DET NATIONALE FORSKNINGS-
OG ANALYSECENTER FOR VELFÆRD

Resultatdokumentation af to mentaliseringsbaserede programmer

Afsluttende evaluering af ’Styrkelse af børn og unges mentaliseringsevne, handlekraft og sociale færdigheder’

Christoffer Scavenius, Jens Hardy Sørensen, Steen Juul Hansen & Steffen Kaspersen

Resultatdokumentation af to mentaliseringsbaserede programmer – Afsluttende evaluering af 'Styrkelse af børn og unges mentaliseringsevne, handlekraft og sociale færdigheder'

© VIVE og forfatterne, 2020

e-ISBN: 978-87-7119-849-2

Forsidefoto: Lars Degnbol/VIVE

Projekt: 301141

VIVE – Viden til Velfærd
Det Nationale Forsknings- og Analysecenter for Velfærd
Herluf Trolles Gade 11, 1052 København K
www.vive.dk

VIVEs publikationer kan frit citeres med tydelig kildeangivelse.

Forord

Denne publikation er en del af projektet *Styrkelse af børn og unges mentaliseringsevne, handlekraft og sociale færdigheder.* Projektet er udbudt af Socialstyrelsen under satspuljeaftalen for 2017 (Socialstyrelsen, 2017) og afprøver to mentaliseringsbaserede programmer til døgnanbragte børn og unge: Robusthedsprogrammet (RP), som leveres af kompetenceleverandøren Komitéen for Sundhedsoplysning, og Mentaliseringsguiden (MG), som leveres af kompetenceleverandøren Center for Mentalisering. Formålet med projektet er at styrke døgnanbragte børn og unges mentaliseringsevne gennem en afprøvning af de to programmer (jf. Børne- og Socialministeriets strategi for udvikling af sociale indsatser; Børne- og Socialministeriet, 2017). VIVE – Det Nationale Forsknings- og Analysecenter for Velfærd og VIA University College har gennemført evalueringen af projektet.

Publikationen er én af fire afsluttende rapporter, som hver evaluerer et af fire emner: implementering (Hansen et al., 2020), fidelitet (Kaspersen et al., 2020), resultater (Scavenius et al., 2020) og omkostninger (Scavenius & Jonsen, 2020). De fire publikationer bygger videre på projektets midtvejsevaluering (Scavenius et al., 2019). Tilsammen genererer publikationerne viden om fire aspekter af de to programmer, som kan skabe grobund for en eventuel videre udbredelse til andre anbringelssteder. Målgruppen for publikationerne er derfor praktikere, ledere og beslutningstagere på anbringelssteder og i kommuner, som overvejer at implementere Mentaliseringsguiden eller Robusthedsprogrammet. Evalueringen er gennemført og udarbejdet af projektleder og chefanalytiker Christoffer Scavenius, psykolog Jens Hardy Sørensen, docent Steen Juul Hansen og lektor Steffen Kaspersen.

Projekter som dette kan ikke gennemføres uden hjælp og velvilje fra en række mennesker. Vi vil derfor gerne sige stor tak til alle ledere, medarbejdere og unge på de 21 anbringelssteder, som deltager i implementering og dataindsamling og dermed har gjort det muligt at evaluere de to mentaliseringsbaserede programmer. Derudover vil vi også gerne takke Henrik Kaustrup fra Foreningen af Døgn- og Dagtilbud for udsatte børn og unge for at bidrage med hans enorme netværk blandt døgntilbud samt Kamille Marie Pöckel, Bente Møller og Kirsten Margaret Højberg fra VIA University College for ihærdig implementeringsstøtte og dataindsamling. Tak til Gitte Jensen, Misja Eiberg, Marie Elbinger Gramstrup, Maja Nørgård Jacobsen, Flemming Zuschlag Christiansen, Marianne Nielsen, Camilla Overbye Roos, Mathias Andersen og Kristina Ford Folkmann, som ihærdigt har koordineret og indsamlet test med unge i undersøgelsen. En række personer skal takkes for deres faglige input og formidlingsmæssige kommentarer i forbindelse med evalueringen: Tak til Mette Laub Petersen og Charlotte Frese fra Komiteen for Sundhedsoplysning; tak til Heino Rasmussen og Janne Østergaard fra Center for Mentalisering og tak til Ida Lautrop-Rasmussen, Pernille Holzmann Poulsen, Mette Agerskov Smith, Bolette Pedersen, Kenneth Sørensen, Eva Husum Schmidt og Kirsten Brøndum fra Socialstyrelsen. Til sidst, men ikke mindst, stor tak til to anonyme reviewere for værdifulde reviewkommentarer til rapporten; deres litteraturkendskab og kritiske kommentarer var uvurderlige i forhold til at præcisere resultaterne og diskussioner.

På vegne af mig selv og forfatterne ønsker jeg dig en god læselyst!

Kræn Blume Jensen
Forsknings- og analysechef for VIVE Social
2020

Indhold

Indledning og sammenfatning 5

1 Metode 7

2 Resultater 11

3 Sammenfatning og diskussion 20

Litteratur 26

Bilag 1 Appendiks 31

Indledning og sammenfatning

Resultatdokumentation er en essentiel del af arbejdet med anbragte børn og unge, og dokumentation og opfølgning blev indført som et lovkrav med anbringelsesreformen i 2006 (Servicestyrelsen, 2007). Dokumentationen skal eksempelvis styrke vidensdeling med kollegaer eller andre anbringelsessteder og således føre til læring og udvikling af indsatsen (Socialstyrelsen, 2016, 2020). Statsrevisorerne (2015) vurderer tilsvarende, at der er betydelig *"risiko for, at kvantitet og økonomi – og ikke kvalitet og effekt – bliver bestemmende for, hvordan kommunerne anbringer udsatte børn",* hvis ikke anbringelsesindsatserne måles og dokumenteres. Sidst, men ikke mindst, stiller Socialtilsynet krav om, at alle anbringelsestilbud kan sandsynliggøre, at deres metoder har en positiv effekt for borgerne.

Denne resultatdokumentationen ser nærmere på to mentaliseringsbaserede programmer: Mentaliseringsguiden og Robusthedsprogrammet. Programmerne er beskrevet i detaljer i Scavenius et al. (2019), så her vil vi blot give en kort opsummering af programmernes indhold og struktur.

Mentaliseringsguiden (MG; Hagelquist, 2015) tilbyder konkrete pædagogiske redskaber og modeller, som er tiltænkt fagfolk, der arbejder med udsatte børn og unge. Programmet leveres af Center for Mentalisering, og Mentaliseringsguiden er udviklet på baggrund af erfaringer, der peger på, at medarbejdernes opmærksomhed på mentale tilstande har stor betydning for arbejdet med udsatte børn og unge. Mentaliseringsguidens fokus er at opbygge kompetencer for hele medarbejdergruppen, så hele miljøet omkring den unge er præget af mentalisering, og så de unge møder mentaliserende voksne i alle funktioner på anbringelsesstedet. Programmet løber over 2,5 år og indeholder uddannelse og observation af medarbejdere, psykoedukative gruppeforløb for unge, supervision af frontmedarbejdernes praksis og ledelsessparring.

Robusthedsprogrammet (RP; Bak, 2017) fokuserer på, at målgruppen bliver bedre til at håndtere udfordringer i livet. Programmet leveres af Komiteen for Sundhedsoplysning og indeholder viden og værktøjer om tanker, følelser og mentalisering. Disse værktøjer har fokus på at træne evnen til at mentalisere og derigennem blive bedre til at begå sig socialt og kunne bedømme, hvad der ligger bag egne og andres handlinger. Programmet er baseret på et billed- og hverdagssprog, hvor der ikke anvendes fagudtryk. Hensigten med sproget er at kunne kommunikere med børn, unge og voksne samt læg- og fagpersoner uanset kompetencer og baggrund. Robusthedsprogrammets implementering og uddannelse løber over et halvt år med 5 undervisningsdage, et gruppeforløb og en medarbejdersupervision.

Resultatdokumentationen ser nærmere på udviklingen i de unges mentaliseringsevne og socioemotionelle kompetencer. Resultaterne undersøger de unges udvikling fra programmernes igangsættelse på anbringelsesstederne i 2018, til den opfølgende måling ved projektets afslutning i 2020, og sammenligner udviklingen på tværs af de to programmer og en sammenlignings-gruppe. Data i denne rapport beskriver således de unges udvikling i perioden, hvor personalet på anbringelsesstederne har implementeret de to mentaliseringsbaserede programmer.

Resultaterne bygger videre på resultaterne fra projektets midtvejsevaluering (Scavenius et al., 2019). Midtvejsevalueringen præsenterede data fra de første test af de unges mentaliseringsevne og socioemonelle kompetencer. Baselinemålingen viste blandt andet, at de unges evne til at forestille sig og beskrive følelser hos sig selv og andre lå under et alderssvarende niveau, hvorimod deres evne til at beskrive andres følelser ud fra ansigtsudtryk lå på et alderssvarende niveau. Ved baselinemålingen rapporterede de unge også markant større socioemotionelle vanskeligheder end andre unge på samme alder (Arnfred et al., 2019).

Hovedresultaterne i denne rapport supplerer resultaterne fra midtvejsevalueringen og finder **primært to tendenser i data:** Vi ser dels nogle positive tendenser i de unges evne til at mentalisere andres følelser ud fra ansigtsudtryk, dels nogle positive tendenser i den generelle socioemotionelle udvikling. Derudover finder vi kun få signifikante ændringer i de unges evne til at mentalisere indre følelser hos sig selv og andre og kun få signifikante forskelle mellem de tre indsatsgrupper.

Resultaterne skal dog læses med forbehold for evalueringsdesignet, uens implementering, ophør i anbringelse og et betydeligt frafald fra evalueringen. Evalueringen er designet som en før-efter-måling af unge på anbringelsessteder, som (1) implementerer Mentaliseringsguiden eller (2) Robusthedsprogrammet eller (3) ikke implementerer et af de to programmer. Implementeringen af programmerne er ikke tilfældig (randomiseret), vi kan derfor ikke med sikkerhed konkludere, om de unges udvikling skyldes programmerne eller en naturlig udvikling, fordi de unge er blevet ældre henover perioden. Derudover bygger evalueringen på test af 102 unge på 21 anbringelsessteder, men i løbet af projektperioden er 43 pct. af de unge fraflyttet anbringelsesstedet, og 43 pct. (N = 44) er frafaldet evalueringen. Ophør og frafaldet kan påvirke resultaterne, specielt for unge i Robusthedsprogrammet og i sammenligningsgruppen, hvor vi har færrest observationer. Samtidig har implementeringen været uens på anbringelsesstederne, 4-5 anbringelsessteder har eksempelvis kun arbejdet med programmerne i lav grad (Hansen et al., 2020). Vi kan derfor ikke konkludere på effekten af de to programmer i forhold til en sammenligningsgruppe. Med de indsamlede data kan vi kun beskrive tendenser i de unges udvikling i de to programmer i forhold til sammenligningsgruppen og generere hypoteser om disse fund, som man bør undersøge nærmere.

Resten af denne rapport er opdelt i et metodeafsnit, resultatafsnit og diskussionsafsnit. Metodeafsnittet begynder med en boks, som beskriver dataindsamlingsproceduren, herefter følger en beskrivelse af målgruppen samt endnu en boks, som beskriver de indsamlede data. Det efterfølgende resultatafsnit illustrerer og beskriver de unges udvikling henover perioden ved hjælp af en række grafer. Alle rapportens resultater sammenfattes og diskuteres i det sidste afsnit, hvor vi også diskuterer begrænsninger i evalueringen.

1 Metode

Afsnittet beskriver evalueringsmetoderne, herunder rekrutterings- og dataindsamlingsprocedure, demografiske og andre baggrundskarakteristika ved målgruppen samt de indsamlede mentaliseringstest og spørgeskemaer. Læsere, som kun er interesseret i resultaterne, kan springe dette afsnit over eller vende tilbage til det senere uden at miste hovedkonklusionerne i denne rapport.

Boks 1.1 Rekrutterings- og dataindsamlingsprocedure

Rekrutteringen af anbringelsessteder er gennemført i to skridt: Først rekrutterede vi unge fra anbringelsessteder, som afprøver Robusthedsprogrammer (RP, 3 anbringelsessteder) eller Mentaliseringsguiden (MG, 11 anbringelsessteder). Dernæst rekrutterede vi unge fra lignende anbringelsessteder til en sammenligningsgruppe (SG). For at finde sammenlignelige anbringelsessteder bad vi både Foreningen for Døgn- og Dagtilbud (FADD) og alle RP- og MG-anbringelsesstederne om at udpege søsterorganisationer eller andre anbringelsessteder med en lignende målgruppe af unge. Proceduren resulterede i 7 anbringelsessteder. Alt i alt medvirker der derfor 21 anbringelsessteder i resultatdokumentationen.

Ved hvert anbringelsessted blev de unge rekrutteret ud fra tre kriterier: For det første skulle anbringelsesstedet vurdere, at den unge med stor sandsynlighed ville være på anbringelsesstedet i 2020, for det andet skulle de unge være mellem 12 og 20 år ved baselinemålingen, og for det tredje måtte de unge ikke have diagnosen infantil autisme. Samtidig skulle de unge selv ønske at deltage i undersøgelsen. Undersøgelsen inkluderer derfor kun en mindre gruppe af alle unge på anbringelsesstederne.[1] Ingen af de unge i undersøgelsen er anbragt på grund af fysiske handikap.

