cut-off på tre måneder ble valgt for å sikre at endringen har vedvart over noe tid, og ikke er en plutselig endring som kamuflerer en generell tendens.

### 3.1.3 Endring i skåringer

Vi har gjort endringer i terskelnivå for hva som skal til for å tilfredsstille kriteriet C2 (g). I forbindelse med prosjektet “Villet egenskade” har Brager-Larsen i sine intervjuer operert med svar på 80% eller høyere for å tilfredsstille dette kriteriet (personlig kommunikasjon, 02.10.19). Vi har valgt å redusere denne terskelen til 50%, for at den skal samsvare med resten av C2-kriteriets terskelnivåer for tilfredsstilling av kriteriet. Dette er på bakgrunn av at de andre spørsmålene under C2 krever “moderat overopptatthet” og at tankene om selvskadning er “moderate, vanskelige eller ekstremt vanskelige å motstå eller tolerere”.

### 3.1.4 Markeringer

Vi har foreslått endringer for å markere ulike deler av testen, og på den måten tydeliggjøre forskjeller på *instruksjoner til pasienten, instruksjoner til intervjueren, spørsmål og mulige oppfølgingsspørsmål som skal stilles til pasienten og avkrysningsbokser for å markere om diagnosekrav er vurdert som møtt eller ikke*. Disse markeringene presenteres i skåringsmanualen og innebærer ulike typer bakgrunnsfarger, markeringer i form av piler, og også egenskaper ved skriften; blokkbokstaver, kursiv og uthevet font. Tanken bak endringene var å gjøre det lettere og tidsbesparende for testleder å gjennomføre intervjuet.

## 3.2 Bakgrunn for valg av deltakere

Alle deltakerne som inngår i denne studien er rekruttert gjennom Oslo Universitetssykehus, henholdsvis gjennom klinikere fra BUP Oslo Nord og BUP Oslo Syd. 18 deltakere er rekruttert gjennom hovedstudien “Villet egenskade”, mens ytterligere syv deltakere ble rekruttert når denne delstudien ble igangsatt. Deltakerne fra Villet egenskadeprosjektet hadde allerede gjennomgått et førstegangsintervju (T1) og ble informert om denne studien når de ble kontaktet for andregangsintervjuet (T2). Noen av deltakerne hadde allerede gjennomført T2. De ble kontaktet per telefon og invitert til å komme tilbake for å delta i denne studien. De

øvrige syv deltakerne ble informert om studien via sin behandler i nevnte psykiatriske enheter.

### **3.3 Utvalget**

#### **3.3.1 Inklusjonskriterier**

Deltakere i alderen 12-18 år som har rapportert at de har selvskadet  $\geq 5$  ganger i løpet av de siste to årene. I utgangspunktet ønsket vi kun å inkludere deltakere som hadde selvskadet  $\geq 5$  i løpet av det siste året, men ettersom noen deltakere fra hovedprosjektet rapporterte ved T2 at de hadde sluttet å skade seg selv siden T1, bestemte vi oss for å utvide denne terskelen. Vi vurderte at dette ikke ville gå ut over vår interater-reliabilitet, da vi fremdeles kunne vurdere deres svarresponser i tråd med skåringsmanualen. For de deltakerne som ikke hadde selvskading i løpet av siste år, baserte administreringen av CANDI seg på de siste to årene. For deltakerne som hadde selvskading i løpet av siste år, ble administreringen av CANDI basert på det siste året.

#### **3.3.2 Eksklusjonskriterier**

Deltakere fra hovedprosjektet som hadde fylt 19 år på ved tidspunkt for T2 ble ekskludert fra våre analyser (n=4). På denne måten kan funnene i større grad generaliseres til ungdomspopulasjonen (12-18 år). De ekskluderte intervjuene ble brukt som treningsintervjuer, der vi øvde oss på å skåre ut intervjuene. Skåringene på disse intervjuene er ikke medregnet i resultatet.

#### **3.3.3 Endelig utvalg**

Totalt ble 25 ungdommer mellom 14-19 år intervjuet. Fire intervjuer ble ekskludert på grunn av alderskriteriet, som nevnt over. Et intervju ble ekskludert grunnet feil i gjennomføringen av selve intervjuet. Et siste intervju ble ekskludert grunnet at deltakeren ikke oppfylte inklusjonskriteriet om selvskading i løpet av de siste to årene, på intervjutidspunktet. Ingen av deltakerne har valgt å trekke seg fra studien. Det endelige utvalget ble dermed på 19 ungdommer, hvor den ene av oss har gjennomført ni intervjuer, og den andre har gjennomført 10 intervjuer. Av de 19 ungdommene i det endelige utvalget, var 18 jenter og én gutt. Gjennomsnittsalderen på deltakerne er 16,4 år.

### **3.4 Studiedesign og prosedyrer**

### 3.4.1 Videofilming av intervjuene

Intervjuene ble foretatt på kontorer i BUP Nord og BUP Syd i Oslo. Samtlige intervjuer ble tatt opp på video. Gjennomføringen av intervjuene og filmingen tok i gjennomsnitt 37 minutter og 44 sekunder. Videofilmene ble oppbevart i henhold til Oslo Universitetssykehus forskningsetiske retningslinjer. Deltakerne som allerede var en del av hovedprosjektet samtykket til filming gjennom “Villet-egenskade”-prosjektet, men ble også spurt verbalt om samtykke til filming før intervjuet ble satt i gang. Deltakerne som kun var med i delprosjektet samtykket til videoopptak gjennom samtykkeskjemaet (se appendiks D). Deltakerne fikk tilbud om at kameraet ble vendt bort, slik at vi fikk lydopptak, men at deltakerne selv ikke var synlige i videofilmen. Noen deltakere valgte å takke ja til dette.

### 3.4.2 Blinding

Etter vi var ferdige med intervjuene, observerte vi samtlige videoopptak og skåret ut intervjuene separat. Vi var blindet for hverandres skåringer, og diskuterte ikke våre resultater med hverandre eller med veiledere og prosjektmedarbeidere i denne prosessen. Observasjonen av videoopptakene ble gjort på samme kontor, men på hver vår skjerm og ved bruk av headset.

### 3.4.3 Randomisering av rekkefølge

Samtlige intervjuer er skåret ut i etterkant, ved hjelp av videoopptakene. Da vi skulle skåre ut intervjuene ble alle videoopptak randomisert i rekkefølge for hver av oss, for å unngå overføringseffekter. Overføringseffekt er når en tidligere behandling eller eksperimentsituasjon påvirker atferden i en senere behandling eller eksperimentsituasjon (Bordens & Abbott, 2011). I vårt tilfelle vil dette referere til at erfaring fra et intervju eller skåring kan påvirke neste intervju eller skåring. Vi ønsket også å minske effektene av overføringseffekten *utmattelse*, som kan føre til slurv og dårligere vurderinger. For å unngå dette observerte vi videofilmene over flere dager.

## 3.5 Statistisk analyse og bearbeiding

### 3.5.1 Databehandling

Alle dataene ble lagt inn i SPSS, versjon 25 og koblet til et ID-nummer.

Cohens Kappa ble brukt til å beregne reliabilitetskoeffisienten for nominale data, hvilket inkluderer diagnosekriteriene. ICC ble brukt til å beregne reliabilitetskoeffisienten for ordinale data og ratiodata. Cronbachs alpha er benyttet for å beregne intern konsistens. Pearson r og Spearman Rho ble brukt for å beregne ulike korrelasjoner. Før vi beregnet korrelasjonene regnet vi om våre to skårer per deltaker, til en gjennomsnittsskåre per deltaker. For noen av analysene ble det beregnet en sumskåre for de ulike kriteriene.

### **3.6 Etiske betraktninger**

Ungdommene vi møtte i forbindelse med dette prosjektet er generelt i en sårbar fase av livet, og spesielt i en sårbar situasjon med tanke på underliggende problematikk. Vi etterstrebet dermed å møte ungdommene på en så varm og respektfull måte som mulig.

#### **3.6.1 Har forskning på NSSI en iatrogen effekt?**

Det er flere studier som indikerer at selvskading kan ha en smitteeffekt, både hos innlagte pasienter (Matthews, 1968) og i den øvrige befolkningen (Muehlenkamp, Hoff, Licht, Azure & Hasenzahl, 2008). Et etisk aspekt ved dette prosjektet er om vurdering av NSSI kan føre til økt selvskadingsatferd, eller inspirere til flere metoder for selvskading. Av etiske grunner er det derfor valgt å gjøre en endring i hovedprosjektets bruk av den norske oversettelsen av CANDI, fra den engelske versjonen. Denne endringen har vi også forholdt oss til. Endringen innebærer å fjerne en liste over mulige selvskadingsmetoder. I den engelske versjonen av CANDI er denne med for å gi eksempler på hva slags type atferd det siktes til ved spørsmål om selvskadingsatferd. Endringen er gjort for å redusere den eventuelle påvirkning en slik liste kan ha på allerede sårbare ungdommer, og for å unngå å gi ideer til måter å skade seg selv på.

På tross av at denne listen ble fjernet, er intervjuet fortsatt sensitivt med tanke på innhold, og med tanke på at det ble utført på en sårbar gruppe. En undersøkelse gjennomført av Whitlock, Pietrusza og Purington (2013), viste imidlertid at mesteparten av deltakerne rapporterte nøytrale eller positive følelser knyttet til gjennomføringen av en spørreundersøkelse som omhandlet selvskading, suicid og psykologisk stress. 2.7% av deltakerne rapporterte negative erfaringer med gjennomføringen av undersøkelsen. Forskerne konkluderte med at deltakelse i NSSI-forskning midlertidig kan øke følelsen av ubehag, særlig for deltakere med en tidligere

relevant historie. Samtidig bidro undersøkelsen også til mer positiv selvrefleksjon og hjelpesøkende atferd.

Muehlenkamp, Swenson, Batejan og Jarvi (2014) fant ingen iatrogene effekter ved detaljerte spørsmål om NSSI. Samlet uttrykte deltakerne lite ubehag og økte positive følelser etter gjennomføringen av undersøkelsen. Mange rapporterte også glede over å ha bidratt til forskning. Resultatene gjaldt både individer som rapporterte en historie med NSSI og individer som ikke hadde en historie med dette. Et viktig funn er også at individer som tidligere hadde drevet med NSSI rapporterte en signifikant nedgang i trang til å utføre selvskading, etter gjennomføringen av spørreundersøkelsen.