Dataindsamlingen er gennemført som en baseline- og eftermåling. Medio 2018 og medio 2020 testede otte psykologer de unges mentaliseringsevne. Testen blev gennemført på anbringelsesstedet og tog omkring en time for hver ung. Alle psykologerne er uddannet i testbatteriet, og alle testene er efterfølgende gennemgået for ensartethed og fejlkodninger. Gennemgangen resulterede ikke i omkodninger. I andet kvartal 2019 indsamlede vi derudover supplerende oplysninger om de unges baggrund fra en medarbejder på anbringelsesstedet, som kender den unge godt.

I løbet af projektet valgte vi at tilpasse rekrutterings- og dataindsamlingsproceduren for at styrke stikprøvestørrelsen i Robusthedsprogrammet. På to anbringelsessteder, som indgår i Robusthedsprogrammet, falder mange af de unge uden for alders- og stabilitetskriteriet (bl.a. et, hvor målgruppen er unge i misbrugsbehandling; Scavenius et al., 2019). I rekrutteringsfasen valgte vi derfor at udvide alderskriteriet til 23 år og slække på kriteriet for stabilitet i anbringelsen for unge i Robusthedsprogrammet. Anbringelsesstederne i Robusthedsprogrammet var derudover udfordret af flere sammenbrud eller planlagte ophør i anbringelsen end anbringelsessteder i Mentaliseringsguiden og sammenligningsgruppen (ibid.). For at imødegå dette besluttede vi at gennemføre en ekstra test af de unges mentaliseringsevne medio 2019. For unge i Robusthedsprogrammet har vi derfor potentielt tre målepunkter. Desværre skete der en fejl i indsamlingen af socioemotionelle vanskeligheder på Sjælland, hvor psykologerne kun indsamlede spørgeskemaet for nogle af de unge. Vi forsøgte efterfølgende at rette op på fejlen ved at indhente manglende målinger ved fremmøde, telefonisk eller over e-mail, men som det fremgår af appendikset, er dataindsamlingen ikke komplet.

Ved baselinemålingen har vi indsamlet 102 test fra unge, hvoraf 19 unge er på anbringelsessteder, som implementerer Robusthedsprogrammet (RP), 62 unge på anbringelsessteder, som implementerer Mentaliseringsguiden (MG), og 21 unge er på sammenlignelige anbringelses-

[1] Ud fra vores kendskab til antallet af unge på de 14 steder, som implementerer Mentaliseringsguiden og Robusthedsprogrammet, er en forsigtig vurdering, at vi har ca. 20 procent af alle unge på anbringelsesstederne med i undersøgelsen.

steder (SG), som ikke implementerer et af de to programmer (se Boks 1.1). Ved midtvejsmålingen blev indsamlet 14 test fra de unge i Robusthedsprogrammet (73 pct.). I forbindelse med eftermålingen har vi indsamlet 58 test (57 pct. af de 102 baselinemålinger), hvoraf 6 (32 pct.) unge er i RP-gruppen, 42 (68 pct.) unge er i MG-gruppen og 10 (48 pct.) unge er i SG. Vi har gennemført en frafaldsanalyse, der peger på, at unge, som falder fra undersøgelsen, i gennemsnit har 0,5 færre anbringelser i bagagen og er 1,6 år ældre, ellers er der ingen overrepræsentation i forhold til køn, alder ved første anbringelse eller længden af anbringelsen ved anbringelsesstedet.[2]

Scavenius et al. (2019) beskriver karakteristika for unge i stikprøven. Overordnet består stikprøven af lidt flere drenge (N = 57 eller 55 pct.) end piger, som ved rekrutteringen var 15,3 år i gennemsnit. Den yngste er 11 år og den ældste 23 år. I gennemsnit er de unge på SG-anbringelsesstederne yngst (14,2 år), efterfulgt af MG (15,0 år) og RP (17,4 år). 4 unge i RP er ældre end målgruppekriteriet på 20 år. Ved første anbringelse var de unge i gennemsnit 12,0 år, og de har siden første anbringelse været anbragt mellem 1 og 5 forskellige steder (i gennemsnit 2,1 steder). I andet kvartal 2019 var 85 pct. af de unge stadig anbragt samme sted, og midt i 2020 var 57 pct. af de unge anbragt på samme sted. Fordelt på de tre grupper var 84 pct., 53 pct. og 95 pct. stadig anbragt på hhv. SG-, RP- og MG-anbringelsesstedet i andet kvartal 2019. Ved eftermålingen, midt i 2020, var 63 pct., 42 pct. og 60 pct. stadig anbragt på hhv. SG-, RP- og MG-anbringelsesstedet.

De unge er ofte anbragt på grund af overgreb, dysfunktionelle familieforhold eller omsorgssvigt. De primære overgreb i hjemmet er: psykiske overgreb (fx verbale overgreb eller trusler mod barnet, 21 pct.), fysiske overgreb (fx vold eller trusler om vold mod barnet, 19 pct.) og seksuelle overgreb (9 pct.). De primære dysfunktionelle forhold er: vold mod mor, far eller anden omsorgsperson (23 pct.), trusler om vold mod mor, far eller anden omsorgsperson (23 pct.), misbrugsproblemer i hjemmet (20 pct.), psykiske lidelser hos forældre (18 pct.), skilsmisse eller separation (16 pct.), forældre afgået ved døden (5 pct.) og kriminel adfærd i hjemmet (5 pct.). De primære omsorgssvigt er: psykisk omsorgssvigt (fx manglende kærlighed, tryghed eller stimulation, 66 pct.) og fysisk omsorgssvigt (fx utilstrækkelig ernæring, beklædning eller hygiejne, 21 pct.). For 6 ud af 7 unge (85 pct.) skyldes anbringelsen én eller flere af disse årsager, for 66 pct. skyldes anbringelsen to eller flere af disse årsager, og for 37 pct. skyldes anbringelsen tre eller flere af disse årsager.

De primære vanskeligheder hos de unge viser sig som: problemer i fritiden (fx med venner, 48 pct.), skoleproblemer (44 pct.), voldelig eller aggressiv adfærd (32 pct.), psykiske lidelser (26 pct.), selvskade (16 pct.) og misbrug (9 pct.). Overordnet er vanskelighederne hos halvdelen af de unge udadreagerende (54 pct.), og hos en tredjedel er vanskelighederne indadreagerende (33 pct.). En læge, psykiater eller psykolog har stillet en diagnose hos 64 pct. af de unge, typisk ADHD (21 pct.), Angst (7 pct.), Autisme spektrum (7 pct.), Selvskade (6 pct.) eller Borderline (6 pct.); andre diagnoser optræder kun i mindre omfang.

[2] I resultatafsnittet beskriver vi resultater fra en robusthedsanalyse, som forsøger at tage højde for forskelle i baselinekarakteristika mellem de unge. Derudover sammenligner vi resultaterne fra robusthedsanalysen med resultater fra den præsenterede analyse for at undersøge om forskelle i antal anbringelser og alder eventuelt kan have påvirket resultaterne.

Boks 1.2 Instrumenter til evaluering af de unges mentaliseringsevne og sociale kompetencer

Til undersøgelsen er udvalgt **fire instrumenter**: tre test til vurdering af mentaliseringsevnen og et spørgeskema til vurdering af sociale vanskeligheder. Instrumenterne er valgt, fordi de opfylder tre kriterier: de er oversat til dansk, de dækker karakteristiske mentaliseringsevner, og de gennemføres under de praktiske begrænsninger i projektet (Scavenius et al., 2019).

Level of Emotional Awareness Scale (LEAS, Lane & Schwartz, 1987; Lane, 2015; Luyten et al., 2012) består af 12 situationer, der hver er beskrevet med 2-4 sætninger, som involverer testpersonen selv og en anden person (Lane et al., 1990). Hver situation omhandler én af fire følelser: glæde, vrede, tristhed eller frygt. **Testen vurderer testpersonens mentalisering af sig selv (LEAS Selv) og den anden (LEAS Anden) samt balancen mellem de to mentaliseringsevne (LEAS Total)**. Hver situation scores på en skala fra 0 til 5. En score på 0 repræsenterer svar med manglende følelser (fx *jeg ville ikke føle noget; jeg ville føle, at han gjorde det med vilje*). En score på 1 repræsenterer et somatisk niveau for følelser (fx *jeg ville få ondt i maven; jeg ville føle mig syg*). En score på 2 repræsenterer et handlingsniveau for følelser (fx *jeg ville føle, at jeg skulle smadre noget*) eller en mere global følelse (*jeg ville føle, det var helt okay; jeg ville føle, det var noget værre skidt*). En score på 3 repræsenterer specifikke, men singulære følelser (fx *jeg ville føle mig glad, ked af det, vred eller bange*). En score på 4 repræsenterer evnen til at have og udtrykke flere enkeltfølelser samtidigt (fx *jeg ville føle mig lidt ked af det, men også vred*). I LEAS Total repræsenterer en score på 5 evnen til sammensatte følelser hos både sig Selv og den Anden (fx *jeg ville føle mig ked af det og vred, og hun ville føle sig vred og lidt bange*). LEAS-skalaen hviler på udviklingspsykologisk teori og måler respondentens udviklingstrin. Skalaen er en såkaldt Guttman-skala (Guttman, 1944; Rasch, 1960), hvor højere niveauer indbefatter evner på lavere udviklingstrin (Lane & Garfield, 2005; Lane et al., 2015; Panksepp 1998; Panksepp & Biven, 2012, 2012a). Hvis en ung fx scorer 3 på LEAS, dvs. kan sætte ord på enkeltemotioner, så kan vedkommende også mentalisere på de lavere niveauer, 1 og 2, hhv. somatisk og handlingsniveau.

Reading the Mind in the Eyes (RME, Baron-Cohen et al., 2001; Bateman & Fonagy, 2016, 2019; Luyten et al., 2012, 2019; Malberg & Rosenbery, 2017) består af 28 billeder af øjenomgivelserne hos voksne mænd og kvinder. Billederne indeholder forskellige emotioner (fx vrede, glæde, afsky) med varierende intensitet, og testpersonen skal til hvert billede vælge et følelsesudtryk ud af fire mulige. **Testen undersøger respondentens evne til at mentalisere emotionelle tilstande hos andre ud fra deres ansigtsudtryk.** Evnen til at aflæse emotionelle tilstande hos andre sker især ud fra iagttagelse af ansigtsudtryk, specielt i øjenregionen (Elfenbein & Ambady, 2002; Ekman, 2003; Niedenthal, 2007; Plutchick, 2003). Forringet afkodning af emotionelle ansigtsudtryk kan have negative konsekvenser for ens sociale interaktion (Luyten et al., 2019). Hvis man kan aflæse, at den anden er vred, vil det medføre en anden type interaktion, end hvis man aflæser, at den anden er glad. Antallet af korrekt aflæste ansigtsudtryk giver en samlede RME-score på mellem 0 og 28.

Test med **Ekmans billeder** (Sørensen, 2009, 2012, 2013)**, som undersøger evnen til at mentalisere egne følelser ud fra andres ansigtsudtryk**. Til forskel fra RME-testen fokuserer denne test på egne følelser frem for følelsen hos personen på billedet. I testen instruerer psykologen den unge med følgende*: "Jeg vil nu vise dig billeder af en ung kvinde med forskellige følelsesudtryk"; "Forestil dig, det er en kvinde, du kender godt – en god ven, en søster eller lignende"; "Kig på billedet og fortæl, hvilke følelser det vækker i dig. Hvilke følelser får du, når du ser udtrykket i billedet?"*. Psykologen viser herefter seks billeder af en ung kvinde med forskellige emotionelle udtryk. Stimulus materialet består af Ekmans billeder af seks emotionelle udtryk (Ekmans, 2003). For hvert billede scores respondentens svar på kvindens følelsesudtryk på en skala fra 0 til 4 ud fra LEAS-kriterierne. Testens validitet er endnu ikke publiceret; der findes således ikke studier, som undersøger, om resultater af testen kan tolkes som 'mentalisering af egne følelser'. Da der findes kun få valide test af mentalisering, som er oversat til dansk, har vi medtaget testen med Ekmans billeder på trods af denne manglende undersøgelse af validitet.

Strengths and Difficulties Questionnaire (SDQ, Goodman 1997; Arnfred et al., 2019) **består af 25 spørgsmål om den unges socioemotionelle vanskeligheder** inden for de sidste 6 måneder. 20

Boks 1.2 Instrumenter til evaluering af de unges mentaliseringsevne og sociale kompetencer

spørgsmål samles til en score for *Sociale vanskeligheder* (skala fra 0 til 40), hvor en højere score indikerer flere vanskeligheder, og 5 spørgsmål samles i *Sociale styrker* (skala fra 0 til 10). I denne evaluering bruger vi et udvidet SDQ-skema med 5 yderligere spørgsmål omkring betydningen af sociale vanskeligheder for *Daglig funktion* (skala fra 0 til 10), hvor en højere score indikerer større vanskeligheder i dagligdagen.

2 Resultater

Dette afsnit viser de unges udvikling i mentaliseringsevne og sociale kompetencer fra medio 2018 til medio 2020 samt beskriver forskelle i udviklingen mellem unge i sammenligningsgruppen (SG), Robusthedsprogrammet (RP) og Mentaliseringsguiden (MG). Resultaterne bygger på lineære regressioner over de tre dataindsamlinger. Resultater for Robusthedsprogrammet er derfor beregnet ud fra observationer i 2018, 2019 og 2020, hvorimod resultater for Mentaliseringsguiden og sammenligningsgruppen er beregnet ud fra observationer i 2018 og 2020. De præcise beregningerne, p-værdier, konfidensintervaller, Cohen's D effektstørrelser (ES) og de statistiske test, som ligger til grund for resultaterne, er beskrevet i fodnoterne og i rapportens appendiks. I dette afsnit beskriver vi blot resultaterne af analysen; i næste afsnit sammenfatter og diskuterer vi disse resultater.

De unges udvikling illustreres i otte figurer, hvor den vandrette akse gengiver udviklingen over tid, og den lodrette akse viser de unges gennemsnitlige mentaliseringsevne (testene LEAS, RME og Ekmans billeder) eller sociale vanskeligheder og styrker (SDQ spørgeskemaet). Hældningen på kurverne i figurerne viser således de unges udvikling henover projektperioden. Niveauet for kurverne i 2018 afspejler forskelle mellem de unge i de tre grupper ved baseline (fx forskelle i køn, alder, diagnoser, anbringelseshistorie eller andre konstante baggrundskarakteristika, som kan være relateret til mentaliseringsevnen eller sociale vanskeligheder). Beskrivelsen af kurverne fokuserer hovedsagligt på udviklingen (hældningen på kurverne) og har mindre fokus på forskelle ved baseline (hvilket er gennemgået i Scavenius et al., 2019). Derudover fokuserer beskrivelsen på tre typer at resultater: statistisk signifikante (p-værdi < 0,05) *ændringer*, marginalt signifikante (p-værdi < 0,10) *tendenser* og statistisk insignifikante (p-værdier ≥ 0,10) resultater, der beskrives som en *uændret* udvikling.

De unges udvikling i LEAS-målingerne for selv, andre og totalt er vist i Figur 2.1 til Figur 2.3. **LEAS-scoren for selvorienteret mentalisering** – dvs. evnen til at kunne reflektere over og benævne følelser hos sig selv – ligger ved baselinemålingen i gennemsnit på 2,3 for hele ungegruppen samlet. Ved den afsluttende måling er niveauet for hele ungegruppen i gennemsnit uændret.[3] Tilsvarende ser vi ingen statistisk signifikant udvikling i selvorienteret mentalisering, når vi opdeler resultaterne på de tre grupper. Sammenligner vi udviklingen i de to programmer med udviklingen i sammenligningsgruppen, er der ingen statistisk forskel mellem grupperne, hvilket vil sige, at vi ikke kan se forskel på hældningerne i de tre kurver, når vi tager højde for den statistiske usikkerhed.[4]

[3] En lineær regression, som estimerer udviklingen for hele ungegruppen samlet (dvs. uden en interaktionseffekt med indsatsgruppen) har en hældning på -0,0 (ES=0,00; p-værdi = 0,99).

[4] Robusthedsanalysen finder tilsvarende resultater. I statistiske modeller, som forsøger at tage højde for frafaldet (multipel imputation modeller) eller forskelle mellem grupperne ved baseline (fixed effects modeller), finder vi den samme forholdsvise udvikling mellem sammenligningsgruppen og de to programmer.

Figur 2.1 Udvikling i mentalisering af egne følelser opdelt på de tre indsatsgrupper

På **LEAS-scoren for mentalisering af andre** – dvs. evnen til at kunne reflektere over og benævne følelser hos andre – lå de unges mentaliseringsniveau i 2018 i gennemsnit på 2,1 for hele ungegruppen samlet. Ved målingen i 2020 var niveauet for hele ungegruppen i gennemsnit uændret.[5] Ser vi på udviklingen i hver af de tre grupper, er scoren i sammenligningsgruppen og Mentaliseringsguiden statistisk uændret, men den stiger i Robusthedsprogrammet med 0,5 point fra 2,0.[6] Sammenligner vi udviklingen i de to programmer med udviklingen i sammenligningsgruppen, henover projektperioden på 2 år, har unge i Robusthedsprogrammet haft en forholdsvis positiv ændring på 0,8 point.[7] For unge i Mentaliseringsguiden ser vi en positiv tendens på 0,4 point i forhold til sammenligningsgruppen, som dog kun er marginalt signifikant (p-værdi på 0,09).[8]

[5] En lineær regression, som estimerer udviklingen for hele ungegruppen samlet (dvs. uden en interaktionseffekt med indsatsgruppen), har en hældning på 0,1 (ES=0,09; p-værdi = 0,61).

[6] ES = 0,83; p-værdi < 0,01. Cohens d effektstørrelsen (ES) skal tolkes med samme forsigtighed for evalueringsdesignet og frafaldet som den point-vise ændring. Se begrænsninger i diskussionsafsnittet.

[7] ES = 1,42; p-værdi < 0,01. Cohens d effektstørrelsen (ES) skal tolkes med samme forsigtighed for evalueringsdesignet og frafaldet som den point-vise ændring. Se begrænsninger i diskussionsafsnittet.

[8] Robusthedsanalysen finder stort set tilsvarende resultater. I statistiske modeller, som forsøger at tage højde for frafaldet (multipel imputation modeller) eller forskelle mellem grupperne ved baseline (fixed effects modeller), finder vi stort set den samme forholdsvise udvikling mellem sammenligningsgruppen og de to programmer. Den eneste forskel er, at den forholdsvise positive udvikling på 0,4 point i Mentaliseringsguiden stiger til 0.5 og bliver signifikant (p-værdi = 0,05) i en fixed-effect model.

Figur 2.2 Udvikling i mentalisering af andres følelser opdelt på de tre indsatsgrupper

Appendix tabel 1 viser de statistiske beregninger

I den **Totale LEAS**-score, som måler evnen til balanceret at kunne reflektere over og benævne følelser hos både sig selv og andre, ligger hele ungegruppen i gennemsnit på niveau 2,6 i 2018. Ved den afsluttende måling er niveauet for hele ungegruppen i gennemsnit uændret.[9] Ser vi på udviklingen i hver af de tre grupper, falder scoren i sammenligningsgruppen med 0,4 point fra 2,9 point,[10] og den er uændret i Robusthedsprogrammet og Mentaliseringsguiden. Sammenligner vi udviklingen i de to programmer med udviklingen i sammenligningsgruppen, har unge i Robusthedsprogrammet haft en forholdsvis positiv ændring på 0,6 point over de 2 år.[11] For unge i Mentaliseringsguiden ser vi en positiv tendens i data på 0,4 point i forhold til sammenligningsgruppen, som dog kun er marginalt signifikant (p-værdi på 0,07).[12]

[9] En lineær regression, som estimerer udviklingen for hele ungegruppen samlet (dvs. uden en interaktionseffekt med indsatsgruppen) har en hældning på 0,1 (ES = 0,08; p-værdi = 0,63).

[10] ES = 0,70; p-værdi = 0,04. Cohens d effektstørrelsen (ES) skal tolkes med samme forsigtighed for evalueringsdesignet og frafaldet som den point-vise ændring. Se begrænsninger i diskussionsafsnittet.

[11] ES = 0,91; p-værdi = 0,02. Cohens d effektstørrelsen (ES) skal tolkes med samme forsigtighed for evalueringsdesignet og frafaldet som den point-vise ændring. Se begrænsninger i diskussionsafsnittet.

[12] Robusthedsanalysen finder stort set tilsvarende resultater. I statistiske modeller, som forsøger at tage højde for frafaldet (multipel imputation modeller) eller forskelle mellem grupperne ved baseline (fixed-effects modeller), finder vi næsten den samme forholdsvise udvikling mellem Sammenligningsgruppen og de to programmer. Den eneste forskel er, at stigningen på 0,6 point i Robusthedsprogrammet falder til 0,5 point og bliver marginalt,insignifikant (p-værdi = 0,7) i en fixed-effect model.

Figur 2.3 Udvikling i balanceret mentalisering af både egne og andres følelser opdelt på de tre indsatsgrupper

Appendix tabel 1 viser de statistiske beregninger

Samlet set tyder resultaterne for de tre LEAS-score på, at evnen til at kunne reflektere over og benævne følelser hos sig selv og andre kan være faldet i sammenligningsgruppen i forhold til de to programmer. Denne tendens er svag i data, og flere estimationer er kun marginalt signifikante, men resultatet peger på, at mentaliseringsevnen ikke er steget for unge i de to programmer, men fordi mentaliseringsevnen i sammenligningsgruppen kan være faldet, ser vi en forholdsvis forbedring i de to programmer relativt til sammenligningsgruppen. Resultatet er som sagt kun marginalt signifikant, og forskellen mellem sammenligningsgruppen og de to programmer kan skyldes evalueringsdesignet og de få observationer i specielt sammenligningsgruppen og Robusthedsprogrammet.

RME-målingen for mentalisering af andres ydre karakteristika er vist i Figur 2.4. RME-scoren fokuserer på evnen til at afkode og benævne ydre udtryk af følelser hos andre, aflæst i øjenområdet. Scoren går fra 0 til 28 point, et point for hvert korrekt genkendt følelse. Læg mærke til, at y-aksen i grafen er skåret mellem 14 og 26 point for at øge læsbarheden. I gennemsnit genkendte de unge 19 ud af 28 følelser ved baselinemålingen, hvilket svarer til, at de kan genkende den korrekte følelse i 68 pct. af tilfældene. Ved den afsluttende måling er genkendelsen af billeder for hele ungegruppen i gennemsnit steget med 1,1 billede, hvilket svarer til, at de kan genkende den korrekte følelse i 72 pct. af tilfældene.[13] Ser vi på udviklingen i hver af de tre grupper, er scoren statistisk uændret i sammenligningsgruppen på 19 korrekt genkendte billeder. Det vil sige, at selvom grafen har en positiv hældning, er denne hældning ikke statistisk

[13] En lineær regression, som estimerer udviklingen for hele ungegruppen samlet (dvs. uden en interaktionseffekt med indsatsgruppen), har en signifikant hældning på 1,1 (ES = 0,32; p-værdi = 0,01).

signifikant; når vi tager højde for den statistiske usikkerhed, kunne grafen for sammenlignings-gruppen lige så godt være flad.[14] I Robusthedsprogrammet stiger genkendelsen af korrekte følelser med 1,1 over 2 år fra 20,6 ved baselinemålingen.[15] I Mentaliseringsguiden ser vi en tendens til, at genkendelsen af korrekte følelser stiger med 1,1 over 2 år fra 18,4 ved baseline-målingen; statistisk er denne udvikling dog kun marginalt signifikant.[16] Sammenligner vi udvik-lingen i de to programmer med udviklingen i sammenligningsgruppen, er der ingen statistisk forskel mellem grupperne, hvilket vil sige, at vi ikke kan se forskel på hældningerne i de tre kurver, når vi tager højde for den statistiske usikkerhed.[17]

Figur 2.4 Udvikling i mentalisering af andres ansigtsudtryk opdelt på de tre indsatsgrupper

Appendix tabel 1 viser de statistiske beregninger

Testen med **Ekmans billeder** er vist i Figur 2.5. Modsat RME-scoren har denne score fokus på at afkode og benævne følelser hos sig selv som reaktion på andres ansigtsudtryk – dvs. den forsøger at opfange indre mentalisering af andres ydre karakteristika. I gennemsnit lå de unge på mentaliseringsniveau 2,0 i 2018. Ved den afsluttende måling er niveauet for hele un-gegruppen i gennemsnit uændret.[18] Tilsvarende ser vi ingen udvikling i scoren på denne test, når vi opdeler resultaterne på de tre grupper. Og vi ser heller ingen forskel i udviklingen, når vi

[14] Stigningen er marginalt insignifikant (p-værdi = 0,14).

[15] ES = 0,32; p-værdi = 0,03. Cohens d effektstørrelsen (ES) skal tolkes med samme forsigtighed for evalueringsdesignet og frafaldet som den point-vise ændring. Se begrænsninger i diskussionsafsnittet.

[16] P-værdien er 0,07. Robusthedsanalysen finder tilsvarende resultater. I en model, som forsøger at tage højde for frafaldet (multipel imputation modeller), finder vi en stigning på 1,2 og en p-værdi på 0,05. I en model, som forsøger at kontrollere for forskelle mellem grupperne ved baseline (fixed-effects modeller), finder vi en stigning på 1,1 og en p-værdi på 0,13.

[17] I statistiske modeller, som forsøger at tage højde for frafaldet (multipel imputation modeller) eller forskelle mellem grupperne ved baseline (fixed-effects modeller), finder vi den samme forholdsvise udvikling mellem sammenligningsgruppen og de to programmer.

[18] En lineær regression, som estimerer udviklingen for hele ungegruppen samlet (dvs. uden en interaktionseffekt med indsats-gruppen), har en marginalt signifikant hældning på -0,3 (ES = 0,32; p-værdi = 0,05).

sammenligner de to programmer med sammenligningsgruppen.[19] Det vil sige, at hældningen på de tre kurver kunne lige så godt være flad, når vi tager højde for den statistiske usikkerhed.