Begge de nevnte studiene er dog gjennomført over internett. Det er mulig at opplevelsen av detaljert utspørring om NSSI oppfattes annerledes når den blir gjort ansikt til ansikt. Resultatene fra ovennevnte studier er likevel i tråd med forskning som viser at vurdering og kartlegging av selvmordstanker og -planer hos ungdom ikke fører til flere suicidforsøk (Gould et al., 2005).

### 3.6.2 REK-godkjenning og oppbevaring av sensitivt datamateriale

Hovedprosjektet «Villet egenskade» er godkjent av REK (Søknadsnr. 2016/617, Villet egenskade hos ungdom 12-18 år). For å fullføre vår del av prosjektet, søkte vi om en utvidelse av den opprinnelige REK-godkjenningen. Vi mottok godkjenning på denne søknaden 28.10.19. Videre utviklet vi et informasjonsskriv til de nyrekrutterte deltakerne (se appendiks E), som er godkjent av REK. Alle opplysningene deltakerne gir er anonymisert og datamaterialet oppbevares hos Oslo Universitetssykehus, i tråd med forskningsetiske standarder for håndtering av sensitivt datamateriale.

### 3.6.3 Informert samtykke

De fleste deltakerne i prosjektet var allerede rekruttert og deltagende i hovedprosjektet “Villet egenskade” (n=18). Disse ungdommene og deres foreldre hadde allerede gitt skriftlig informert samtykke til deltakelse i hovedprosjektet. Etersom vårt prosjekt er et delprosjekt av hovedprosjektet “Villet egenskade”, samt at vår godkjenning er en utvidet REK-godkjenning av den opprinnelige REK-godkjenningen til hovedprosjektet, innhentet vi ikke nytt skriftlig samtykke fra de allerede rekrutterte ungdommene. Mange av ungdommene ble intervjuet av



oss under T2 i “Villet egenskade”-prosjektet. Noen av ungdommene hadde allerede gjennomført T2. Disse ble kontaktet per telefon og invitert til å komme tilbake for å gjennomføre CANDI-intervjuet en gang til. Alle ungdommene fikk muntlig informasjon om delprosjektet, og samtykket muntlig.

Noen av deltakerne ble rekruttert til delprosjektet fra BUP Oslo Syd og BUP Oslo Nord, uten at de hadde vært deltakende i hovedprosjektet (n=7). Samtlige av disse ungdommene var i individual- og gruppeterapi på det aktuelle tidspunktet. Disse deltakerne fikk informasjon om delprosjektet via sine behandlere, og både foreldre og ungdommen ga først muntlig samtykke til deltakelsen. Etter muntlig samtykke ble de invitert til intervju, hvor de fikk mer inngående informasjon om både hovedprosjektet og delprosjektet, samt informasjon om anonymisering, taushetsplikt, oppbevaring av sensitivt datamateriale og rettighetene de har som deltakere. Denne informasjonen fikk de både muntlig og skriftlig (se appendiks E). Skriftlig samtykke ble innhentet fra både foreldre og ungdommen. Ifølge helseforskningsloven §17 (2008) kan ungdom over 16 år selv samtykke til deltakelse i forskning som ikke innebærer legemsinngrep eller legemiddelutprøving. Ettersom “Villet egenskade”-prosjektet og vårt metodeprosjekt undersøker svært sensitive temaer, ble det likevel vurdert at det var nødvendig å innhente informert samtykke fra både ungdommen og foreldrene, selv når de var over 16 år.

#### 3.6.4 Informasjon til ansvarlig behandler

Ungdommene som kun var deltakende i delprosjektet fikk tilbud om at informasjonen som kom frem gjennom intervjuet kunne bli delt med deres behandler, tilsvarende tilbudet som ungdommene i hovedprosjektet har fått. På den måten kan deltakelse i prosjektet komme pasient og behandler til gode i videre behandling. Noen av deltakerne ønsket dette, og deres behandler fikk tilbakemeldinger fra intervjuet over telefon. Noen av disse ungdommene brukte også tid i forkant og/eller i etterkant på å snakke med sin behandler om deltakelsen i prosjektet.

#### 3.6.5 Pasientsikkerhet og forebygging av suicidal atferd

I forbindelse med hovedprosjektet ble det utviklet en protokoll for å ivareta pasientsikkerheten og forebygging av suicidal atferd (se appendiks F). Protokollens hensikt er å gi ungdommene som deltar i studien optimal sikkerhet og beskyttelse mot suicidalitet gjennom deltakelsen av studien. Videre er protokollen i tråd med Nasjonale retningslinjer for

selvmordsforebygging i psykisk helsevern (Sosial og helsedirektoratet, 2008). Deltakerne i hovedprosjektet har fått en grundig utredning i risiko for suicidalitet før de ble inkludert i studien. De nyrekrutterte deltakerne til delprosjektet har ikke fått en like grundig utredning når det gjelder dette. Disse deltakerne kan likevel vurderes som ivaretatt ut ifra helseloven (Helse- og omsorgstjenesteloven, 2011), da alle var under behandling i BUP på intervju tidspunktet. Utover dette har vi forholdt oss til samme protokoll for pasientsikkerhet og forebygging av suicidalatferd.

## 4. Resultater

*Se appendiks B (CANDI) for mer inngående informasjon om de ulike spørsmålene og kriteriene.*

### 4.1 Interrater-reliabilitet

*Resultater med absolutt enighet vil i det følgende bli presentert som tekst. De resterende resultatene vil bli presentert i tabeller.*

#### 4.1.1 A-kriteriet

Spørsmålene under A-kriteriet omfatter generell selvskadingsatferd og -hyppighet, samt spørsmål om suicidintensjon.

**Tabell 1**

*Cohens Kappa og ICC (consistency) for A-kriteriet*

Spørsmål	Kappa	ICC
1	1.00	
2		1.00
3	1.00	
4		1.00
5	.922	
6	1.00	

Spørsmål	Kappa	ICC
7	1.00	
8	1.00	
A-kriteriet	1.00	

#### 4.1.2 D-kriteriet

Spørsmålene under D-kriteriet innebærer å utelukke sosialt aksepterte grunner for skading av kroppsvev.

Dette kriteriet inneholder to spørsmål, samt et spørsmål om D-kriteriet er oppfylt. Vi oppnådde en  $\kappa = 1.00$  på samtlige av disse.

#### 4.1.3. B-kriteriet

Spørsmålene under B-kriteriet dreier seg i hovedsak om å kartlegge selvskadingens funksjoner. Dette kan være forventninger om å redusere eller få lettelse fra negative følelser, eller forventninger om å kunne påvirke andre mennesker eller sin situasjon. Informanten blir bedt om å angi tilstedeværelse og hyppighet av ulike typer forventninger, før de skader seg selv.

Vi oppnådde en absolutt enighet på dette kriteriet ( $\kappa = 1.00$ ). Nærmere bestemt viste resultatene på itemnivå for B-kriteriet en  $\kappa = 1.00$  på samtlige nominale spørsmål, og ICC = 1.00 for samtlige spørsmål på intervallskala.

#### 4.1.4 C1-kriteriet

C1-kriteriet omfatter spørsmål om følelser som foreligger i forkant av selvskadingen. Informanten blir bedt om å angi tilstedeværelse og hyppighet av ulike typer følelser før de skader seg selv.

Vi oppnådde en absolutt enighet på C1-kriteriet ( $\kappa = 1.00$ ), og  $\kappa = 1.00$  på samtlige nominale spørsmål på item-nivå, og ICC = 1.00 for samtlige spørsmål på intervallskala på item-nivå.

#### 4.1.5 C2-kriteriet

Spørsmålene som går under C2-kriteriene fanger opp informantens eventuelle overopptatthet av selvskading. Nærmere bestemt hvor mye tid som brukes på å tenke på selvskading, samt grad av kontroll og intensitet på disse tankene.

## Tabell 2

*Cohens kappa og ICC (consistency) for C2-kriteriet*

Spørsmål	Kappa	ICC	95% Konfidensintervall
1	1.00		
2		.987	.966 - .995
3		.946	.866 - .979
4		.942	.857 - .977
5		.957	.829 - .983
6		1.00	-
7		.978	.944 - .992
8		1.00	-
9		1.00	-
C2-kriteriet	1.00		

### 4.1.6 C3-kriteriet og C-kriteriet totalt

C3-kriteriet har spørsmål som dreier seg om selvskadingstanker og -trang som opptrer selv om det ikke handles på tankene og/eller trangen. Spørsmålene prøver å fange opp hyppighet, varighet, intensitet og grad av evne til å motstå denne type tanker/trang. C-kriteriet totalt er oppfylt dersom en eller flere av C1, C2 eller C3 er fylt.

Også her oppnådde vi en ICC = 1.00 for samtlige spørsmål under C3-kriteriet. På spørsmålet om C3-kriteriet er oppfylt viste resultatene en  $\kappa = 1.00$ . For C-kriteriet totalt oppnådde vi en  $\kappa = 1.00$ .

#### 4.1.7 E-kriteriet

E-kriteriet fanger opp i hvor stor grad selvskadingen påvirker og svekker egen fungering, familiefunksjon, sosial funksjon og funksjon i skole, jobb og andre viktige områder i livet.

**Tabell 3**

*Cohens kappa og ICC (consistency) for E-kriteriet*

Spørsmål	Kappa	ICC	95% Konfidensintervall
1		.921	.806 - .969
2		.929	.825 - .972
3		.934	.837 - .974
4		.861	.675 - .944
E-kriteriet	1.00		

#### 4.1.8 F-kriteriet og nåværende NSSID

F-kriteriet dreier seg om å utelukke at selvskadingen bedre kan forklares av andre lidelser eller diagnoser. Til slutt vurderes det om alle kriteriene er oppfylte, og følgelig om kriteriene for nåværende NSSID er oppfylt.

På F-kriteriet oppnådde vi en  $\kappa = 1.00$  for samtlige spørsmål på itemnivå og på kriterienivå. Vi oppnådde også en  $\kappa = 1.00$  på vurderingen om kriteriene for NSSID er møtt.

## 4.2 Intern konsistens

For å beregne intern konsistens har vi benyttet Cronbachs alpha. Beregningen viste en akseptabel intern konsistens med  $\alpha = .72$ , for hele instrumentet.

## 4.3 Korrelasjoner

Vi har benyttet oss av både Pearson  $r$  ( $r$ ) og Spearman Rho ( $\rho$ ) i undersøkelsen av ulike korrelasjoner. Pearson er brukt på intervallskalaer, dikotome variabler og sumskårer, og Spearman er brukt på ordinalskalaer.

#### 4.3.1 Korrelasjoner med oppfylt diagnose

Oppfylt diagnose korrelerer sterkt med sum antall følelser som er oppgitt (C1),  $r = .561$ ,  $p \leq .013$ . Oppfylt diagnose korrelerer også sterkt med sumskåre av funksjonssvikt (E-kriteriet),  $r = .561$ ,  $p \leq .012$ . Sterkest korrelasjon er det mellom forstyrret familiefungering og oppfylt diagnose ( $\rho = .467$ ,  $p \leq .044$ ), og forstyrret sosial fungering og oppfylt diagnose ( $\rho = .687$ ,  $p \leq .001$ ). Antall funksjoner møtt (B-kriteriet) målt opp mot oppnådd diagnose, var rett over terskelen for klinisk signifikansnivå ( $r = 0.454$ ,  $p \leq 0.051$ ).

#### 4.3.2 Korrelasjoner med E-kriteriet

Sum av antall funksjoner møtt (B-kriteriet) korrelerer sterkt med oppfylt E-kriterium,  $r = .673$ ,  $p \leq .002$ . Sum antall følelser (C1-kriteriet) har en veldig sterk korrelasjon med oppfylt E-kriterium,  $r = .884$ ,  $p \leq .000$ . Sumskåre av styrke og varighet på selvskadingstrang uten at man handler på trangen (C3-kriteriet), korrelerer også sterkt med sumskåre av E-kriteriet,  $r = .620$ ,  $p \leq .014$ .

#### 4.3.3. Korrelasjoner med selvskadingsmetoder

Antall selvskadingsmetoder korrelerer moderat med sumskåre av overopptatthet (C2),  $r = .468$ ,  $p \leq .044$ . Antall selvskadingsmetoder korrelerer også sterkt med antall selvskadingsepisoder,  $r = .549$ ,  $p \leq .015$ .

### 4.4 En oversikt over kohorter

Under følger en kort oversikt over antall kriterier som var oppfylt for utvalget vårt, samt en kort oversikt over fordelingen av ulike selvskadingsmetoder.

#### Tabell 4

*Antall deltakere som møter de ulike kriteriene*

Kriterier	Deltakere som møter kriteriene
A-kriteriet	89,5% (n=17)
B-kriteriet	100% (n=19)
C1	100 % (n=19)
C2	78,9% (n=15)
C3	73,7% (n=14)
C-kriteriet	100% (n=19)
D-kriteriet	100% (n=19)
E-kriteriet	73,7% (n=14)
F-kriteriet	94,7% (n=18)
Oppfylt NSSID	57,9% (n=11)

Kutting er den selvskadingsmetoden som rapporteres hyppigst, hvor 94,7% (n=18) rapporterer dette som minst én av metodene. Mer spesifikt rapporterer 52,6% (n=10) kun kutting, og 5,3% (n=1) rapporterer kun slåing. Av deltakere som rapporterer to eller flere selvskadingsmetoder er det 15,8% (n=3) som rapporterer at de kutter og klorer/skraper seg, 5,3% (n=1) som kutter og slår, samt 21,1% (n=4) rapporterer at de benytter seg av mer enn to selvskadingsmetoder, inkludert kutting.

## 5. Diskusjon

### 5.1 Resultatene

#### 5.1.1 Interrater-reliabilitet og intern konsistens

Hovedformålet med denne studien var å undersøke interrater-reliabiliteten til CANDI, og resultatene her var svært lovende med en absolutt enighet på kriterienivå for alle kriteriene, og på diagnosnivå. Videre ble resultatene oppnådd med begrenset trening og gjennomført innenfor en rimelig tidsramme, noe som også gir støtte til gjennomførbarheten til intervjuet. Disse resultatene er også i tråd med Gratz et al. (2015) sine funn. I tillegg oppnådde vi en adekvat intern konsistens på  $\alpha = .72$ . Dette samsvarer godt med Gratz (2015), som fant en  $\alpha = .71$ .

Selv om resultatene viste at raterne var helt enige på kriterienivå, viste resultatene på item-nivå noe diskrepans blant raterne, hvor itemene under E-kriteriet var målene det ble oppnådd lavest interrater-reliabilitet på. En av årsakene til dette kan være kravet om å sikre seg at økt subjektivt ubehag, forstyrret familie-/sosial fungering og svekkelse i arbeid-/skoleprestasjoner, skyldes selvskadingen i seg selv og ikke andre lidelser. I virkeligheten kan dette argumenteres for å være et kunstig skille, gitt NSSIs høye forekomst med andre diagnoser (Nock et al., 2006). Flere av deltakerne i vår studie rapporterte også andre psykiske problemer ved siden av NSSI. Under våre intervjuer ba vi deltakeren knytte spørsmålene under E-kriteriet til selve selvskadingen og ikke til andre psykiske plager. Selv om dette skillet ble gjort, utgjorde dette skillet fortsatt en utfordring i noen tilfeller både for deltaker og rater. For eksempel kunne det bli rapportert om store forstyrrelser i skolegang med oppstart rundt samme tid som selvskadingen, men at disse forstyrrelsene skyldtes andre plager enn selvskadingen. Hvordan slike tilfeller ble vurdert, varierte noe blant raterne. I tillegg ble det ofte gitt svar som falt under flere svarkategorier. Selv om skåringsmanualen har retningslinjer for dette, var slike tilfeller ofte opphav til forvirring, noe som kan ha økt variansen blant raterne.

Ved siden av E-kriteriet hadde vi størst variasjon i skårer på to av itemene under C2-kriteriet. En av utfordringene ved skåringen av dette kriteriet, var at flere av deltakerne rapporterte en stor variasjon innen det siste året, samt stor variasjon blant de ulike selvskadingsepisodene. Dette medførte en utfordring både under selve intervjuet, samt for skåringen etterpå, og kan ha vært opphav til ulike typer varians blant raterne. Raterne stilte for eksempel tidvis ulike oppfølgingsspørsmål og la vekt på ulike utsagn under skåringen. Når det kommer til spørsmålene under kriterium C3, oppnådde vi imidlertid en absolutt enighet, noe som tyder på at spørsmålene under dette kriteriet ikke var like sårbare for variasjoner og/eller at vi var mer presise i spørsmålsformuleringen og i bruken av skåringsmanualen i etterkant.

Når det gjelder kriterium B, C1, D og i stor grad også F-kriteriet, ble disse administrert som et muntlig spørreskjema, og vi oppnådde dermed absolutt enighet på disse kriteriene. Et intervju med en slik kombinasjon av flere strukturerte spørsmål, har utvilsomt bidratt til at vi fikk en absolutt enighet på kriterie- og diagnosenivå. På tross av dette oppnådde vi imidlertid også svært god interrater-reliabilitet på de semistrukturerte spørsmålene, noe som kan tyde på at vi



ved utarbeidelse av skåringsmanual har klart lyktes i å fange opp viktige skiller i alvorlighetsgrad.

Videre, når det kommer til varians i resultatene på item-nivå, kan disse stamme fra utfordringer knyttet til spørsmålenes semistrukturerte natur, intervjuets retrospektive design, begrenset intervjutrening i forkant, samt utforming og bruk av skåringsmanualen. De semistrukturerte spørsmålene oppfordrer for eksempel til oppfølgingsspørsmål som kan være ulike blant ulike intervjuere. Dette, i kombinasjon med et intervju som går langt tilbake i tid og kan gi store variasjoner i svaralternativ, kan skape opphav til ulike former for varians blant intervjuerne, både systematiske og usystematiske. Begrenset trening og er også en faktor som kan ha økt diskrepans blant raterne, noe som vil utdypes ytterligere senere i diskusjonsdelen.

I tillegg var skåringsmanualen mangelfull når det gjaldt å ta høyde for svar som gikk under flere svarkategorier og variasjoner blant de ulike selvskadingsepisodene. Relatert til dette, ble skåringsmanualen også utarbeidet med begrenset erfaring fra møter med selvskadende ungdom. Selv om vi tilstrebet å gjøre skåringsmanualen så realistisk som mulig, basert på videoklipp av tidligere administrerte intervjuer og egne pilotintervjuer, var tilgangen til slikt materiale begrenset, noe som følgelig ga begrensede svareksempler som var basert på faktiske utsagn. Stänicke (2019, s. 51) har problematisert det faktum at forhåndsdefinerte svarkategorier som ikke er basert på ungdommens egne utsagn, som er teoridrevne og/eller basert på populasjoner av voksne med borderline personlighetsforstyrrelse, ikke er tilstrekkelig relevant for selvskadende ungdom. Det foreligger også svært få studier som undersøker ungdommers egne perspektiver på selvskadingen (Stänicke, 2019, s. 51), noe som gjorde at vi ikke hadde kvalitativ litteratur å støtte oss på ved utarbeidelsen av svarresponsene til kriteriene. Videre er CANDI et intervju som innebærer dimensjonale data for de ulike kriteriene, noe som også kan hevdes å utgjøre en ekstra utfordring når det kommer til konseptualisering av de ulike responsene.