Figur 2.5 Udvikling i mentalisering af andres ansigtsudtryk på egne følelser opdelt på de tre indsatsgrupper

SDQ-målingerne for sociale vanskeligheder, sociale styrker og daglig funktion er vist i Figur 2.6 til Figur 2.8. Den **samlede problemscore i SDQ** går fra 0 til 40 point, hvor flere point angiver flere socioemotionelle vanskeligheder. En score mellem 0 og 14 point svarer til en normalpopulation uden sociale vanskeligheder, 15-17 point svarer til let forhøjede vanskeligheder, og en score over 18 point svarer til store eller meget store vanskeligheder (Arnfred et al., 2019). Læg mærke til, at y-aksen i grafen er skåret mellem 12 og 17 point for at øge læsbarheden. På den samlede problemscore opnår de unge i gennemsnit 15,1 point ved den første måling. Ved den afsluttende måling er gennemsnittet for hele ungegruppen faldet med 1,4 point.[20] Opdeler vi udviklingen på de tre grupper, er scoren statistisk set uændret i sammenligningsgruppen og i Robusthedsprogrammet.[21] Det vil sige, at når vi tager højde for den statistiske usikkerhed i datamaterialet, kunne hældningen på kurverne for sammenligningsgruppen og Robusthedsprogrammet lige så godt være flad. I Mentaliseringsguiden falder niveauet af socioemotionelle vanskeligheder derimod med 1,6 point over 2 år fra 15,2 point ved baseline-

[19] Robusthedsanalysen finder tilsvarende resultater. I statistiske modeller, som forsøger at tage højde for frafaldet (multipel imputation modeller) eller forskelle mellem grupperne ved baseline (fixed-effects modeller), finder vi den samme forholdsvise udvikling mellem sammenligningsgruppen og de to programmer.

[20] En lineær regression, som estimerer udviklingen for hele ungegruppen samlet (dvs. uden en interaktionseffekt med indsatsgruppen), har en hældning på -1,4 (ES = 0,26; p-værdi = 0,04).

[21] Appendiks tabellen viser et fald på 1,4 point (p-værdi=0,48) i sammenligningsgruppen og et fald på 0,9 (p-værdi = 0,61) i Robusthedsprogrammet.

målingen og går således i gennemsnit fra kategorien 'let forhøjede vanskeligheder' til kategorien 'ingen vanskeligheder'.[22] Sammenligner vi udviklingen i de to programmer med udviklingen i sammenligningsgruppen, er der ingen statistisk forskel mellem grupperne, hvilket vil sige, at vi ikke kan se forskel på hældningerne i de tre kurver, når vi tager højde for den statistiske usikkerhed.[23]

Figur 2.6 Udvikling i socioemotionelle vanskeligheder opdelt på de tre indsatsgrupper

SDQ-scoren for sociale styrker går fra 0 til 10 point, hvor flere point angiver flere sociale styrker. En score mellem 6 og 10 point svarer til normale sociale styrker, 0-5 point svarer til lave eller meget lave sociale styrker (Arnfred et al., 2019). Læg mærke til, at y-aksen i grafen er skåret mellem 5 og 9 point for at styrke læsbarheden. På den score for sociale styrker opnår de unge i gennemsnit 7,5 point ved den første måling i 2018. Ved den afsluttende måling er gennemsnittet for hele ungegruppen uændret. Opdeler vi udviklingen på de tre grupper, er scoren statistisk set uændret i Sammenligningsgruppen, Robusthedsprogrammet og Mentaliseringsguiden.[24] Det vil sige, at når vi tager højde for den statistiske usikkerhed i datamaterialet kunne hældningen på kurverne for alle tre grupper lige så godt være flad. Sammenligner vi

[22] ES=0,30; p-værdi=0,04. Cohens d effektstørrelsen (ES) skal tolkes med samme forsigtighed for evalueringsdesignet og frafaldet som den point-vise ændring. Se begrænsninger i diskussionsafsnittet.

[23] Robusthedsanalysen finder tilsvarende resultater. I statistiske modeller, som forsøger at tage højde for frafaldet (multipel imputation modeller) eller forskelle mellem grupperne ved baseline (fixed-effects modeller), finder vi den samme forholdsvise udvikling mellem sammenligningsgruppen og de to programmer.

[24] Appendiks tabellen viser et fald på 0,2 point (p-værdi = 0,87) i sammenligningsgruppen, et fald på 0,4 (p-værdi = 0,76) i Robusthedsprogrammet og et fald på 0,3 i Mentaliseringsguiden (p-værdi = 0,23).

udviklingen i de to programmer med udviklingen i sammenligningsgruppen, er der ingen statistisk forskel mellem grupperne, hvilket vil sige, at vi ikke kan se forskel på hældningerne af kurverne, når vi tager højde for den statistiske usikkerhed.[25]

Figur 2.7 Udvikling i sociale styrker opdelt på de tre indsatsgrupper

Appendix tabel 1 viser de statistiske beregninger

SDQ-scoren for daglig funktion går fra 0 til 10 point. Scoren ser på vanskeligheder med følelser, koncentration, opførsel eller social interaktion, som påvirker daglig trivsel, familieliv, venskaber, indlæring eller fritidsaktiviteter. Flere point angiver flere vanskeligheder. En score på 0 point svarer til, at vanskelighederne ikke påvirker den unges daglige trivsel, familieliv, venskaber, indlæring eller fritidsaktiviteter, 1 point svarer til, at vanskelighederne giver en lettere påvirkning i dagligdagen, 2-3 point svarer til, at vanskelighederne giver en stor påvirkning i dagligdagen, og over 4 point svarer til, at vanskelighederne har en stor påvirkning på daglige funktioner (Arnfred et al., 2019). Læg mærke til, at y-aksen i grafen er skåret mellem 0 og 4 point for at styrke læsbarheden. På scoren for daglig funktion opnår de unge i gennemsnit 1,9 point ved den første måling. Ved den afsluttende måling er gennemsnittet for hele ungegruppen uændret.[26] Opdeler vi udviklingen på de tre grupper, er scoren statistisk uændret i alle tre grupper.[27] Det vil sige, at når vi tager højde for den statistiske usikkerhed i datamaterialet, kunne hældningen på kurverne for alle tre grupper lige så godt være flad. Sammenligner vi

[25] Robusthedsanalysen finder tilsvarende resultater. I statistiske modeller, som forsøger at tage højde for frafaldet (multipel imputation modeller) eller forskelle mellem grupperne ved baseline (fixed-effects modeller), finder vi den samme forholdsvise udvikling mellem Sammenligningsgruppen og de to programmer.

[26] En lineær regression, som estimerer udviklingen for hele ungegruppen samlet (dvs. uden en interaktionseffekt med indsatsgruppen), har en hældning på -0,3 (ES=0,07; p-værdi = 0,53).

[27] Appendiks tabellen viser en stigning på 0,3 point (p-værdi = 0,68) i sammenligningsgruppen, et fald på 0,7 (p-værdi = 0,49) i Robusthedsprogrammet og et fald på 0,3 i Mentaliseringsguiden (p-værdi = 0,63).

udviklingen i de to programmer med udviklingen i sammenligningsgruppen, er der ingen statistisk forskel mellem grupperne, hvilket vil sige, at vi ikke kan se forskel på hældningerne af kurverne, når vi tager højde for den statistiske usikkerhed.[28]

Figur 2.8 Udvikling i daglig funktion opdelt på de tre indsatsgrupper

Appendix tabel 1 viser de statistiske beregninger

[28] Robusthedsanalysen finder tilsvarende resultater. I statistiske modeller, som forsøger at tage højde for frafaldet (multipel imputation modeller) eller forskelle mellem grupperne ved baseline (fixed-effects modeller), finder vi den samme forholdsvise udvikling mellem sammenligningsgruppen og de to programmer.

3 Sammenfatning og diskussion

Resultatdokumentationen undersøger udviklingen i døgnanbragte unges mentaliseringsevne og socioemotionelle kompetencer i løbet af projektperioden fra medio 2018 til medio 2020, hvor medarbejderne på anbringelsesstedet tilbydes undervisning i en mentaliseringsbaseret indsats. Dataindsamlingen er gennemført som en baseline-, under- og eftermåling, hvor vi har indsamlet tre mentaliseringstest (LEAS, RME og test med Ekmans Billeder) og et spørgeskema om de unges generelle socioemotionelle kompetencer og vanskeligheder (SDQ). Resultaterne undersøger således de unges udvikling fra programmernes igangsættelse på anbringelsesstederne i 2018, til den opfølgende måling ved projektets afslutning i 2020, og sammenligner udviklingen på tværs af de to mentaliseringsbaserede programmer (Mentaliseringsguiden, MG, og Robusthedsprogrammet, RP) og en sammenligningsgruppe, SG.

Rapporten finder **primært to tendenser i data:** Vi ser dels nogle positive tendenser i de unges evne til at mentalisere andres følelser ud fra ansigtsudtryk (se RME-testen), dels nogle positive tendenser i den generelle socioemotionelle udvikling (se totalscoren for SDQ) for alle unge i evalueringen. Derudover finder vi kun få signifikante ændringer i de unges evne til at mentalisere indre følelser hos sig selv og andre (se LEAS og testen med Ekmans billeder) og kun få signifikante forskelle mellem de tre indsatsgrupper, og disse forskelle mellem grupperne er usikre og skal læses med forbehold for en række begrænsninger i data (se begrænsninger senere i dette afsnit).

For unge i alle tre grupper (MG, RP og SG) ser vi positive tendenser i mentalisering af ydre karakteristika, målt med RME. RME-scoren måler evnen til korrekt at genkende følelser i øjenomgivelser (Malberg & Rosenbery, 2017; Luyten, Lemma & Target, 2019). Ved baselinemålingen lå RME-scoren i gennemsnit på 19,0, svarende til 68 pct. korrekt genkendte følelser i andres øjne. Ved den afsluttende måling lå scoren i gennemsnit 1,1 point højere, hvilket svarer til, at de unge kan genkende den korrekte følelse i 72 pct. af tilfældene.[29] Blandt en nogenlunde jævnaldrende målgruppe (402 almindelige skoleelever med en gennemsnitsalder på 16,7 år) fandt Hünefeldt et al. (2013) en korrekt genkendelse af 66 pct. af billederne i RME-testen, blandt en lidt yngre målgruppe (129 børn med en gennemsnitsalder på 11,7 år) fandt Koizumi & Takagishi (2014) en korrekt genkendelse af 65 pct. af billederne i RME-testen, og blandt en lidt ældre gruppe af drenge med adfærdsproblemer i Danmark fandt Gillespie et al. (2019) en korrekt genkendelse af 65 pct. af billederne i RME-testen. For voksne i normalbefolkningen fandt Baron-Cohen et al. (2001) en korrekt genkendelse af 73 pct. af billederne. De unge i undersøgelsen lå derfor under niveauet for voksne, men på niveau med andre unge ved baselinemålingen, men på niveau med voksne og over niveauet for andre unge ved den opfølgende måling.[30]

Resultaterne fra både LEAS- og RME-testen viser, at de unge er gode til at mentalisere andres ydre karakteristika, men deres evne til at mentalisere indre karakteristika hos dem selv og andre er derimod relativt ringe i forhold til unge på samme alder. De har eksempelvis sværere ved at sætte ord på egne og andres følelser end andre unge har (målt med LEAS), men de er lige så gode eller bedre end andre unge til at genkende følelser i øjenomgivelser (målt med RME). Dette benævnes ofte ”borderline empati paradokset” (Bączkowski & Cierpiałkowska 2015; Fertuck et al., 2009; Fonagy & Bateman, 2019; Rosenblau et al., 2015; Scott et al., 2011).

[29] En lineær regression, som estimerer udviklingen for hele ungegruppen samlet (dvs. uden en interaktionseffekt med indsatsgruppen), har en hældning på 1,1 (ES = 0,32; p-værdi = 0,01).

[30] I en lineær regression, som estimerer udviklingen for hele ungegruppen samlet (dvs. uden en interaktionseffekt med indsatsgruppen), finder vi, at genkendelsen ved den opfølgende måling er signifikant forskellig fra 66 pct. genkendelse (p-værdi < 0,01), men ikke signifikant forskellig fra 73 pct. genkendelse (p-værdi = 0,50).

En mulig forklaring på dette paradoks kan være, at omsorgssvigtede børn, tidligt I deres opvækst, har været hyperopmærksomme på en ustabil omsorgsgivers sindstilstand, men ikke fået hjælp til at forstå denne tilstand og afkode og regulere egne følelser (Bateman & Fonagy, 2019; Gunderson & Lyons-Ruth, 2008). Paradokset mellem mentalisering af egne følelser og andres ansigtsudryk er uændret henover projektperioden.

Der findes kun ganske få publicerede studier af mentaliseringsbaserede programmer til teenagere, som også undersøger udviklingen i de unges mentaliseringsevne (Byrne, Murphy & Connon, 2020). Tre studier (Rossouw & Fonagy, 2012; Bo et al., 2017, 2019) finder dog små til mellemstore effekter af mentaliseringsbaserede interventioner på unges mentaliseringsevne. Til forskel fra dette projekt undersøger de tre studier intensive individuelle eller gruppeforløb til unge med borderline eller undvigende personlighedsforstyrrelse. Et større dansk randomiseret, kontrolleret forsøg med mentaliseringsbaseret træning i grupper til unge borderlinepatienter finder derimod ingen effekt af den menatilseringsbaserede træning i forhold til standard individuel behandlingen i psykiatrien (Beck et al., 2020). Begge interventioner bestod af intensiv træning og støtte leveret af uddannede psykologer eller psykiatere over et år. Til sammenligning har de to programmer i dette studie kun få gruppesessioner med de unge og ingen individuelle sessioner. Det kan derfor være nødvendigt at øge antallet af gruppesessioner med de unge for at opnå en tydeligere effekt på mentaliseringsevnen.