Til slutt er både CANDI og skåringsmanualen basert på de foreslåtte DSM-5-kriteriene for NSSID. Det er foreligger fortsatt en debatt rundt klassifiseringen til NSSI, og om fenomenet best forstås innenfor et transdiagnostisk rammeverk eller som en enhetlig diagnose. Det er dermed mulig at det finnes en bedre operasjonalisering av begrepet, som i sin tur vil få følger for hvordan det måles og forstås. Selv om det er utenfor rekkevidden til denne oppgaven å diskutere hvordan fenomenet best klassifiseres, er den nødvendig å nevne, da den både har

preget vårt arbeid med skåringsmanualen, samt hvordan resultatene tolkes og sees opp imot lignende studier. Vi har tatt utgangspunkt i hvordan NSSI operasjonaliseres som en enhetlig diagnose, og våre resultater på kriterienivå for denne diagnosen vil i det følgende sees opp imot lignende studier og diskuteres med tanke på klinisk nytte.

### *5.1.2 Kriterier og klinisk nytte*

Til tross for noe diskrepans blant raterne på itemnivå, var resultatene reliable mellom skårene, og raterne var som nevnt helt enige på kriterienivå. Videre var E-kriteriet det kriteriet som viste seg å skille best mellom deltakerne som fikk NSSID fra de som ikke oppfulgte kriteriet for diagnosen (n=5). Disse funnene er i tråd med tidligere studier, både hos ungdom (Zetterqvist et al., 2020; Zetterqvist, Lundh, Dahlström & Svedin, 2013) og voksne (Gratz et al., 2015; Victor, Davis & Klonsky, 2017) som også har målt kriteriet på en direkte måte. Det har på motsatt side blitt stilt spørsmålstegn ved om E-kriteriet bedre konseptualiseres mer indirekte, ved å *assosiere* NSSI til ubehag og svekkelse, heller enn som en direkte kausal kobling, da flere oppgir at NSSI reduserer ubehag og svekkelse heller enn å forårsake det (Wilkinson & Goodyer, 2011). Studier som har målt ubehag og svekkelse på denne mer indirekte måten, har oppnådd andre resultater, hvor flere i NSSI-utvalget fyller E-kriteriet (Brausch, 2019; Washburn et al., 2015).

Gitt NSSIs høye forekomst med andre lidelser (Nock et al., 2006) og assosiasjon til emosjonell dysregulering (Klonsky, 2007; Jacobson & Gould, 2007), kan det imidlertid argumenteres for at en slik måte å operasjonalisere E-kriteriet på heller fanger opp en global form for ubehag og svekkelse eller andre plager, som ikke nødvendigvis alene er knyttet til NSSI. I tråd med dette, fremviser individer med selvskading flere markører på overordnet ubehag og svekkelse enn individer som ikke driver med selvskading (Klonsky et al., 2003). Det er også funnet lignende mønstre ved sammenligning av individer som møter et direkte målt E-kriterium sammenlignet med individer med subterskel NSSI, hvor oppfylt E-kriterium blant annet er assosiert med antall metoder brukt, intrapersonlige funksjoner, emosjonell dysregulering og depresjon- og angstsymptomer (Gratz et al., 2015). Dette, sammen med funnet om at den direkte vurderingen av E-kriteriet skiller best mellom individene med subterskel NSSI og NSSID, kan dermed antyde at en slik direkte konseptualisering av kriteriet er nyttig i differensieringen av klinisk vs. subklinisk NSSI (Brausch, 2019). Særlig dersom utgangspunktet er klassifiseringen av NSSI som en enhetlig diagnose og ikke som et

transdiagnostisk symptom. Samtidig kan det stilles spørsmålstegn til hvor hensiktsmessig et krav om at funksjonssvekkelse må skyldes NSSI alene, er i en behandlingssammenheng. Faren med å kun knytte E-kriteriet direkte til NSSI, er at individer som ikke oppgir en slik direkte kobling, vil kunne bli undervurdert i klinikken. NSSI er et alvorlig fenomen med potensielt fatale følger, og må tas alvorlig uavhengig av om en potensiell diagnose oppfylles eller ikke. En mulig løsning på E-kriteriet kan være å dele NSSID opp i “med og uten direkte funksjonsnedsettelse” for å belyse forskjeller innad i selvskadingsgruppen, og videre informere behandling.

Videre, når det gjelder kriterium B og C1, ble disse møtt av alle deltakerne. Den høye tilstedeværelsen av disse kriteriene er også i overensstemmelse med andre studier som har undersøkt DSM-5-kriteriene for NSSI (Buelens et al., 2020; Zetterqvist et al., 2013; Grazt et al., 2015; Washburn et al., 2015), og stiller dermed spørsmålstegn til den kliniske nytten til disse kriteriene. Når det kommer til B-kriteriet, forklarte Brausch, Muehlenkamp og Washburn (2016) ut ifra sin studie at det ser ut til å være sjeldent at et individ som har en klinisk relevant frekvens av NSSI ikke vil møte minst én funksjon for atferden. Det ble videre foreslått at en mulig måte å forbedre de diagnostiske kriteriene på, kan være å spesifisere den funksjonelle naturen til NSSI innen kriterium A-symptomene heller enn å ha den funksjonelle naturen til NSSI som et eget symptomkriterium.

Et annet forslag til endring av B-kriteriet, vil være å omformulere det fra at minst én funksjon må være tilstede, til at et minimum antall unike funksjoner skal være tilstede (Brausch et al., 2016). En slik endring vil også være enkel å innføre i CANDI-instrumentet, da intervjuet allerede inkluderer en lang liste med mulige funksjoner. Våre funn viste videre en signifikant korrelasjon mellom antall funksjoner møtt (B-kriteriet) og kriteriet for ubehag og svekkelse (E-kriteriet). I tillegg til dette var antall funksjoner møtt målt opp mot oppnådd diagnose rett over terskelen for hva som kan betegnes som klinisk signifikant ( $r = .454, p \leq .051$ ). Dette kan indikere at det foreligger en sammenheng, som eventuelt kan avdekkes ved et større utvalg. Disse funnene er videre i tråd med tidligere forskning som har dokumentert at individer med moderat og alvorlig NSSI rapporterer flere funksjoner enn de med lavere alvorlighetsgrad (Lloyd-Richardson et al., 2007). Det kan dermed argumenteres for at å revidere B-kriteriet ved å stille krav til antall funksjoner som må møtes, vil kunne øke klinisk nytte av dette kriteriet. Et annet forslag er å tydeligere skille mellom de intrapersonlige og interpersonlige funksjonene, da disse kan representere ulike subgrupper innenfor individer med NSSI

(Klonsky, Glenn, Styer, Olino & Washburn, 2015). Denne studien har ikke foretatt analyser på dette, men tidligere studier har funnet at ulike NSSI-funksjoner kan ha ulike implikasjoner for behandling, prognose og suicidrisiko (Klonsky & Olino, 2008; Nock & Prinstein, 2005; Klonsky et al., 2015). Videre forskning for å undersøke en potensiell reorganisering av B-kriteriet er nødvendig (Brausch et al., 2016).

Videre fant vi lignende korrelasjoner for C1-kriteriet, som måler hvilke følelser som er tilstede i forkant av selvskadingen, med en signifikant korrelasjon med både oppnådd diagnose og med kriteriet for ubehag og svekkelse (E-kriteriet). Disse resultatene indikerer at rapportering av antall følelser også har betydning for diagnose og alvorlighetsgrad av NSSI (vurdert ut ifra ubehag og svekkelse). Dette kan videre ses i lys av affektreguleringshypotesen (Klonsky, 2007; Jacobson & Gould, 2007), og kan antyde at selvskadingen etterhvert kan bli en mer generalisert form for affektregulering hos personer med alvorlig NSSI.

Ved siden av C1, utgjøres C-kriteriet av C2 og C3, som måler henholdsvis overopptatthet med NSSI som er vanskelig å kontrollere, og frekvente tanker på NSSI, selv når de ikke blir handlet på. I vårt utvalg var det 21,1 % (n=4) som ikke fylte C2-kriteriet og 26,3 % (n=5) som ikke fylte kriterium C3. Disse funnene er også i overensstemmelse med tidligere funn i et klinisk utvalg av ungdom med NSSI (Washburn et al., 2015). Totalt sett fylte alle C-kriteriet, ettersom det kun er nødvendig å fylle et av kriterium C1, C2 eller C3. Ettersom de aller fleste fyller C1-kriteriet, betyr dette i praksis at kriterium C2 og C3 på nåværende tidspunkt blir tillagt lite eller ingen diagnostisk vekt. Det kan argumenteres for at en organisering hvor de ulike områdene under C-kriteriet blir vektlagt hver for seg, bedre vil kunne belyse selvskadingens heterogenitet, og potensielt også skille mellom ulike subgrupper. I tråd med dette er overopptatthet av selvskading blitt beskrevet av klinikere som et prototypisk symptom på NSSID, noe som kan implisere at dette kriteriet bør tillegges mer relevans i diagnostisering (Lengel & Mullins-Sweatt, 2013). Videre fant vi en signifikant korrelasjon mellom antall selvskadingsmetoder og grad av overopptatthet. Dette er i overensstemmelse med tidligere studier og støtter antakelsen om at individer med høyere alvorlighetsgrad av NSSI (vurdert ut ifra antall selvskadingsmetoder og grad av vevsskade), i større grad overveier og tenker på selvskadingen i forkant, sammenlignet med mildere former for NSSI (Fikke et al., 2011; Lloyd-Richardson et al., 2007). Videre studier på hvordan de ulike underkriteriene til kriterium C korrelerer med eksekutive funksjoner og impulsivitet vil kunne belyse dette nærmere, og potensielt kunne støtte opp under en oppdeling av C-kriteriet.

Kriterium D ble møtt av alle deltakerne i vår studie, noe som er i overensstemmelse med tidligere studier som har rapportert at hovedandelen av individer som møter de andre kriteriene for NSSID, rapporterer vanlige former for NSSI (som kutting og slåing) og at veldig få begrenser atferden til neglebiting og hudplukking (Gratz et al., 2015; Victor et al., 2017).

Noe relatert til kriterium D, er kriterium F, som i vårt utvalg ekskluderte 5,3 % (n=1) fra å møte NSSID. Zetterqvist (2020) fant at alle deltakerne i et ungdomsutvalg med NSSI møtte F-kriteriet, og andre studier på lignende utvalg har funnet at 0,8 % ikke møtte NSSID som følge av å ikke fylle F-kriteriet (Washburn et al., 2015). Oppsummert ser likevel både D- og F-kriteriene ut til å være nyttige for diagnostisk validitet ved å hjelpe klinikere å identifisere tydelig NSSI-atferd og for å utelukke andre potensielle lidelser (Brausch, 2019, s. 78).