Den sidste tendens i data er en positiv udvikling i generelle socioemotionelle kompetencer. For unge i alle tre grupper ser vi et samlet fald i socioemotionelle vanskeligheder, målt med SDQ.[31] Faldet er tydeligst i Mentaliseringsguiden, hvor niveauet af socioemotionelle vanskeligheder i gennemsnit falder med 1,6 point over 2 år fra 15,2 point ved baselinemålingen. SDQ-scoren går således i gennemsnit fra kategorien 'let forhøjede vanskeligheder' til kategorien 'ingen vanskeligheder' (Arnfred et al., 2019).[32] SDQ-scoren indfanger generelle socioemotionelle vanskeligheder; resultaterne peger således i retning af, at programmet i gennemsnit har styrket de unges generelle trivsel og normaliseret socioemotionelle vanskeligheder. Størrelsen af ændringen svarer til, hvad andre studier af mentaliseringsbaserede programmer til anbragte børn har fundet på SDQ (Midgley et al., 2019a, 2019b).

Der kan være mindst **tre hypoteser eller forklaringer på, at vi ikke observerer en tydeligere forbedring i de unges evne til at mentalisere** egne og andres følelser henover projektperioden. En hypotese er, at mentaliseringsevnen er en kompleks evne, som er svær at indfange og ændre. Opdraget for evalueringen var bl.a. at undersøge de unges mentaliseringsevne og socioemotionelle kompetencer, samtidig med at dataindsamlingen var så nem som muligt for medarbejdere og unge på anbringelsesstederne. Vi har efter bedste evne forsøgt at indsamle data om mentaliseringsevnen med et let tilgængeligt testbatteri, som tidligere har vist lovende resultater (Sørensen, 2013). Men målet om at undersøge udviklingen i de unges mentaliseringsevne kan have været for ambitiøst i forhold til stikprøvestørrelsen og rammen for evalueringen. Mentaliseringsevnen medierer endvidere trivsel, angst, depression og andre sociale og emotionelle kompetencer (Scavenius et al., 2019). I denne evaluering har vi ikke haft mulighed for at udvide testbatteriet med flere test, men fremadrettet kunne man også sandsynliggøre resultater af programmerne på udviklingen af andre kompetencer.

[31] En lineær regression, som estimerer udviklingen for hele ungegruppen samlet (dvs. uden en interaktionseffekt med indsatsgruppen), har en hældning på -1,4 (ES = 0,26; p-værdi = 0,04).

[32] En score under 14 betegnes af Arnfred et al. (2019) som inden for normalområdet. I en lineær regression, som estimerer udviklingen med en interaktionseffekt for hver indsatsgruppe (se Bilagstabel 1.6), estimerer vi, at unge i Mentaliseringsguiden i gennemsnit ender på 13,8 point (95 CI = [12,2;15,3]). Det vil sige, at i gennemsnit ligger de inden for normalområdet, men gennemsnittet er ikke signifikant forskellig fra en score på 14 point (p-værdi = 0,75), så en stor del af de unge ligger stadig uden for normalområdet.

En anden hypotese er, at det tager tid for personalet at implementere og dygtiggøre sig i programmet. Medarbejderne på anbringelsesstederne, specielt medarbejdere, som har modtaget undervisning i Mentaliseringsguiden, fortæller eksempelvis, at de først efter halvandet år er begyndt at opbygge en rutine i metoderne, og den manglende rutine kan være medvirkende til, at de unge endnu ikke har fået fuldt udbytte af programmet. Tilsvarende finder vi i projektets implementeringsrapport (Hansen et al., 2020), at anbringelsesstederne stadig arbejder med implementeringsprocessen, og ved projektets afslutning har de udviklet medarbejderkompetencer og tilpasset organisation og ledelse i forskelligt omfang. En gennemgribende implementering tager tid, og 4-5 anbringelsessteder har kun implementeret programmerne i lav grad. Projektets fidelitetsevaluering (Kaspersen et al., 2020) konkluderer ligeledes, at der i de formelt strukturerede aktiviteter på anbringelsesstederne, såsom behandlingskonferencer og planlagte samtaler med de unge, er mulighed for bedre at udnytte programmernes øvelser og begreber. Det vil sige, at programmerne stadig kan anvendes mere systematisk og med højere fidelitet. Den delvise implementering og reducerede fidelitet kan således forklare, at vi endnu ikke har set en stigning i mentaliseringsevnen hos de unge.

En tredje hypotese kan findes i niveauet på LEAS-scoren. I dette studie ligger de unges mentaliseringsscore i gennemsnit mellem niveau 2 og 3, hvilket er omtrent et point lavere end i en normalpopulation (se referencer i Scavenius et al., 2019).[33] Unge med et udviklingsniveau omkring 2 eller derunder har tegn på *"signifikante problemer med basale mentale funktioner"* (Speranza & Midgley, 2017, s. 307) og en usikker tilknytning til omsorgspersoner (McWilliams & Schedler, 2017; Speranza & Mayes, 2017). Fra et udviklingspsykologisk perspektiv har de unge på niveau 2 en emotionsbevidsthed svarende til et handleniveau for emotioner, hvor følelser udtrykkes ved handlinger – fx "*jeg ville føle, at jeg skulle smadre noget"* – eller følelserne afkodes som overordnede, globale fornemmelser – fx "*jeg ville føle, at det var skidt"*. I gennemsnit har de unge i undersøgelsen således meget svært ved at forestille sig og beskrive følelser hos sig selv eller andre, fx "*jeg ville føle vrede"* (svarende til LEAS-niveau 3), og de kan derfor også have svært ved at tilegne sig nye mentaliseringsevner alene gennem sprogbårne interventioner. I projektets fidelitetsrapport (Kaspersen et al., 2020) beskriver vi tilsvarende, at anbringelsesstederne tilpasser øvelserne til de unges funktions- og kognitive niveau. For de børn, der primært udtrykker følelser ved handlinger, kunne flere mentaliseringsbaserede handlinger eller øvelser eventuelt styrke den kropslige mentalisering og forbedre udbyttet af mentaliseringsbaserede samtaler (Fonagy et al., 2017; Fonagy & Target, 1996; Fotopoulou & Tsakiris, 2017; Halfon & Bulut, 2019; Shai & Belsky, 2011; Tessier et al., 2016; Verheugt-Pleiter, Zevalkink & Schmeets, 2008). Det vil sige, at programmerne med fordel kunne udvides med flere kropslige øvelser for at nå børn og unge med et meget lavt, sprogligt mentaliseringsniveau.

Resultaterne læses med forbehold for mindst syv **begrænsninger. For det første** er tildelingen af undervisning i programmerne ikke randomiseret; anbringelsesstederne har selv søgt om mulighed for kompetenceudvikling. Der kan derfor være systematiske forskelle mellem anbringelsessteder i de tre grupper, som også har påvirket de unges mentaliseringsevne og socioemotionelle kompetencer henover projektperioden. Med det ikke-randomiserede evalueringsdesign kan vi ikke isolere effekten af programmerne fra andre systematiske forskelle mellem grupperne. Resultaterne beskriver derfor kun tendenser i de unges udvikling i de to programmer i forhold til sammenligningsgruppen og genererer hypoteser om disse fund, som man bør undersøge nærmere.

[33] For normalpopulationen ligger LEAS total i gennemsnit mellem 3,0 og 3,6. I en lineær regression, som estimerer udviklingen for hele ungegruppen samlet (dvs. uden en interaktionseffekt med indsatsgruppen), finder vi en samlet LEAS score på 2,6 point (95 pct. CI = [2,5;2,7]) ved den opfølgende måling, hvilker er signifikant forskellig fra en score på 3,0 point (p-værdi < 0,01).

For det andet er et betydeligt antal unge faldet fra undersøgelsen. Ved baselinemålingen indsamlede vi test fra alle 102 unge. Ved midtvejsmålingen indsamlede vi kun test fra de unge i Robusthedsprogrammet, her lykkeses det at teste 14 unge. Ved eftermålingen indsamlede vi igen test fra alle, men her lykkedes det kun at teste 58 unge (6 unge i RP-gruppen, 42 unge i MG-gruppen og 10 unge i SG). I alt bygger analysen således på 174 observationer af døgnanbragte unge. Frafaldet skyldes oftest ugyldige kontaktoplysninger på de unge, som er fraflyttet anbringelsesstedet, eller at de unge ikke ønskede at deltage i den opfølgende måling. Frafaldsprocenten fra 2018 til 2020 er 43 pct. (N = 44). Frafaldsanalysen peger på, at de unge, som faldt fra undersøgelsen, har lidt færre anbringelser i bagagen og er ældre, ellers er der ingen overrepræsentation i forhold til køn, alder ved første anbringelse eller længden af anbringelsen ved anbringelsesstedet. Generelt kan vi ikke afgøre, om frafaldet har påvirket resultaterne; det afhænger af, hvilke unge som er faldet fra – om de unge, der faldt fra, ville have scoret højt eller lavt på testene ved den opfølgende måling og således ville have trukket resultaterne i en anden retning. Specielt for sammenligningsgruppen og Robusthedsprogrammet er resultaterne meget usikre på grund af den lille stikprøve, hvilket kan skævvride resultaterne og mindske den statistiske styrke.[34]

For det tredje undersøger evalueringen en mindre gruppe af unge på anbringelsesstederne. De unge i undersøgelsen var i 2018 anbragt på 21 anbringelsessteder, heraf 3 anbringelssteder, som implementerer Robusthedsprogrammet, 11 anbringelssteder, som implementerer Mentaliseringsguiden, og 7 sammenlignelige anbringelssteder, som ikke implementerer et af de to mentaliseringsbaserede programmer. Ikke alle unge fra de 21 anbringelssteder har deltaget i evalueringen (Scavenius & Jonsen, 2020). Vi bad anbringelsesstederne udvælge unge, som ved baselinemålingen var mellem 12 og 23 år, ikke havde diagnosen infantil autisme og med stor sandsynlighed ville være anbragt hos dem i hele projektperioden. Derudover blev de unge spurgt, om de ville deltage i dataindsamlingen. De unge i evalueringen er således ikke repræsentative for alle unge på anbringelsesstederne, og de tendenser, vi ser i data, afspejler kun de unge, som ønskede at deltage i undersøgelsen og falder inden for inklusionskriterierne.

For det fjerde har 43 pct. af de unge forladt anbringelsesstedet, på trods af at vi har forsøgt at indsamle data blandt unge, som med stor sandsynlighed ville være på anbringelsesstedet i hele projektperioden. For unge i Robusthedsprogrammet er 42 pct. stadig på anbringelsesstedet, for unge i Mentaliseringsguiden er 60 pct. stadig på anbringelsesstedet, og for unge i sammenligningsgruppen er 63 pct. stadig på anbringelsesstedet. Ophør i anbringelsen har naturligvis konsekvenser for, hvor længe de unge er i kontakt med pædagoger og andet personale, der er uddannet i Robusthedsprogrammet eller Mentaliseringsguiden, og dermed også konsekvenser for de unges udbytte af programmerne. Tilsvarende beskriver Hansen et al. (2020), at implementeringen på to af de tre steder, som implementerer Robusthedsprogrammet, har haltet, hvilket også kan have påvirket de unges udbytte af programmet. Den flade udvikling i eksempelvis LEAS-scoren kan således skyldes, at en stor del af de unge har modtaget en relativ lav 'dosis' af programmet.

For det femte kan baggrundskarakteristika have betydning for de unges udbytte af programmerne. I forhold til de unges mentaliseringsniveau ved baselinemålingen er der nogle forskelle mellem sammenligningsgruppen, Robusthedsprogrammet og Mentaliseringsguiden, men forskellene er ikke entydige. Ved baselinemålingen havde de unge i sammenligningsgruppen et lidt højere mentaliseringsniveau end de unge i Robusthedsprogrammet og Mentaliseringsguiden, målt ved LEAS-testen. Målt med RME-testen, testen med Ekmans billeder og SDQ-

[34] Robusthedsanalyserne, hvor manglende observationer imputeres med multiple imputation, tyder på, at resultaterne er stabile over for frafaldet.

målingerne var der derimod ingen statistisk forskel på grupperne ved baseline. De unge i sammenligningsgruppen og Mentaliseringsguiden er derimod yngre end unge i Robusthedsprogrammet, se Scavenius et al. (2019). Vi har gennemført robusthedsanalyser, som kontrollerer for disse forskelle ved baseline. Resultaterne af robusthedsanalyserne er sammenlignelige med resultaterne i resultatafsnittet, hvilket tyder på, at baggrundskarakteristika ikke har påvirket udbyttet af programmerne i vores stikprøve.

For det sjette kan programmerne have påvirket medarbejdernes mentaliseringsevne, men dette har endnu ikke haft en afsmittende effekt på de unges mentaliseringsevne. Som beskrevet i de to programmer (Bak, 2017; Hagelquist, 2015; Scavenius et al., 2020; Kaspersen et al., 2020) går forandringsmekanismen primært gennem medarbejdernes mentalisering af de unge. Kun gennem medarbejdernes egen mentaliseringsevne kan medarbejderne styrke de unges mentaliseringsevne. I denne resultatdokumentation har vi udelukkende undersøgt andet led i forandringsmekanismen. Ved opstarten af projekt vurderede vi, at det ville være umuligt at teste et tilstrækkeligt antal medarbejdere på anbringelsesstederne, dels på grund af logistiske udfordringer, men også fordi der var modstand i dele af personalegruppen mod at blive testet. I anbringelsesstedernes egen resultatdokumentation eller i fremtidige analyser af mentaliseringsbaserede programmer vil det derfor være relevant at supplere testene af børnene og de unges mentaliseringsevne med test af medarbejdernes mentaliseringsevne for at få et mere komplet billede af det mentaliserende miljø på anbringelsesstederne.