Kriterium A beskrives som frekvens- og nylighetskriteriet, men kriteriet måler også antall NSSI-metoder, om det har forekommet vevsskade og eventuelle selvmordsforsøk. 26,3% (n=5) av deltakerne fra det endelige utvalget rapporterte at de hadde sluttet med selvskading og dermed ikke hadde NSSI i løpet av det siste året, men at de hadde drevet med selvskading i løpet av de siste to årene. Ettersom hovedformålet med denne studien var å undersøke interrater-reliabilitet, valgte vi å utvide nylighetsgrensen til to år istedenfor et år for disse deltakerne, da dette ville øke antall deltakere vi kunne utføre reliabilitetstesting på. Med dette som utgangspunkt, viste våre funn at 10,5% (n=2) ikke fikk oppfylt NSSID på bakgrunn av å ikke fylle A-kriteriet. Våre resultater kan imidlertid ikke sammenlignes med andre studier, da disse tar utgangspunkt i siste år. Videre rapporterte 94,7% (n=18) av vårt utvalg selvskading mer enn 10 ganger i løpet det året de skadet seg. Dette er også i tråd med tidligere studier som har foreslått at å øke frekvensgrensen til 10 episoder, vil bedre kunne skille individer med subterskel NSSID og de som fyller diagnosen (Muehlenkamp & Brausch, 2016). Et krav på kun fem episoder det siste året har av mange blitt hevdet å være for liberalt, og kan bidra til en overpatologisering av individer som driver med NSSI mer periodevis (Brausch, 2019, ss. 74-75). Det kan dermed stilles spørsmålstegn til den diagnostiske validiteten til det nåværende A-kriteriet, og om en økning av frekvensterskelen vil kunne bidra til en forbedring av denne.

Selv om det for oppfylging av A-kriteriet også kreves at det skal foreligge vevsskade, vurderes ikke alvorlighetsgrad av vevsskaden, og hvor mange ulike metoder som benyttes, samt antall selvmordsforsøk, blir heller ikke vektlagt. I vårt utvalg var det variasjon blant deltakerne i hvor mange metoder de hadde brukt, hvorvidt det forelå tidligere selvmordsforsøk, samt om vevsskaden innebar overflatiske blødninger, eller dype, pleietrengende sår. Videre viste resultatene en signifikant korrelasjon mellom antall episoder og antall metoder, noe som kan implisere at antall metoder har betydning for alvorlighetsgrad (vurdert ut fra antall episoder, se Muehlenkamp et al., 2017). Det kan dermed argumenteres for at et skille på disse områdene vil kunne bidra til å skille mellom NSSI-subgrupper, som i sin tur blant annet vil kunne informere behandling.

Oppsummert, samsvarer flere av resultatene våre på kriterienivå med tidligere studier. Nærmere bestemt gjelder dette særlig 1) E-kriteriet er det kriteriet som best skiller de individene med NSSID fra de med subterksel NSSI, 2) alle fyller B og C og D-kriteriet, 3) antall funksjoner møtt og følelser i forkant korrelerer med E-kriteriet og kan på den måten kunne knyttes til alvorlighetsgrad av NSSI, 4) antall selvskadingsmetoder korrelerer med grad av overopptatthet, og kan kunne bidra til å skille mellom ulike subgrupper av NSSI, 5) A-kriteriet demonstrerte at de fleste deltakerne, og alle som fyller NSSID, har 10 eller flere selvskadingsepisoder i løpet av det året de har skadet seg selv.

På denne måten kan resultatene, i samsvar med tidligere kritikk, bidra til å belyse svakheter ved den nåværende organiseringen og vektleggingen av kriteriene til NSSID. Fremtidige studier med flere mål på NSSID og andre markører, vil være nyttig for å undersøke om en annen organisering og vektlegging av kriteriene bedre vil kunne belyse selvskadingens heterogenitet, samt være mer hensiktsmessig i en eventuell diagnostisering av NSSID. Selv om resultatene på mange måter samsvarer med andre studier, har vi i denne studien ikke andre mål på selvskading enn CANDI, og kan derfor ikke trekke konklusjoner angående konstruktvaliditeten til intervjuet. Dette, sammen med metodiske svakheter som vil bli diskutert i det følgende, begrenser generaliserbarheten til resultatene.

## **5.2 Begrensninger**

### **5.2.1 Utvalget**

Når det kommer til utvalget i denne studien, består det av en relativ homogen gruppe, med tanke på demografiske variabler som alder, kjønn og bosted. 68,4% (n=13) av deltakerne var enten 16 eller 17 år på intervju tidspunktet. Muligheten for å generalisere funnene kunne ha vært økt dersom det hadde vært flere deltakere i flere forskjellige aldersgrupper, særlig i den yngre delen av ungdomsgruppen. Videre var en overvekt av deltakerne i prosjektet jenter, noe som begrenser generaliserbarheten til begge kjønn. I tillegg er deltakerne rekruttert fra samme by, og det er slik sett ikke et variert utvalg. Samtidig er de hentet fra to ulike områder i byen, som representerer to ulike sosioøkonomiske områder. Til slutt utgjør utvalget vårt en avgrenset gruppe sett opp mot andre kliniske grupper og i populasjonssammenheng. En utvidet replikeringsstudie vil dermed være nødvendig i fremtiden for å kunne generalisere resultatene ytterligere.

Oppsummert, utgjør et lite og homogent utvalg en svakhet ved denne studien. Koo og Li (2016) hevdet at man bør ha minst 30 heterogene deltakere og minst 3 ratere når det er mulig, ved gjennomføring av reliabilitetsstudier. Walter, Eliasziw og Donner (1998) hevder imidlertid at det totale nummeret på antall observasjoner kan minimeres med et relativt lite antall av observatører, så lenge reliabiliteten er høy.

### 5.2.2 Raterne

Ved siden av begrensninger knyttet til utvalg, foreligger det også svakheter knyttet til raterne av denne studien. Raterne utarbeidet skåringsmanualen sammen, samt trente og diskuterte ulike intervjuer i forkant av reliabilitetstesting. Dette kan ha vært opphav til observatørbias, som kan opptre når observatøren(e) er klar over studiens mål, og deres observasjoner blir påvirket av denne kunnskapen (Bordens & Abbott, 2011). Dette kan videre ha bidratt til den høye interrater-reliabiliteten, ved at raterne gjennom prosessen har fått innsikt i hva den andre personen ofte vektlegger i en vurdering, og dermed ubevisst har blitt påvirket av dette under egne vurderinger. Fremtidige studier med uavhengige ratere vil dermed være nyttig for å redusere effekten av observatørbias.

Videre begrenses resultatene av den lave treningsmengden raterne i studien hadde. En større treningsmengde vil blant annet kunne ha redusert eventuelle læringseffekter, som er former for overføringseffekter (Bordens & Abbott, 2011). Gjentatte intervjuer kan føre til en læringseffekt som vil gjøre at de siste intervjuene blir gjennomført bedre og mer presist enn

de første. Erfaring med skåring vil også kunne gjøre at kvaliteten på vurderingene endrer seg i takt med erfaring fra å ha observert og skåret ut mange intervjuer. Overføringseffekten *kontrast* kan også ha påvirket resultatene. Skåring av et intervju av en person med alvorlig selvskading, vil kunne føre til en undervurdering i skåringen av det påfølgende intervjuet, dersom dette intervjuet er av en person med mindre alvorlig selvskading. Effekten av ovennevnte overføringseffekter ble forsøkt minimert med randomisering, men vil også kunne ha blitt redusert ytterligere med økt treningsmengde og erfaring.

I noe tråd med det ovennevnte ble et av intervjuene skåret ut på nytt av én av raterne. Dette skyldtes at intervjuobjektet hadde hatt store endringer i sin atferd og sine plager i løpet av det siste året. Én av raterne tok utgangspunkt i hvordan lidelsestrykket var tidligere på året, når det var som verst, mens den andre rateren tok utgangspunkt i hvordan lidelsestrykket var akkurat nå. Vi oppdaget diskrepansen blant raterne tidlig under gjennomgang av intervjuene etter skåring, og vurderte i samråd med prosjektkoordinator av hovedstudien at det var nødvendig å ta utgangspunkt i samme tidsperiode slik at vi ikke vurderte to ulike symptombilder. For å minske grad av hukommelse og observatørbias, ble ikke intervjuet skåret ut umiddelbart etter diskrepansen ble oppdaget. Andre intervjuer ble også gjennomgått i mellomtiden. Selv om ny skåring av intervjuet bidro til økt interrater-reliabilitet for studien som helhet, var dette intervjuet ett av de vi oppnådde lavere enighet på, noe som kan tyde på at vi lyktes i å redusere effekten av hukommelse og observatørbias i dette bestemte tilfellet.

I sum belyser likevel hendelsen svakheten ved begrenset trening og følgelig noe uklare retningslinjer fra skåringsmanual. Økt treningsmengde ville kunne ha bidratt til å løse dilemmaer som det ovennevnte, i forkant av testingen. Dette ville i sin tur kunne ha bidratt til utvikling av tydeligere retningslinjer i skåringsmanual, samt bidratt til mer klargjørende og presise spørsmål ved tvetydige utsagn under selve intervjuingen. Hendelsen førte til at skåringsmanualen ble gjort tydeligere på vurdering av denne type tilfeller i etterkant. Dette er også den eneste endringen i skåringsmanualen som er gjort i etterkant av skåringen av intervjuene. Til tross for at økt trening på mange måter kunne ha styrket studien, indikerer også den høye interrater-reliabiliteten at CANDI kan gjennomføres med begrenset trening.