For det syvende kan vi ikke udelukke, at COVID-19 og nedlukningen i 2020 har påvirket den afsluttende måling. På anbringelsesstederne blev alle rutiner lavet om, hvilket har været svært for de unge, som har oplevet en angst og frygt for den usikre situation. På den anden side fortæller et par anbringelsessteder også, at nedlukningen har betydet, at de unge ikke skulle i skole, og personalet derfor har haft mulighed for stabilt at arbejde med og mentalisere de unge over en længere periode. COVID-19 og nedlukningen kan således have påvirket anbringelsesstederne forskelligt og dermed påvirket resultaterne.

Resultaterne af de to programmer på de unges udvikling akkompagneres af en omkostningsvurdering af de to programmer (Scavenius & Jonsen, 2020). Overordnet peger resultaterne af omkostningsvurderingen på, at Mentaliseringsguiden er 5 gange dyrere, fordi programmet er 5 gange længere end Robusthedsprogrammet. Vi vil gerne advare mod, at man bruger resultaterne af Resultat- og Omkostningsvurderingen til at konkludere, at de to programmer har samme effekt, og man derfor lige så godt kan iværksætte det billigste program. Det ville være en fejlagtig konklusion på baggrund af de indsamlede data. Som beskrevet ovenfor er evalueringen ikke gennemført som et randomiseret kontrolleret forsøg, der var et betydeligt frafald fra dataindsamlingen, og få unge er med i sammenligningsgruppen og Robusthedsprogrammet. Datamaterialet er derfor ikke stærkt nok til at sammenligne resultaterne fra Robusthedsprogrammet med resultaterne fra Mentaliseringsguiden. Begge programmer viser derimod positive tendenser i forhold til de unges mentaliseringsevne og sociale kompetencer. Derudover er de to programmer markant forskellige, hvad angår længde og operationalisering af mentaliseringsteorien, så valget af program bør nærmere bero på, om den eksisterende pædagogiske praksis på anbringelssteder er kompatibel med det ene eller det andet program, som beskrevet i projektets implementeringsevaluering (Hansen et al., 2020).

Resultaterne af **de unges udvikling står i slående kontrast til medarbejdernes vurdering** af programmerne og oplevelser med de unge. Både Kaspersen et al. (2020) og Hansen et al. (2020), samt anekdoter og casebeskrivelser fra medarbejderne, illustrerer, at medarbejderne har en anden oplevelse af programmerne. De fortæller eksempelvis, at programmerne har

medført et fælles sprog, de er blevet mere mentaliserende over for de unge, og de unge responderer positivt på den nye tilgang. Forskellene kan skyldes, at de kvantitative data undersøger specifikke evner hos de unge, hvorimod de kvalitative fortællinger inkluderer et helhedsbillede af de unge. På den anden side inkluderer de kvantitative dataindsamlinger information fra en bredere gruppe af unge, hvorimod casebeskrivelser kun fortæller om specifikke oplevelser med få unge. Den ene type data er ikke generelt bedre end den anden; de supplerer hinanden. I dette projekt har vi ikke haft mulighed for systematisk at supplere den kvantitative dataindsamling med interview eller observationer af de unge. Men de positive tendenser, som vi har observeret i de kvantitative dataindsamlinger og medarbejdernes fortællinger, illustrerer et behov for, at dokumentationen af resultater på anbringelsessteder fremadrettet gennemføres med både kvantitative og kvalitative data.

Litteratur

Arnfred, J., Svendsen, K., Rask, C., Jeppesen, P., Fensbo, L., Houmann, T., ... & Bilenberg, N. (2019). Danish norms for the Strengths and Difficulties Questionnaire. *Danish Medical Journal*, *66*(6), 1-7.

Bączkowski, B. M. & Cierpiałkowska, L. (2015). Mentalization within close relationships: The role of specific attachment style. *Polish Psychological Bulletin, 46*(2), 285-299.

Bak, P.L. (2017). *Robusthed - en redskabsbog for professionelle og forældre*. Aarhus: Forlaget Klim.

Baron-Cohen, S., Wheelwright, S., Lawson, J., Griffin, R., Ashwin, C., Billington, J. & Chakrabarti, B. (2005). Empathizing and systemizing in autism spectrum conditions. In: Volkmar, F. R., Paul, R., Klin, A. & Cohen, D. (Eds.), *Handbook of Autism and Pervasive Developmental Disorders.* New York: John Wiley & Sons, Ltd.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y. & Plumb I. (2001). The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry, 42*(2), 241-251.

Bateman, A. & Fonagy, P. (2016). *Mentalization-Based Treatment for Personality Disorders*. London: Oxford Press.

Bateman, A. & Fonagy, P. (2019). *Handbook of Mentalization in Mental Health Practice*. Washington, DC: American Psychiatric Publishing, Inc.

Beck, E., Bo, S., Jørgensen, M. S., Gondan, M., Poulsen, S., Storebø, O. J., ... & Simonsen, E. (2020). Mentalization-based treatment in groups for adolescents with borderline personality disorder: a randomized controlled trial. *Journal of Child Psychology and Psychiatry*, *61*(5), 594-604.

Bo, S., Sharp, C., Beck, E., Pedersen, J., Gondan, M. & Simonsen, E. (2017). First empirical evaluation of outcomes for mentalization-based group therapy for adolescents with BPD. *Personality Disorders: Theory, Research, and Treatment*, *8*(4), 396.

Bo, S., Bateman, A. & Kongerslev, M. T. (2019). Mentalization-based group therapy for adolescents with avoidant personality disorder: Adaptations and findings from a practice-based pilot evaluation. *Journal of Infant, Child, and Adolescent Psychotherapy*, *18*(3), 249-262.

Byrne, G., Murphy, S., & Connon, G. (2020). Mentalization-based treatments with children and families: A systematic review of the literature. *Clinical Child Psychology and Psychiatry*, *25*(4) 1022–1048.

Cohen, J. (1988). *Statistical Power for the Behavioral Sciences.* Hillsdale, NJ: Erlbaum.

Eifenbein, H. A. & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin, 128*(2), 203-235.

Ekman, P. (2003). *Emotions Revealed. Understanding Faces and Feelings.* London: Weidenfeld & Nicolson.

Fertuck, E. A., Jekal, A., Song, I., Wyman, B., Morris, M. C., Wilson, S. T., Brodsky, B. S. & Stanley, B. (2009). Enhanced 'Reading the Mind in the Eyes' in borderline personality disorder compared to healthy controls. *Psychological Medicine, 39*(12), 1979-1988.

Fonagy, P., Luyten, P., Allison, E. & Campbell, C. (2017). What we have changed our minds about: Part 1. Borderline personality disorder as a limitation of resilience. *Borderline Personality Disorder and Emotion Dysregulation, 4*(1), 11.

Fonagy, P. & Target, M. (1996). Playing with reality: I. Theory of mind and the normal development of psychic reality. *International Journal of Psycho-analysis*, *77*, 217-233.

Fotopoulou, A. & Tsakiris, M. (2017). Mentalizing homeostasis: The social origins of interoceptive inference. *Neuropsychoanalysis*, *19*(1), 3-28.

Gillespie, S. M., Kongerslev, M. T., Bo, S. & Abu-Akel, A. M. (2020). Schizotypy and psychopathic tendencies interactively improve misattribution of affect in boys with conduct problems. *European Child & Adolescent Psychiatry*, Publ. online 31 May 2020, 1-13.

Goodman, G. (2002). *The Internal World of Attachment*. Hillsdale, NJ: The Analytic Press.

Griffiths, H., Noble, A., Duffy, F., & Schwannauer, M. (2017). Innovations in Practice: Evaluating clinical outcome and service utilization in an AMBIT-trained Tier 4 child and adolescent mental health service. *Child and Adolescent Mental Health*, *22*(3), 170-174.

Gunderson, J. G. & Lyons-Ruth, K. (2008). BPD's interpersonal hypersensitivity phenotype: A gene-environment-developmental model. *Journal of Personality Disorders*, *22*(1), 22-41.

Gutmann, L. (1944). A basis for scaling qualitative data. *American Sociological Review, 9*(2), 139-150.

Hagelquist, J. (2015). *Mentaliseringsguiden*. København: Hans Reitzels Forlag.

Halfon, S. & Bulut, P. (2019). Mentalization and the growth of symbolic play and affect regulation in psychodynamic therapy for children with behavioral problems. *Psychotherapy Research*, *29*(5), 666-678.

Hansen, S. J., Kaspersen, S., Scavenius, C., Pöckel, K. M. & Sørensen, J. H. (2020). *Erfaringer med implementering af to mentaliseringsbaserede programmer – Afsluttende evaluering af 'Styrkelse af børn og unges mentaliseringsevne, handlekraft og sociale færdigheder*. København: VIVE – Det Nationale Forsknings- og Analysecenter for Velfærd.

Halfon, S. & Bulut, P. (2019). Mentalization and the growth of symbolic play and affect regulation in psychodynamic therapy for children with behavioral problems. *Psychotherapy Research, 29*(5), 666-678.

Hauber, K., Boon, A. E. & Vermeiren, R. (2017). Examining changes in personality disorder and symptomology in an adolescent sample receiving intensive mentalization based treatment: a pilot study. *Child and Adolescent Psychiatry and Mental Health*, *11*(1), 1-7.

Hünefeldt, T., Laghi, F., Ortu, F. & Belardinelli, M.O. (2013). The relationship between 'theory of mind' and attachment-related anxiety and avoidance in Italian adolescents. *Journal of Adolescents, 36*(3), 613-621.

Kaspersen, S., Hansen, S. J., Scavenius, C., Pöckel, K. M. & Sørensen, J. H. (2020). *Fidelitetsvurdering af to mentaliseringsbaserede programmer på anbringelsessteder – Afsluttende evaluering af 'Styrkelse af børn og unges mentaliseringsevne, handlekraft og sociale færdigheder.* København: VIVE – Det Nationale Forsknings- og Analysecenter for Velfærd.

Koizumi, M. & Takagishi, H. (2014). The Relationship between Child Maltreatment and Emotion Recognition. *PLoS One, 9*(1), e86093, 1-4.

Lane, R. D. & Garfield, D. A. S. (2005). Becoming aware of feelings: Integration of cognitive development, neuroscience, and psychoanalytic perspectives. *Neuro-Psychoanalysis, 7*(1), 5-30.

Lane, R. D., Quinlan, D. M., Schwartz, G. E., Walker, P. A. & Zeitlin, S. B. (1990). The Levels of Emotional Awareness Scale: A cognitive-developmental measure of emotion. *Journal of Personality Assessment*, *55*(1-2), 124-134.

Lane, R. D., Ryan, L., Nadel, L. & Greenberg, L. (2015). Memory reconsolidation, emotional arousal, and the process of change in psychotherapy: new insights from brain science. *Behavioral and Brain Sciences, 38,* 1-64.

Lane, R. D. & Schwartz, G. E. (1987). Level of emotional awareness: A cognitive- developmental theory and its application to psycho-pathology. *American Journal of Psychiatry, 144,* 133-143.

Luyten, P., Fonagy, P., Lowyck, B. & Vermote, R. (2012). Assessment of Mentalization. In: Bateman, A. W. & Fonagy, P. (Eds.), *Handbook of Mentalizing in Mental Health Practice.* Washington, DC: American Psychiatric Publishing, Inc.

Luyten, P., Lemma, A. & Target, M. (2019). Depression. In: Bateman, A. & Fonagy, P. (Eds), *Handbook of Mentalizing in Mental Health Practice*. Washington, DC. American Psychiatric Publishing.

Malberg, N. & Rosenberg, L. (2017). Profile of Mental Functioning for Children. In: Lingiardi, V. & McWilliams, N. (Eds.), *The Psychodynamic Diagnostic Manual (PDM-2).* New York: Guilford Press.

McWilliams, N. & Shedler, J. (2017). Personality syndromes - P axis. In: V. Lingiardi & McWilliams, N. (Eds.), *Psychodynamic Diagnostic Manual (PDM-2).* New York: Guilford Press.

Midgley, N., Besser, S. J., Fearon, P., Wyatt, S., Byford, S., & Wellsted, D. (2019a). The Herts and Minds study: feasibility of a randomised controlled trial of Mentalization-Based Treatment versus usual care to support the wellbeing of children in foster care. *BMC Psychiatry*, *19*(1), 215.

Midgley, N., Cirasola, A., Austerberry, C., Ranzato, E., West, G., Martin, P., ... & Park, T. (2019b). Supporting foster carers to meet the needs of looked after children: A feasibility

and pilot evaluation of the Reflective Fostering Programme. *Developmental Child Welfare, 1*(1), 41-60.

Niedenthal, P. M. (2007). Embodying emotions. *Science, 316*(5827), 1002-1005.

Panksepp, J. & Biven, L. (2012). *The Archeology of Mind*. New York, NY: Norton.

Panksepp, J. & Biven, L. (2012a). A meditation on the affective neuroscientific view of human and animalian MindBrains. In: Fotopoulou, A., Pfaff, D. & Conway, M. A. (Eds), *From the Couch to the Lab. Trends in Psychodynamic Neuroscience.* Oxford: Oxford University Press.

Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford: Oxford University Press.

Plutchik, R. (2003). *Emotions and Life. Perspectives from Psychology, Biology and Evolution*. Washington DC: American Psychological Association.

Rasch, G. (1960). *Probabilistic Model for some Intelligence and Achievement Tests*. København: Danish Institute for Educational Research.

Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. 2. ed. Thousand Oaks, CA: Sage.

Rosenblau, G., Kliemann, D., Heekeren, H.K., Dziobek, I. (2015). Approximating Implicit and Explicit Mentalizing with Two Naturalistic Video-Based Tasks in Typical Development and Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders, 45*(4), 953–965.

Rossouw, T. I. & Fonagy, P. (2012). Mentalization-based treatment for self-harm in adolescents: A randomized controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry, 51*(12), 1304-1313.

Scavenius, C., Kaspersen, S., Hansen, S. J. & Sørensen, J. H. (2019*). Opstart af to mentaliseringsbaserede programmer. Midtvejsevaluering af 'Styrelse af bøn og unges mentaliseringsevne, handlekraft og sociale færdigheder'.* København: VIVE – Det Nationale Forsknings- og Analysecenter for Velfærd.

Scavenius, C. & Jonsen, E. H. (2020). *Omkostningsvurdering af to mentaliseringsbaserede programmer.* København: VIVE – Det Nationale Forsknings- og Analysecenter for Velfærd.

Scavenius, C., Sørensen, J. H., Hansen, S. J. & Kaspersen, S. (2020). *Resultatdokumentation af to mentaliseringsbaserede programmer. Afsluttende evaluering af 'Styrkelse af børn og unges mentaliseringsevne, handlekraft og sociale færdigheder'*. København: VIVE – Det Nationale Forsknings- og Analysecenter for Velfærd.

Scott, L. N., Levy, K. N., Adams, R. B. Jr. & Stevenson, M. T. (2011). Mental state decoding abilities in young adults with borderline personality disorder traits. *Personality Disorders: Theory, Research, and Treatment*, *2*(2), 98-112.

Servicestyrelsen (2007). *Håndbog om anbringelsesreformen.* 2. udg. Odense: Servicestyrelsen.

Shai, D. & Belsky, J. (2011). When words just won't do: Introducing parental embodied mentalizing. *Child Development Perspectives*, *5*(3), 173-180.

Socialstyrelsen (2016). *Resultatdokumentation og evaluering. Håndbog for sociale tilbud.* Hentet fra: https://socialstyrelsen.dk/udgivelser/resultatdokumentation-og-evaluering-2013-handbog-for-sociale-tilbud-2 [8. 12. 2020].

Socialstyrelsen (2020). *Håndbog I lovende praksis. Vidensbaseret udvikling af sociale indsatser.* Hentet fra: https://socialstyrelsen.dk/udgivelser/handbog-i-lovende-praksis-haefte-0-vidensbaseret-udvikling-af-sociale-indsatser [8. 12. 2020].

Speranza, A. M. & Mayes, L. (2017). Mental Health and Developmental Disorders. In: Lingiardi, V. & McWilliams, N. (Eds.), *Psychodynamic Diagnostic Manual (PDM-2).* New York: The Guilford Press.

Speranza, A. M. & Midgley, N., (2017). Profile of Mental Functioning for Adolescents. In: Lingiardi, V. & McWilliams, N. (Eds.), *Psychodynamic Diagnostic Manual (PDM-2).* New York. The Guilford Press.

Sørensen, J. H. (2009). *Mentaliseringsbaseret behandling i teori og praksis.* København. Hans Reitzels Forlag.

Sørensen, J. H. (2012). Mentalisering af krop, handling og emotion: implicit og eksplicit mentalisering i differentieret neuroaffektiv behandling. In: Hart, S. (Ed.), *Neuroaffektiv Psykoterapi med voksne.* København. Hans Reitzels Forlag.

Sørensen, J. H. (2013). *Programmer for PTSD. Region Syddanmarks tværfaglige udrednings- og behandlingsprogrammer for personer med Posttraumatic Stress Disorder (PTSD), Kompleks PTSD og Komorbide diagnoser*. Vejle: Afdeling for Traume- og Torturoverlevere, Psykiatrien i Region Syddanmark.

Tessier, V. P., Normandin, L., Ensink, K. & Fonagy, P. (2016). Fact or fiction? A longitudinal study of play and the development of reflective functioning. *Bulletin of the Menninger clinic*, *80*(1), 60-79.

Verheugt-Pleiter, A. J., Zevalkink, J. E. & Schmeets, M. G. (2008). *Mentalizing in Child Therapy: Guidelines for clinical practitioners.* London: Karnac Books.

Bilag 1 Appendiks

Appendikset beskriver først den statistiske metode og derefter de statistiske beregninger, som ligger til grund for resultatafsnittet i rapporten. Appendikset kræver kendskab til estimation af regressionsmodeller og er skrevet til læseren, som ønsker at forstå alle detaljer i den statistiske model.

Den **statistiske metode** anvender et quasi-eksperimentelt, før-efter design med multiple respondenter i indsatsgrupper, hvor gruppernes udvikling over tid og forskellen mellem de quasi-eksperimentielle grupper approksimerer effekten af indsatserne. Mere specifikt estimerer vi følgende hierarkiske lineære regressionsmodel: $Y_{tik} = \gamma_0 + \gamma_1 RP_i + \gamma_2 MG_i + \delta_0\, tid_t + \delta_1(tid_t \times RP_i) + \delta_2(tid_t \times MG_i) + \zeta^3_k + \zeta^2_{ik} + \varepsilon_{tik}$. Se StataCorp (2019) eller Raudenbush & Bryk (2002), for en introduktion til hierarkiske regressionsmodeller. I modellen er outcome, Y_{tik}, forklaret af indsatsgruppen (RP, MG og referencegruppen, SG) og 2-årig kontinuert tid siden 1. marts 2018, $tid_t=(t-t_{01.03.18})/2\times365$. Parameteren ζ^3_k er et random skæringspunkt for anbringelsessted k og ζ^2_{ik} er et random skæringspunkt for individ i på anbringelsessted k. Fejlleddene (ζ^3_k,ζ^2_{ik} og ε_{tik}) antages i modellen at have middelværdi på nul og være individuelt ukorreleret, så deres individuelle varianser summer til den samlede varians, betinget på modellens forklarende variable.

Bilagstabel 1.1 viser fem parameterestimater fra modellen. Parameteren δ_0 estimerer SG's udvikling over 2 år, summen af parametrene $\delta_0 + \delta_1$ estimerer RPs udvikling over 2 år, og summen af parametrene $\delta_0 + \delta_2$ estimerer MGs udvikling over 2 år. Parametrene δ_1 og δ_2 estimerer derfor forskellen mellem SG og henholdsvis RP og MG. For hvert parameterestimat rapporterer vi Cohen's D for at kunne afgøre den relative effektstørrelse i forhold til standardafvigelsen (SD) ved baselinemålingen. Værdier omkring 0,2 tolkes typisk som en lille effekt, værdier omkring 0,5 tolkes typisk som en medium effekt, og værdier omkring 0,8 tolkes typisk som en stor effekt (Cohen, 1988).

Bilagstabel 1.1 Baseline: Antal observationer, gennemsnit og standardafvigelser

	Baseline (N = 102)								
	SG (N = 21)			RP (N = 19)			MG (N = 63)		
Test og spørgeskemaer	N_i	Cl	M(SD)	N_i	Cl	M(SD)	N_i	Cl	M(SD)
LEAS, Selv	21	7	2,62(0,43)	19	3	2,21(0,80)	62	11	2,19(0,59)
LEAS, Andre	21	7	2,28(0,42)	19	3	2,00(0,67)	62	11	1,99(0,59)
LEAS, Total	21	7	2,94(0,37)	19	3	2,63(0,81)	62	11	2,54(0,57)
RME	21	7	19,29(3,81)	19	3	20,58(2,93)	62	11	18,55(3,47)
Test med Ekmans billeder	21	7	1,91(0,97)	19	3	1,82(0,86)	62	11	2,00(0,85)
Samlet problemscore (SDQ)	20	7	15,95(5,04)	19	3	13,84(4,96)	62	11	15,16(5,56)
Følelsesmæssige symptomer	20	7	4,00(2,27)	19	3	3,37(1,95)	62	11	4,27(2,30)
Adfærdsmæssige symptomer	20	7	3,10(2,27)	19	3	2,79(2,04)	62	11	2,63(1,91)
Hyperaktivitet/opmærksomhedsvanskeligheder	20	7	5,95(2,31)	19	3	4,95(2,09)	62	11	5,37(2,28)
Vanskeligheder ift. jævnaldrende	20	7	2,90(1,48)	19	3	2,74(1,48)	62	11	2,89(1,74)
Sociale styrkesider (SDQ)	20	7	7,45(1,88)	19	3	7,58(1,54)	62	11	7,44(2,00)
Trivsel og funktion (SDQ)	20	7	1,45(2,11)	19	3	2,37(2,29)	61	11	1,82(2,43)
Ked af det/ulykkelig	20	7	0,35(0,59)	19	3	0,21(0,54)	61	11	0,31(0,62)
Familie	20	7	0,25(0,55)	19	3	0,74(0,81)	61	11	0,26(0,60)
Venskaber	20	7	0,15(0,49)	19	3	0,42(0,61)	61	11	0,26(0,57)
Indlæring	20	7	0,45(0,60)	19	3	0,42(0,61)	61	11	0,67(0,83)
Fritid	20	7	0,25(0,55)	19	3	0,58(0,84)	61	11	0,31(0,67)

Note: Observationer (N = observationer, N_i = individer, Cl = anbringelsessteder (clusters)); Gennemsnit (M) og std. dev. (SD).

Bilagstabel 1.2 Første opfølgning: Antal observationer, gennemsnit og standardafvigelser

	Første opfølgende test (N = 14)		
	RP (N = 14)		
Test og spørgeskemaer	N_i	Cl	M(SD)
LEAS, Selv	14	3	2,48(0,57)
LEAS, Andre	14	3	2,44(0,58)
LEAS, Total	14	3	2,67(0,55)
RME	14	3	21,21(3,81)
Test med Ekmans billeder	14	3	1,67(0,97)
Samlet problemscore (SDQ)	8	2	11,63(5,24)
Følelsesmæssige symptomer	8	2	2,25(1,58)
Adfærdsmæssige symptomer	8	2	2,38(2,13)
Hyperaktivitet/opmærksomhedsvanskeligheder	8	2	3,63(1,77)
Vanskeligheder ift. jævnaldrende	8	2	3,38(2,00)
Sociale styrkesider (SDQ)	8	2	7,63(2,00)
Trivsel og funktion (SDQ)	8	2	1,00(1,69)
Ked af det/ulykkelig	8	2	0,13(0,35)
Familie	8	2	0,25(0,46)
Venskaber	8	2	0,13(0,35)
Indlæring	8	2	0,50(0,76)
Fritid	8	2	0,00(0,00)

Note: Observationer (N = observationer, N_i = individer, Cl = anbringelsessteder (clusters)); Gennemsnit (M) og std. dev. (SD). SG (N = 0) og MG (S = 0).

Bilagstabel 1.3 Anden opfølgning: Antal observationer, gennemsnit og standardafvigelser

	Anden opfølgende test (N = 58)								
	SG (N = 10)			RP (N = 6)			MG (N = 42)		
Test og spørgeskemaer	N_i	Cl	M(SD)	N_i	Cl	M(SD)	N_i	Cl	M(SD)
LEAS, Selv	10	6	2,20(0,58)	6	2	1,97(0,82)	41	11	2,23(0,59)
LEAS, Andre	10	6	1,96(0,57)	6	2	2,27(0,48)	41	11	2,02(0,56)
LEAS, Total	10	6	2,51(0,46)	6	2	2,62(0,43)	41	11	2,51(0,57)
RME	10	6	20,20(4,13)	6	2	21,00(2,61)	41	11	19,71(3,82)
Test med Ekmans billeder	10	6	1,46(1,03)	6	2	0,90(0,72)	41	11	1,74(1,04)
Samlet problemscore (SDQ)	9	5	14,56(6,41)	6	2	13,83(4,17)	35	11	13,71(5,02)
Følelsesmæssige symptomer	9	5	2,89(2,57)	6	2	3,17(1,83)	35	11	3,66(2,29)
Adfærdsmæssige symptomer	9	5	3,11(1,83)	6	2	2,83(1,94)	35	11	2,14(1,67)
Hyperaktivitet/opmærksom-hedsvanskeligheder	9	5	5,67(2,74)	6	2	4,50(1,97)	35	11	5,34(1,89)
Vanskeligheder ift. jævnald-rende	9	5	2,89(1,69)	6	2	3,33(1,21)	35	11	2,57(1,88)
Sociale styrkesider (SDQ)	9	5	7,44(1,81)	6	2	6,50(3,08)	35	11	7,69(1,41)
Trivsel og funktion (SDQ)	9	5	1,56(2,40)	6	2	1,33(1,86)	35	11	1,54(2,23)
Ked af det/ulykkelig	9	5	0,44(0,88)	6	2	0,00(0,00)	35	11	0,26(0,56)
Familie	9	5	0,33(0,50)	6	2	0,00(0,00)	35	11	0,29(0,57)
Venskaber	9	5	0,11(0,33)	6	2	0,50(0,84)	35	11	0,37(0,65)
Indlæring	9	5	0,44(0,73)	6	2	0,50(0,55)	35	11	0,31(0,53)
Fritid	9	5	0,22(0,67)	6	2	0,33(0,82)	35	11	0,31(0,68)

Note: Observationer (N = observationer, N_i = individer, Cl = anbringelsessteder (clusters)); Gennemsnit (M) og std. dev. (SD)

Bilagstabel 1.4 Opsummering af resultater fordelt på tre indsatsgrupper. Resultatdokumentation med den hierarkiske lineære regressionsmodel