### 5.2.3 Studiedesignet



Som tidligere nevnt er flere av spørsmålene i CANDI utformet, og følgelig også administrert som et muntlig spørreskjema, blant disse F-kriteriet. At F-kriteriet, som innebærer en rekke potensielle psykiske lidelser og medisinske tilstander som kan forklare den selvskadende atferden, blir vurdert kjapt med enkle ja/nei-spørsmål, er en metodisk svakhet med denne studien og CANDI som intervju. Ideelt sett burde studien hatt andre diagnostiske mål for å vurdere tilstedeværelsen av andre diagnoser som kan ekskludere NSSID. Dette ville ha gjort studien mer omfattende, men samtidig økt validiteten og følgelig også generaliserbarheten til resultatene. Det er imidlertid naturlig å se for seg at CANDI i klinisk sammenheng vil bli brukt som en del av en større utredning, noe som gjør at denne utformingen av F-kriteriet sannsynligvis ikke vil medføre store problemer i klinikken.

I noe tråd med det ovennevnte, er det på siste side av CANDI to supplerende målinger der det skal vurderes global validitet og global alvorlighetsgrad. Disse skalaene ble ikke inkludert i vår interrater-reliabilitetstesting. Årsaken er basert på at hovedformålet med denne oppgaven var å måle interrater-reliabiliteten ved å vurdere svarresponser som omfatter NSSI-begrepet. Hvordan validitet av svarresponser skal vurderes omfatter et annet fagfelt, og vi valgte derfor - i tråd med oppgavens omfang og tidsbegrensning - å utelukke denne skalaen fra skåringsmanualen og reliabilitetstesting. Når det gjelder vurdering av global alvorlighet av NSSI, er denne skalaen noe basert på skalaen for global validitet, og vi valgte derfor å også utelukke vurdering av global alvorlighet. Et mål på disse skalaene ville vært optimalt for å kunne vurdere eventuelle subgrupper, samt validiteten av svarresponsene, som i sin tur ville kunne gjort resultatene mer egnet for generalisering.

En annen begrensning ved studien, er at den er retrospektiv i sitt design, noe som er problematisk ettersom hukommelsen for bestemte hendelser av NSSI, sannsynligvis dempes og endrer seg over tid (Jacobson & Gould, 2007). En slik basering på hukommelsen, kan videre utgjøre en trussel for validiteten. Dette gjelder særlig frekvenskriteriet hvor deltakeren blir bedt om å anslå antall selvskadingsepisoder siste år (Jacobson & Gould, 2007). Et annet aspekt som kan være en trussel mot validiteten, er deltakernes unge alder og, for noen av dem, den nylige oppstarten av NSSI. Noen av deltakerne hadde ikke drevet med selvskading lenge, og følgelig kan det tenkes at de heller ikke har reflektert mye over årsaker til hvorfor de gjør det, hva de føler i forkant og hvor mye det opptar dem. Noe i samsvar med dette er også effekten av sosial ønskedighet, som refererer til deltakernes tendens til å rapportere sosiale ønskelige trekk og atferder, og å fornekte slike trekk og atferder som er ikke-sosialt ønskelige

(Krumpal, 2013). Gitt intervjuets sensitive natur, er det ikke utenkelig at denne formen for bias kan ha farget deltakernes responser. I tilfelle NSSI kan dette være særlig relevant for rapportering av interpersonlige funksjoner (Nock, 2008), da selvskading som en måte å påvirke andre på kan oppfattes som manipulativt. I tillegg kan det å selv rapportere at selvskadingen utføres for å påvirke andre, virke mot sin hensikt ved å undergrave signaleffekten til atferden (Nock, 2008). På tross av dette, ble det under intervjuene i denne studien rapportert interpersonlige grunner av flere deltakere, dog ikke like hyppig og tilstedeværende som de intrapersonlige funksjonene.

Til slutt, for å løse validitetsutfordringene nevnt over, kan det være hensiktsmessig å redusere tidsperspektivet, samt foreta en test-retest av intervjuet for å undersøke både intervjuets og diagnosens stabilitet. Videre vil en multimetodisk tilnærming i fremtiden være nødvendig for å ytterligere kunne validere NSSID som diagnose. Det forutsettes at metodene benytter samme operasjonalisering av selvskadingsbegrepet og konsekvent undersøker alle de foreslåtte DSM-5-kriteriene for å oppnå dette.

## **6. Konklusjon**

Gjennom denne pilotstudien på en klinisk ungdomspopulasjon med NSSI, fant vi en svært god interrater-reliabilitet, samt en adekvat intern konsistens for den norske versjonen av det klinisk strukturerte intervjuet CANDI. Videre bidro resultatene til å belyse tidligere diskuterte svakheter ved kriteriene, med tanke på klinisk nytte. Den nåværende organiseringen og vektleggingen av kriteriene ser ut til å være lite egnet til å fange opp heterogeniteten og alvorlighetsgraden innad i selvskadingsgruppen. I fremtiden kan det dermed vurderes om det bør foretas endringer i takt med økt kunnskap på feltet og i tråd med en voksende forskningslitteratur. Fremtidige replikasjonsstudier med uavhengige ratere og et større utvalg, samt med flere mål på NSSI, vil også være nødvendig for å øke generaliserbarheten av resultatene. Til tross for metodiske svakheter, fremstår CANDI likevel som et viktig tilskudd i selvskadingslitteraturen, da det direkte tar for seg alle de foreslåtte DSM-5 kriteriene på NSSID, er anvendelig og har vist god intern konsistens og interrater-reliabilitet både på unge voksne og ungdommer med NSSI.

## Referanseliste

- Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine, 18*(3), 91-93. DOI: 10.1016/j.tjem.2018.08.001
- American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). DOI: 10.1176/appi.books.9780890425596
- Andover, M. S., Pepper, C. M., Ryabchenko, K. A., Orrico, E. G. & Gibb, B. E. (2005). Self-mutilation and symptoms of depression, anxiety, and borderline personality disorder. *Suicide and Life-Threatening Behavior, 35*(5), 581–591. DOI: 10.1521/suli.2005.35.5.581
- Bentley, K. H., Nock, M. K. & Barlow, D. H. (2014). The four-function model of nonsuicidal self-injury: Key directions for future research. *Clinical Psychological Science, 2*(5), 638-656. DOI: 10.1177/2167702613514563
- Blake, D., Weathers, F. W., Nagy, L. M., Kaloupek D. G., Gusman F. D., Charney, D. S. & Keane, T. M. (1995). The development of a Clinician-Administered PTSD Scale. *Journal of Traumatic Stress, 8*(1), 75–90. DOI: 10.1007/bf02105408
- Blake, D., Weathers, F. W., Nagy, L. M., Kaloupek D., Klauminzer, G., Charney, D., Kean, T. & Buckley, T. C. (2000). Instruction manual: Clinician-administered PTSD scale (CAPS). *National Center for Posttraumatic Stress Disorder*. Hentet fra: [https://www.ptsd.va.gov/professional/assessment/documents/CAPSmanual.pdf?fbclid=IwAR3d1r8QFXhr07\\_zWtAPftZyiPW\\_AXs1mtj5zmQ6BgUOIE\\_MAzCS6Zgm-Kw](https://www.ptsd.va.gov/professional/assessment/documents/CAPSmanual.pdf?fbclid=IwAR3d1r8QFXhr07_zWtAPftZyiPW_AXs1mtj5zmQ6BgUOIE_MAzCS6Zgm-Kw)
- Bordens, K. S. & Abbott, B. B. (2011). *A Process Approach: Research Design and Methods* (8. utg.). New York: McGraw-Hill.
- Brausch, A. M. (2019). Diagnostic Classification of Nonsuicidal Self-Injury. I Washburn, J.J. (Red.), *Nonsuicidal Self-Injury: Advances in Research and Practice* (1.utg., ss. 71-87). New York: Routledge

- Brausch, A. M., Muehlenkamp J. J. & Washburn, J. J. (2016). Nonsuicidal Self-injury Disorder: Does Criterion B Add Diagnostic Utility? *Psychiatry Research*, 244, 179-184.  
DOI: 10.1016/j.psychres.2016.07.025
- Bresin, K. & Schoenleber, M. (2015). Gender differences in the prevalence of nonsuicidal self-injury: A meta-analysis. *Clinical Psychology Review*, 38, 55-64.  
<https://doi.org/10.1016/j.cpr.2015.02.009>
- Brown, M. Z., Comtois, K. A. & Linehan, M. M. (2002). Reasons for suicide attempts and nonsuicidal self-injury in women with borderline personality disorder. *Journal of Abnormal Psychology*, 111(1), 198–202. <https://doi.org/10.1037/0021-843X.111.1.198>
- Brown, R. C. & Plener, P. L. (2017). Non-Suicidal Self-Injury in Adolescence. *Current Psychiatry Reports*, 19(3), 20. DOI: 10.1007/s11920-017-0767-9
- Buelens, T., Luyckx, K., Kiekens, G., Gandhi, A., Muehlenkamp, J. J. & Claes, L. (2020). Investigating the DSM-5 criteria for non-suicidal self-injury disorder in a community sample of adolescents. *Journal of Affective Disorders*, 260, 314-322.  
<https://doi.org/10.1016/j.jad.2019.09.009>
- Burke, T. A., Ammerman, B. A., Hamilton, J. L. & Alloy, L. B. (2017). Impact of Non-Suicidal Self-Injury Scale: Initial Psychometric Validation. *Cognitive Therapy and Research*, 41(1), 130-142. DOI: 10.1007/s10608-016-9806-9
- Ceci, S. J. & Bruck, M. (1999). The suggestibility of children's memory. *Annual Review of Psychology*, 50, 419-439. DOI: 10.1146/annurev.psych.50.1.419
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37-46. DOI: 10.1177/001316446002000104
- Cohen, R. J., Swerdlik, M. E. & Sturman, E. D. (2013). *Psychological Testing and Assessment. An Introduction to Tests and Measurements* (8. utg.). New York: McGraw-Hill Education.

- Cooper, J., Kapur, N., Webb, R., Lawlor, M., Guthrie, E., Mackway-Jones, K. & Appleby, L. (2005). Suicide after deliberate self-harm: a 4-year cohort study. *The American Journal of Psychiatry*, 162(2), 297–303. DOI: 10.1176/appi.ajp.162.2.297
- Hauber, K., Boon, A. & Vermeiren, R. (2019). Non-suicidal Self-Injury in Clinical Practice. *Frontiers in Psychology*, 10(502). doi: [10.3389/fpsyg.2019.00502](https://doi.org/10.3389/fpsyg.2019.00502)
- DiClemente, R. J., Ponton, L. E. & Hartley, D. (1991). Prevalence and correlates of cutting behavior: Risk for HIV transmission. *Journal of the American Academy of Child and Adolescent Psychiatry*, 30(5), 735 - 739. [https://doi.org/10.1016/S0890-8567\(10\)80007-3](https://doi.org/10.1016/S0890-8567(10)80007-3)
- Ezpleta, L., de la Osa, N., Doménech, J. M., Navarro, J. B., Losilla, J. M. & Júdez, J. (1997). Diagnostic Agreement between Clinicians and the Diagnostic Interview for Children and Adolescents - DICA-R - in an Outpatient Sample. *Child Psychology & Psychiatry & Allied Disciplines*, 38(4), 431-440. DOI: 10.1111/j.1469-7610.1997.tb01528.x
- Favazza, A. R. & Conterio, K. (1989). Female habitual self-mutilators. *Acta Psychiatrica Scandinavica*, 79(3), 283–289. DOI: 10.1111/j.1600-0447.1989.tb10259.x
- Field, A. (2009). *Discovering statistics using SPSS* (3. utg.). London: SAGE Publications.
- Fikke, L. T., Melinder, A. & Landrø, N. I. (2011). Executive functions are impaired in adolescents engaging in non-suicidal self-injury. *Psychological Medicine*, 41(3), 601-610. DOI: 10.1017/S0033291710001030
- Frick, P. J., Barry, C. T. & Kamphaus, R. W. (2010). *Clinical Assessment of Child and Adolescent Personality and Behavior*. (3. utg.). New York: Springer
- Fylan, F. (2005). Semi-structured interviewing. I Miles, J. & Gilberts, P. (Red.), *A Handbook of Research Methods for Clinical and Health Psychology*. New York: Oxford University Press.

- Gould, M. S., Marrocco, F. A., Kleinman, M., Thomas, J. G., Mostkoff, K., Cote, J. & Davies, M. (2005). Evaluating iatrogenic risk of youth suicide screening programs: A randomized controlled trial. *Journal of the American Medical Association*, 293(13), 1635-1643. DOI: 10.1001/jama.293.13.1635
- Gratz, K. L. (2003). Risk Factors for and Functions of Deliberate Self-Harm: An Empirical and Conceptual Review. *Clinical Psychology - Science and Practice*, 10(2), 192-205. DOI: 10.1093/clipsy.bpg022
- Gratz, K. L., Dixon-Gordon, K. L., Chapman, A. L. & Tull, M. T. (2015). Diagnosis and Characterization of DSM-5 Nonsuicidal Self-Injury Disorder Using the Clinician-Administered Nonsuicidal Self-Injury Disorder Index. *Assessment*, 22(5), 527-539. DOI: 10.1177/1073191114565878
- Gudjonsson, G. H. (2003). The psychology of interrogations and confessions. A handbook. West Sussex: John Wiley & Sons.
- Hallgren, K. A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23-34. DOI: 10.20982/tqmp.08.1.p023
- Hargus, E., Hawton, K. & Rodham, K. (2009). Distinguishing between subgroups of adolescents who self-Harm. *Suicide and Life-Threatening Behaviour*, 39(5), 518-537. DOI: 10.1521/suli.2009.39.5.518
- Hawton, K., Rodham, K., Evans, E. & Weatherall, R. (2002). Deliberate self harm in adolescents: Self report survey in schools in England. *The British Medical Journal*, 325(7374), 1207-1211. DOI: 10.1136/bmj.325.7374.1207
- Hawton, K., Saunders K. E. A. & O'Connor R. C. (2012). Self-harm and suicide in adolescents. *The Lancet*, 379(9834), 2373-2382. DOI: 10.1016/S0140-6736(12)60322-5

Helseforskningsloven. 2008. Lov om medisinsk og helsefaglig forskning (LOV-2008-06-20-44). Hentet fra: [https://lovdata.no/dokument/NL/lov/2008-06-20-44#KAPITTEL\\_4](https://lovdata.no/dokument/NL/lov/2008-06-20-44#KAPITTEL_4)

Helse- og omsorgstjenesteloven. 2011. Lov om kommunale helse- og omsorgstjenester. (LOV-2011-06-24-30). Hentet fra: [https://lovdata.no/dokument/NL/lov/2011-06-24-30#KAPITTEL\\_4](https://lovdata.no/dokument/NL/lov/2011-06-24-30#KAPITTEL_4)

International Society for the Study of Self-injury. (2018, Mai). What is self-injury? Hentet fra: <https://itriples.org/about-self-injury/what-is-self-injury>

Jacobson, C. M. & Gould, M. (2007). The epidemiology and phenomenology of non-suicidal self-injurious behavior among adolescents: a critical review of the literature. *Archives of Suicide Research, 11*(2), 129-147. DOI: 10.1080/13811110701247602

Jacobson C. M., Muehlenkamp J. J., Miller A. L. & Turner J. B. (2008). Psychiatric impairment among adolescents engaging in different types of deliberate self-harm. *Journal of Clinical Child & Adolescent Psychology, 37*(2), 363–375. DOI: 10.1080/15374410801955771

Karterud, S., Pedersen, G., Løvdahl, H. & Friis, S. (1998). S-GAF: Global funksjonsskåring - Splittet versjon. *Klinikk for psykiatri, Ullevål sykehus*. Hentet fra: <https://oslo-universitetssykehus.no/seksjon/nasjonalt-kompetansetjeneste-for-personlighetspsykiatri-napp/Documents/S-GAF%20veiledning.pdf>

Klonsky, D. E. (2007). The functions of deliberate self-injury: A review of the evidence. *Clinical Psychology Review, 27*(2), 226-239. DOI: 10.1016/j.cpr.2006.08.002

Klonsky, D. E., Glenn, C. R., Styer, D. M., Olino, T. M. & Washburn, J. J. (2015). The functions of nonsuicidal self-injury: converging evidence for a two-factor structure. *Child and Adolescent Psychiatry and Mental Health, 9*(44). DOI 10.1186/s13034-015-0073-4

- Klonsky, D. E & Muehlenkamp, J. J. (2007). Self-injury: A research review for the practitioner. *Journal of Clinical Psychology*, 63(11), 1045-1056. DOI: 10.1002/jclp.20412
- Klonsky, D. E. & Olino, T. M. (2008). Identifying clinically distinct subgroups of self-injurers among young adults: a latent class analysis. *Journal of Consulting and Clinical Psychology*, 76(1), 22-27. doi: 10.1037/0022-006X.76.1.22.
- Klonsky, D. E., Oltmanns, T. F. & Turkheimer, E. (2003). Deliberate Self-Harm in a nonclinical population: Prevalence and psychological correlates. *American Journal of Psychiatry*. 160 (8), 1501-1508. DOI: 10.1176/appi.ajp.160.8.1501
- Koo, T. K. & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155-163. doi: 10.1016/j.jcm.2016.02.012
- Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *The European Journal of Developmental Research*, 47, 2025-2047. DOI: 10.1007/s11135-011-9640-9
- Lengel, G. J. & Mullins-Sweatt, S. N. (2013). Nonsuicidal self-injury disorder: Clinician and expert ratings. *Psychiatry Research*, 210(3), 940-944. DOI: 10.1016/j.psychres.2013.08.047
- Leo, D. D. & Heller T. S. (2004). Who are the kids who self-harm? An Australian self-report school survey. *The Medical Journal of Australia*, 181(3), 140-144. DOI: 10.5694/j.1326-5377.2004.tb06204.x
- Lloyd-Richardson, E. E., Perrine, N., Dierker, L. & Kelley, M. L. (2007). Characteristics and functions of non-suicidal self-injury in a community sample of adolescents. *Psychological Medicine*, 37(8), 1183-1192. DOI: 10.1017/S003329170700027X



- Madge, N., Hewitt, A., Hawton, K., de Wilde, E. J., Corcoran, P., Fekete, S., ...Ystgaard, M. (2008). Deliberate self-harm within an international community sample of young people: comparative findings from the Child & Adolescent Self-harm in Europe (CASE) Study. *The Journal of Child Psychology and Psychiatry*, 49(6), 667-677. DOI: 10.1111/j.1469-7610.2008.01879.x
- Magnussen, S. (2017). *Vitnepsykologi 2.0*. Abstrakt forlag: Oslo.
- Matthews, P. C. (1968). Epidemic Self-Injury in an Adolescent Unit. *International Journal of Social Psychiatry*, 14(2), 125 –133. DOI: 10.1177/002076406801400206
- McGraw, K. O. & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30–46. DOI: 10.1037/1082-989X.1.1.30
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, 22 (3), 276 - 282. DOI: 10.11613/BM.2012.031
- Melinder, A. M. D. (2011). Gjentatte avhør: Flere forhold har betydning for påliteligheten i barns utsagn. *Tidsskrift for Norsk Psykologforening*, 48(11), 1091- 1095
- Moran, P., Coffey, C., Romaniuk, H., Olsson, C., Borschman, R., Carlin, J. B. & Patton, G. C. (2012). The natural history of self-harm from adolescence to young adulthood: a population-based cohort study. *The Lancet*, 379(9812), 236-243. DOI: 10.1016/S0140-6736(11)61141-0
- Muehlenkamp, J. J. (2005). Self-Injurious Behavior as a Separate Clinical Syndrome. *American Journal of Orthopsychiatry*, 75(2), 324-33. DOI: 10.1037/0002-9432.75.2.324
- Muehlenkamp, J. J. & Brausch A. M. (2016). Reconsidering Criterion A for the Diagnosis of Non-Suicidal Self-Injury Disorder. *Journal of Psychopathology and Behavioral Research*, 38, 547-558. DOI: 10.1007/s10862-016-9543-0