	Ændring i SG over 2 år							
Test og spørgeskemaer	N	N_i	Cl	δ_0	std. err.	p-value	[95% CI]	ES
LEAS, Selv	173	102	21	-0,4	0,29	0,171	[-0,97;0,17]	0,64
LEAS, Andre	173	102	21	-0,34	0,21	0,11	[-0,76;0,08]	0,59
LEAS, Total	173	102	21	-0,43	0,21	0,043*	[-0,84;-0,01]	0,7
RME	173	102	21	1,66	1,13	0,142	[-0,56;3,88]	0,47
Ekmans billeder	173	102	21	-0,45	0,3	0,126	[-1,04;0,13]	0,52
Samlet problemscore (SDQ)	159	101	21	-1,42	1,98	0,475	[-5,30;2,47]	0,27
Følelsesmæssige symptomer	159	101	21	-1	0,61	0,102	[-2,21;0,20]	0,45
Adfærdsmæssige symptomer	159	101	21	0,34	0,86	0,694	[-1,34;2,02]	0,17
Hyperaktivitet/opmærksomhedsvanskeligheder	159	101	21	-0,45	0,55	0,417	[-1,53;0,63]	0,2
Vanskeligheder ift. jævnaldrende	159	101	21	-0,07	0,59	0,903	[-1,24;1,09]	0,04
Sociale styrkesider (SDQ)	159	101	21	-0,2	1,25	0,872	[-2,65;2,25]	0,11
Trivsel og funktion (SDQ)	158	100	21	0,26	0,62	0,675	[-0,96;1,48]	0,11
Ked af det/ulykkelig	158	100	21	0,1	0,24	0,689	[-0,37;0,57]	0,16
Familie	158	100	21	0,13	0,14	0,336	[-0,14;0,40]	0,2
Venskaber	158	100	21	-0,04	0,16	0,803	[-0,35;0,27]	0,07
Indlæring	158	100	21	0,02	0,27	0,955	[-0,52;0,55]	0,02
Fritid	158	100	21	-0,04	0,23	0,867	[-0,50;0,42]	0,06

Note: Antal observationer [N = observationer (Level 1), N_i = individer (Level 2), Cl = anbringelsessteder (Level 3)]; cluster robuste standardfejl (std. err.) og 95-procents-konfidensinterval (95 pct. CI). Mixed effekts estimater af den hierarkiske lineære regressionsmodel. Two-sidet tests med p-værdier (* p < 0,05; ** p < 0,01) tilpasset for robuste standardfejl. Cohen's D effektstørrelser (ES) divideret med total sample SD ved baseline.

Bilagstabel 1.5 Opsummering af resultater fordelt på tre indsatsgrupper. Resultatdokumentation med den hierarkiske lineære regressionsmodel

				Ændring i RP over 2 år				
Test og spørgeskemaer	N	N_i	Cl	$\delta_0+\delta_1$	std, err,	p-value	[95% CI]	ES
LEAS, Selv	173	102	21	0,05	0,04	0,197	[-0,03;0,12]	0,08
LEAS, Andre	173	102	21	0,48	0,16	0,003**	[0,16;0,80]	0,83
LEAS, Total	173	102	21	0,13	0,12	0,284	[-0,11;0,36]	0,21
RME	173	102	21	1,12	0,51	0,029*	[0,12;2,13]	0,32
Ekmans billeder	173	102	21	-0,44	0,27	0,106	[-0,97;0,09]	0,5
Samlet problemscore (SDQ)	159	101	21	-0,89	1,75	0,612	[-4,31;2,54]	0,17
Følelsesmæssige symptomer	159	101	21	-0,35	0,21	0,087	[-0,75;0,05]	0,16
Adfærdsmæssige symptomer	159	101	21	-0,15	0,93	0,875	[-1,98;1,68]	0,07
Hyperaktivitet/opmærksomhedsvanskeligheder	159	101	21	-0,98	0,08	0,000**	[-1,15;-0,82]	0,44
Vanskeligheder ift. jævnaldrende	159	101	21	0,55	0,62	0,376	[-0,66;1,76]	0,34
Sociale styrkesider (SDQ)	159	101	21	-0,35	1,15	0,763	[-2,61;1,92]	0,18
Trivsel og funktion (SDQ)	158	100	21	-0,66	0,94	0,485	[-2,51;1,19]	0,28
Ked af det/ulykkelig	158	100	21	-0,11	0,09	0,22	[-0,28;0,07]	0,18
Familie	158	100	21	-0,55	0,22	0,015*	[-0,99;-0,11]	0,83
Venskaber	158	100	21	-0,01	0,29	0,966	[-0,58;0,56]	0,02
Indlæring	158	100	21	0,14	0,2	0,475	[-0,25;0,54]	0,19
Fritid	158	100	21	-0,32	0,27	0,236	[-0,86;0,21]	0,47

Note: Antal observationer [N = observationer (Level 1), N_i = individer (Level 2), Cl = anbringelsessteder (Level 3)]; cluster robuste standardfejl (std. err.) og 95-procents-konfidensinterval (95 pct. CI). Mixed effekts estimater af den hierarkiske lineære regressionsmodel. Two-sidet tests med p-værdier (* p < 0,05; ** p < 0,01) tilpasset for robuste standardfejl. Cohen's D effektstørrelser (ES) divideret med total sample SD ved baseline.

Bilagstabel 1.6 Opsummering af resultater fordelt på tre indsatsgrupper. Resultatdokumentation med den hierarkiske lineære regressionsmodel

				Ændring i MG over 2 år				
Test og spørgeskemaer	N	N_i	Cl	$\delta_0+\delta_2$	std, err.	p-value	[95% CI]	ES
LEAS, Selv	173	102	21	0,07	0,09	0,401	[-0,10;0,25]	0,12
LEAS, Andre	173	102	21	0,05	0,1	0,593	[-0,14;0,24]	0,09
LEAS, Total	173	102	21	0	0,1	0,977	[-0,20;0,20]	0
RME	173	102	21	1,07	0,59	0,07	[-0,09;2,23]	0,31
Ekmans billeder	173	102	21	-0,22	0,19	0,245	[-0,58;0,15]	0,25
Samlet problemscore (SDQ)	159	101	21	-1,58	0,76	0,037*	[-3,07;-0,10]	0,3
Følelsesmæssige symptomer	159	101	21	-0,58	0,34	0,089	[-1,25;0,09]	0,26
Adfærdsmæssige symptomer	159	101	21	-0,55	0,18	0,002**	[-0,89;-0,21]	0,28
Hyperaktivitet/opmærksomhedsvanskeligheder	159	101	21	-0,08	0,26	0,746	[-0,60;0,43]	0,04
Vanskeligheder ift. jævnaldrende	159	101	21	-0,31	0,24	0,2	[-0,77;0,16]	0,19
Sociale styrkesider (SDQ)	159	101	21	0,32	0,27	0,228	[-0,20;0,85]	0,17
Trivsel og funktion (SDQ)	158	100	21	-0,29	0,6	0,629	[-1,48;0,89]	0,12
Ked af det/ulykkelig	158	100	21	-0,05	0,14	0,707	[-0,32;0,22]	0,09
Familie	158	100	21	0,02	0,16	0,885	[-0,28;0,33]	0,03
Venskaber	158	100	21	0,09	0,13	0,477	[-0,16;0,34]	0,16
Indlæring	158	100	21	-0,34	0,09	0,000**	[-0,52;-0,16]	0,46
Fritid	158	100	21	0,01	0,16	0,959	[-0,31;0,33]	0,01

Note: Antal observationer [N = observationer (Level 1), N_i = individer (Level 2), Cl = anbringelsessteder (Level 3)]; cluster robuste standardfejl (std. err.) og 95-procents-konfidensinterval (95pct. CI). Mixed effekts estimater af den hierarkiske lineære regressionsmodel. Two-sidet tests med p-værdier (* $p < 0,05$; ** $p < 0,01$) tilpasset for robuste standardfejl. Cohen's D effektstørrelser (ES) divideret med total sample SD ved baseline.

Bilagstabel 1.7 Opsummering af resultater fordelt på tre indsatsgrupper. Resultatdokumentation med den hierarkiske lineære regressionsmodel

				Gevinst i RP i forhold til SG				
Test og spørgeskemaer	N	N_i	CI	δ_1	std, err,	p-value	[95% CI]	ES
LEAS, Selv	173	102	21	0,45	0,29	0,127	[-0,13;1,02]	0,72
LEAS, Andre	173	102	21	0,82	0,27	0,002**	[0,29;1,35]	1,42
LEAS, Total	173	102	21	0,55	0,24	0,022*	[0,08;1,03]	0,91
RME	173	102	21	-0,54	1,24	0,664	[-2,97;1,90]	0,15
Ekmans billeder	173	102	21	0,02	0,4	0,966	[-0,77;0,80]	0,02
Samlet problemscore (SDQ)	159	101	21	0,53	2,64	0,841	[-4,65;5,71]	0,1
Følelsesmæssige symptomer	159	101	21	0,65	0,65	0,314	[-0,62;1,92]	0,29
Adfærdsmæssige symptomer	159	101	21	-0,48	1,27	0,703	[-2,97;2,00]	0,24
Hyperaktivitet/opmærksomhedsvanskeligheder	159	101	21	-0,53	0,56	0,339	[-1,62;0,56]	0,24
Vanskeligheder ift. jævnaldrende	159	101	21	0,62	0,86	0,47	[-1,06;2,30]	0,38
Sociale styrkesider (SDQ)	159	101	21	-0,15	1,7	0,931	[-3,48;3,19]	0,08
Trivsel og funktion (SDQ)	158	100	21	-0,92	1,13	0,416	[-3,14;1,30]	0,39
Ked af det/ulykkelig	158	100	21	-0,2	0,26	0,423	[-0,71;0,30]	0,34
Familie	158	100	21	-0,68	0,26	0,010**	[-1,19;-0,16]	1,03
Venskaber	158	100	21	0,03	0,33	0,934	[-0,62;0,68]	0,05
Indlæring	158	100	21	0,13	0,34	0,703	[-0,53;0,79]	0,17
Fritid	158	100	21	-0,28	0,36	0,429	[-0,99;0,42]	0,41

Note: Antal observationer [N = observationer (Level 1), N_i = individer (Level 2), CI = anbringelssteder (Level 3)]; cluster robuste standardfejl (std. err.) og 95-procents-konfidensinterval (95 pct. CI). Mixed effekts estimater af den hierarkiske lineære regressionsmodel. Two-sidet tests med p-værdier (* p < 0,05; ** p < 0,01) tilpasset for robuste standardfejl. Cohen's D effektstørrelser (ES) divideret med total sample SD ved baseline.

Bilagstabel 1.8 Opsummering af resultater fordelt på tre indsatsgrupper. Resultatdokumentation med den hierarkiske lineære regressionsmodel

				Gevinst i MG i forhold til SG				
Test og spørgeskemaer	N	N_i	CI	δ_2	std, err,	p-value	[95% CI]	ES
LEAS, Selv	173	102	21	0,47	0,3	0,12	[-0,12;1,07]	0,76
LEAS, Andre	173	102	21	0,4	0,24	0,094	[-0,07;0,86]	0,68
LEAS, Total	173	102	21	0,43	0,23	0,067	[-0,03;0,89]	0,71
RME	173	102	21	-0,59	1,28	0,645	[-3,09;1,92]	0,17
Ekmans billeder	173	102	21	0,24	0,35	0,5	[-0,45;0,92]	0,27
Samlet problemscore (SDQ)	159	101	21	-0,17	2,12	0,938	[-4,32;3,99]	0,03
Følelsesmæssige symptomer	159	101	21	0,42	0,7	0,548	[-0,96;1,80]	0,19
Adfærdsmæssige symptomer	159	101	21	-0,89	0,88	0,311	[-2,60;0,83]	0,44
Hyperaktivitet/opmærksomhedsvanskeligheder	159	101	21	0,36	0,61	0,552	[-0,83;1,56]	0,16
Vanskeligheder ift. jævnaldrende	159	101	21	-0,23	0,64	0,715	[-1,49;1,02]	0,14
Sociale styrkesider (SDQ)	159	101	21	0,52	1,28	0,682	[-1,98;3,03]	0,28
Trivsel og funktion (SDQ)	158	100	21	-0,55	0,87	0,524	[-2,25;1,15]	0,24
Ked af det/ulykkelig	158	100	21	-0,15	0,28	0,594	[-0,69;0,39]	0,25
Familie	158	100	21	-0,11	0,21	0,597	[-0,52;0,30]	0,17
Venskaber	158	100	21	0,13	0,2	0,524	[-0,27;0,53]	0,23
Indlæring	158	100	21	-0,36	0,29	0,21	[-0,92;0,20]	0,48
Fritid	158	100	21	0,05	0,29	0,868	[-0,51;0,61]	0,07

Note: Antal observationer [N = observationer (Level 1), N_i = individer (Level 2), CI = anbringelsessteder (Level 3)]; cluster robuste standardfejl (std. err.) og 95-procents-konfidensinterval (95 pct. CI). Mixed effekts estimater af den hierarkiske lineære regressionsmodel. Two-sidet tests med p-værdier (* p < 0,05; ** p < 0,01) tilpasset for robuste standardfejl. Cohen's D effektstørrelser (ES) divideret med total sample SD ved baseline.

VIDEN VELFÆRD

DET NATIONALE FORSKNINGS-
OG ANALYSECENTER FOR VELFÆRD