- Muehlenkamp, J. J., Brausch, A. M. & Washburn, J. J. (2017). How much is enough? Examining Frequency Criteria for NSSI Disorder in Adolescent Inpatients. *Journal of Consulting and Clinical Psychology, 85*(6), 611-619. DOI: 10.1037/ccp0000209
- Muehlenkamp, J. J., Claes, L., Havertape, L. & Plener, P. L. (2012). International prevalence of adolescent non-suicidal self-injury and deliberate self-harm. *Child and Adolescent Psychiatry and Mental Health, 6*(10). DOI: 10.1186/1753-2000-6-10
- Muehlenkamp, J. J. & Gutierrez, P. M. (2004). An investigation of differences between self-injurious behavior and suicide attempts in a sample of adolescents. *Suicide and Life-Threatening Behavior, 34*(1), 12-23. DOI: 10.1521/suli.34.1.12.27769
- Muehlenkamp, J. J., Hoff, E. R., Licht, J., Azure, J. & Hasenzahl, S. J. (2008). Rates of Non-Suicidal Self-Injury: A Cross-Sectional Analysis of Exposure. *Current Psychology, 27*, 234-241. DOI: 10.1007/s12144-008-9036-8
- Muehlenkamp, J. J., Swenson, L. P., Batejan, K. L. & Jarvi, S. M. (2014). Emotional and Behavioral Effects of Participating in an Online Study of Nonsuicidal Self-Injury: An Experimental Analysis. *Clinical Psychological Science, 3*(1), 26-37. DOI: 10.1177/2167702614531579
- Nock, M. K. (2008). Action speak louder than words: An elaborated theoretical model of the social functions of self-injury and other harmful behaviors. *Applied and Preventive Psychology 12*(4), 159–168. DOI:10.1016/j.appsy.2008.05.002.
- Nock, M. K., Joiner, T. E. Jr., Gordon, K. H., Lloyd-Richardson, E. & Prinstein, M. J. (2006). Non-suicidal self-injury among adolescents: Diagnostic correlates and relation to suicide attempts. *Psychiatry Research, 144*(1), 65–72. DOI: 10.1016/j.psychres.2006.05.010

- Nock, M. K. & Prinstein, M. J. (2004). A Functional Approach to the Assessment of Self-Mutilative Behavior. *Journal of Consulting and Clinical Psychology*, 72(5), 885-890. doi:10.1037/0022-006X.72.5.885
- Nock, M. K. & Prinstein, M. J. (2005). Contextual Features and Behavioral Functions of Self-Mutilation Among Adolescents. *Journal of Abnormal Psychology*, 114(1), 140-146. DOI: [10.1037/0021-843X.114.1.140](https://doi.org/10.1037/0021-843X.114.1.140)
- O'Leary-Kelly, S. W. & Vokurka, R. J. (1998). The empirical assessment of construct validity. *Journal of Operations Management*, 16(4), 387-405. DOI:10.1016/S0272-6963(98)00020-5
- Pallant, J. (2010). *SPSS Survival Manual (4th ed.)*. England: McGraw-Hill Education.
- Regier, D. A., Narrow, W. E., Clarke, D. E., Kraemer, H. C., Kuramoto, S. J., Kuhl, E. A. & Kupfer, D. J. (2013). DSM-5 field trials in the United States and Canada, part II: Test-retest reliability of selected categorical diagnoses. *American Journal of Psychiatry*, 170(1), 59-70. DOI: [10.1176/appi.ajp.2012.12070999](https://doi.org/10.1176/appi.ajp.2012.12070999)
- Rogers, R. (1995). *Diagnostic and Structured Interviewing: A Handbook for Psychologists*. Florida: Psychological Assessment Resources, Inc.
- Selby, E. A., Kranzler, A., Fehling, K. B. & Panza, E. (2015). Nonsuicidal self-injury disorder: The path to diagnostic validity and final obstacles. *Clinical Psychology Review*, 38, 79-91. DOI: [10.1016/j.cpr.2015.03.003](https://doi.org/10.1016/j.cpr.2015.03.003).
- Shaffer, D. & Jacobson, C. (2009). Proposal to the DSM-V childhood disorder and mood disorder work groups to include non-suicidal self-injury (NSSI) as a DSM-V disorder. American Psychiatric Association. Hentet fra: <http://www.dsm5.org/Pages/Default.aspx>.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420. DOI: [10.1037//0033-2909.86.2.420](https://doi.org/10.1037//0033-2909.86.2.420)

- Siegal, M., Waters, L. J. & Dinwiddy, L. S. (1988). Misleading children: Causal attributions for inconsistency under repeated questioning. *Journal of Experimental Child Psychology*, 45(3), 438–456. DOI: 10.1016/0022-0965(88)90041-0
- Skegg, K. (2005). Self-harm. *The Lancet*, 366(9495), 1471-1483. DOI: 10.1016/S0140-6736(05)67600-3
- Sosial og helsedirektoratet. (2008). Nasjonale retningslinjer for forebygging av selvmord i psykisk helsevern. IS-1511. Hentet fra:  
<https://www.helsedirektoratet.no/retningslinjer/forebygging-av-selvmord-i-psykisk-helsevern>
- Spitzer, R. L., Williams, J. B., Gibbon, M. & First, M. B. (1992). The Structured Clinical Interview for DSM-III-R (SCID). I: History, Rationale, and Description. *Archives of General Psychology*, 49(8), 624-629. DOI:10.1001/archpsyc.1992.01820080032005
- Stänicke, L. I. (2019). *The punished self, the unknown self, and the harmed self - towards a more nuanced understanding of self-harm in adolescence*. (Doktoravhandling). Universitetet i Oslo, Oslo.
- Suyemoto, K. L. (1998). The functions of self-mutilation. *Clinical Psychology Review*, 18(5), 531-554. DOI: 10.1016/s0272-7358(97)00105-0
- Tavakol, M. & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53-55. DOI: 10.5116/ijme.4dfb.8dfd
- Tinsley, H. E. & Weiss, D. J. (1975). Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology*, 22(4), 358–376. DOI: 10.1037/h0076640
- Turner, B. J., Chapman, A. L. & Layden, B. K. (2012). Intrapersonal and Interpersonal Functions of Non suicidal Self-Injury: Associations with Emotional and Social

- Functioning. *Suicide and Life-Threatening Behavior*, 42(1), 36-55. DOI: 10.1111/j.1943-278X.2011.00069.x.
- Tørmoen, A. J. (2017). *Self-Harm among Adolescents: From Identification to Tailored Treatment* (Doktoravhandling). Universitetet i Oslo, Oslo.
- Universitetet i Oslo. (2018, 6. juni). Villet egenskade - atferd og symptomer på borderline personlighetsforstyrrelse hos ungdom som er i kontakt med psykisk helsevern. Hentet fra: <https://www.med.uio.no/klinmed/forskning/sentre/nssf/forskning/nssf-prosjekter/borderline-personlighetsforstyrrelse-ungdom/>
- Victor, S. E., Davis, T. & Klonsky, E. D. (2017). Descriptive Characteristics and Initial Psychometric Properties of the Non-Suicidal Self-Injury Disorder Scale. *Archives of Suicide Research*, 21(2), 265-278. DOI: 10.1080/13811118.2016.1193078
- Walsh, B. (2007). Clinical Assessment of Self-injury: A Practical Guide. *Journal of Clinical Psychology*, 63(11), 1057-1068. DOI: 10.1002/jclp.20413
- Walter, S. D., Elisziw, M. & Donner, A. (1998). Sample size and optimal designs for reliability studies. *Statistics in Medicine*, 17(1), 101-110. DOI: 10.1002/(sici)1097-0258(19980115)17:1<101::aid-sim727>3.0.co;2-e
- Washburn, J. J., Potthoff, L. M., Juzwin, K. R. & Styer, D. M. (2015). Assessing DSM-5 nonsuicidal self-injury disorder in a clinical sample. *Psychological Assessment*, 27(1), 31-41. DOI: 10.1037/pas0000021
- Watson, C. G. (1990). Psychometric posttraumatic stress disorder measurement techniques: A review. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 2(4), 460-469. DOI: 10.1037/1040-3590.2.4.460
- Watson, C. G., Juba, M. P., Manifold, V., Kucala, T. & Anderson, P. E. (1991). The PTSD interview: Rationale, description, reliability, and concurrent validity of a DSM-III-based

- technique. *Journal of Clinical Psychology*, 47(2), 179–188. DOI: 10.1002/1097-4679(199103)47:2<179::aid-jclp2270470202>3.0.co;2-p
- Weathers, F. W., Keane, T. M. & Davidson, J. R. (2001). Clinician-administered PTSD scale: A review of the first ten years of research. *Depression and Anxiety*, 13(3), 132–156. DOI: 10.1002/da.1029
- Whitlock, J., Pietrusza, C. & Purington, A. (2013). Young Adult Respondent Experiences of Disclosing Self-Injury, Suicide-Related Behavior, and Psychological Distress in a Web-Based Survey. *Archives of Suicide Research*, 17(1), 20-32. DOI: 10.1080/13811118.2013.748405
- Wilkinson, P. & Goodyer, I. (2011). Non-suicidal self-injury. *European Child and Adolescent Psychiatry*. 20(2), 103-108. DOI: 10.1177/070674371405901101
- Young, J. G., O'Brien, J. D., Gutterman, E. M. & Cohen, P. (1987). Research on the Clinical Interview. *Journal of the American Academy of Child & Adolescent Psychiatry*, 26(5), 613-620. DOI: 10.1097/00004583-198709000-00002
- Zetterqvist, M. (2015). The DSM-5 diagnosis of nonsuicidal self-injury disorder: a review of the empirical literature. *Child and Adolescent Psychiatry and Mental Health*, 9(31). DOI: 10.1186/s13034-015-0062-7
- Zetterqvist, M., Lundh, L. G., Dahlström, Ö. & Svedin, C. G. (2013). Prevalence and Function of Non-Suicidal Self-Injury (NSSI) in a Community Sample of Adolescents, Using Suggested DSM-5 Criteria for a Potential NSSI Disorder. *Journal of Abnormal Child Psychology* 41, 759–773. DOI: 10.1007/s10802-013-9712-5
- Zetterqvist, M., Perini, I., Mayo, L. M. & Gustafsson, P. A. (2020). Nonsuicidal Self-Injury Disorder in Adolescents: Clinical Utility of the Diagnosis Using the Clinical Assessment of Nonsuicidal Self-Injury Disorder Index. *Frontiers in Psychiatry*, 11(8). DOI: 10.3389/fpsy.2020.00008