

# Effekt av artikulatøriske bevissthetstrening på skriftspråklige ferdigheter

*En randomisert kontrollert undersøkelse og en  
singel case studie*

Anne Cathrine Thurmann-Moe



Thesis submitted for the degree of Ph.D.  
Department of Special Needs Education

Faculty of Educational Sciences

UNIVERSITY OF OSLO  
2020

© Anne Cathrine Thurmann-Moe, 2021

*Doktoravhandlinger forsvart ved  
Det utdanningsvitenskapelige fakultet, Universitetet i Oslo.  
Nr. 336*

ISSN 1501-8962

Det må ikke kopieres fra denne boka i strid med åndsverkloven eller med avtaler om kopiering inngått med Kopinor, interesseorgan for rettighetshavere til åndsverk.

Omslag: Hanne Baadgaard Utigard.  
Grafisk produksjon: Representralen, Universitetet i Oslo.

## Sammendrag

Både årsakene til dysleksi og måten vanskene framtrer kan være sammensatte og heterogene og dette øker behovet både for ulike metodiske innfallsvinkler i planlegging av tiltak og individuell tilpasning av undervisningen.

Dette prosjektet har hatt to hovedformål. For det første å undersøke effekten av artikulatorisk bevissthetstrening på fonologisk bevissthet og lese- og stavekompetanse. For det andre å drøfte funksjonalitet av Single Case Design (SCD) i evaluering av spesialundervisning etter enkeltvedtak i henhold til Opplæringslovens §5.1. Prosjektet omfatter to eksperimentelle studier, begge gjennomført i samarbeid med lokale skoler. I den første studien (artikkel 1) var utvalget 129 elever på første trinn med skårer under kritiske grense på den obligatoriske nasjonale leseprøven (UDIR 2018). Designet for studien var en randomisert kontrollert undersøkelse. For hver skole ble barna som oppfylte kriteriene fordelt til enten en intervensjonsgruppe eller en kontrollgruppe. Intervensjonsgruppen fikk fem uker med trening i små grupper (20 timer) der målet var å lære barna å bruke artikulatorisk bevissthet som verktøy i grunnleggende lesing og staving. Kontrollgruppen fikk ordinær tilpasset opplæring. Resultatene viste ingen signifikante forskjeller mellom gruppene.

I den andre studien (artikkel 2 og 3) var deltakerne mellom 10 og 15 år og diagnostisert med dysleksi (N=13). Intervensjonsprogrammet var hovedsakelig det samme som i den første studien, men deltakerne fikk åtte ukers individuell trening (32 timer). Det ble brukt et SCD med totalt 18 målinger av lese- og staveferdighet. Generelt viste resultatene signifikant framgang i råskårer og lesenøyaktighet i intervensjon- og postfasen sammenliknet med baselinefasen. Imidlertid var det også signifikant effekt på en ekstern oppgave målt på samme tidspunkter, noe som indikerer mulig testeffekt av de gjentatte kartleggingene. Det konkluderes med at flertallet av elevene har vist god framgang, men at forekomsten av mulig testeffekt svekker grunnlaget for sikre konklusjoner. I artikkel 2 diskuteres både egenskaper og metodiske utfordringer ved SCD i et bredere perspektiv. Her drøftes også særlig bruk av SCD i vurdering av effekt av tiltak for enkeltelever og det konkluderes med at bruk av SCD kan bidra til bedret praksis rundt evaluering av spesialundervisning i skolen.

## Artikkel 1-3

### Artikkel 1

Thurmann-Moe, A.C., Melby-Lervåg, M & Lervåg, A. Effects of Articulatory Consciousness Training in First Graders with a Reading Delay: A Randomised Control Trial.

Status: Fagfelleurdert og akseptert for publisering i *Scandinavian Journal of Educational Research*

### Artikkel 2

Thurmann-Moe, A.C., Melby-Lervåg, M & Lervåg, A. Single Case Design i evaluering av spesialpedagogiske tiltak: En gjennomgang av aktuelle analysemetoder for vurdering av effekt.

Status: Fagfelleurdert og akseptert for publisering i *Spesialpedagogikk*

### Artikkel 3

Thurmann-Moe, A.C., Melby-Lervåg, M & Lervåg, A. The Impact of Articulatory Consciousness Training on Reading and Spelling Literacy in Students with Severe Dyslexia: An Experimental Single Case Study.

Status: Sendt inn til *Annals of Dyslexia* 11.10.2020

## Forord

Først og fremst rettes det en stor takk til elever, lærere og foreldre i de involverte kommunene som på forskjellige måter har bidratt og deltatt i dette prosjektet. En spesiell takk til ansatte i pedagogisk psykologisk tjeneste for stor innsats i forbindelse med de ukentlige kartleggingene av elevene i studie 2.

Takk til Norges Forskningsråd som gjennom stipendordningen for Offentlig Ph.d. bidrar til praksisnær forskning og til å øke kvaliteten på tilbudene i offentlig sektor. Takk til arbeidsgiver Statped som har lagt til rette for gjennomføring av prosjektet både gjennom delfinansiering og praktisk tilrettelegging. Spesiell takk til tidligere og nåværende avdelingsleder i avdeling for språk og tale i Statped sørøst, Anne Berit Andreassen og Marit Bollingmo, som gjennom positivt engasjement og gunstig tilrettelegging av arbeidsoppgaver har lagt til rette for realisering av prosjektet.

Stor takk til Arne Lervåg og Monica Melby-Lervåg for veiledning gjennom alle disse årene! Takk også til Øistein Anmarkrud for gode innspill i forbindelse med gjennomlesing i avslutningsfasen og til anonyme fagfeller som gjennom sine kommentarer har bidratt til å øke kvaliteten på artiklene.

Takk til alle snille nåværende og tidligere kollegaer i avdeling Språk/Tale for diskusjoner, interesse og engasjement i forhold til prosjektet. Spesiell takk til Christina, Henriette, Ingrid og Stig for praktisk hjelp i forbindelse med datainnsamlingen i studie 1.

Tusen takk til Institutt for Spesialpedagogikk og alle hyggelige folk der for tilgang til kontorplass, trivelig lunsjmiljø, fredagsvafler og veldig god te!

### Referanser til de publiserte artiklene:

Thurmann-Moe, A. C., Melby-Lervåg, M., & Lervåg, A. (2021). Effects of Articulatory Consciousness Training in First Graders with a Reading Delay: A Randomised Control Trial. *Scandinavian Journal of Educational Research*, 1-18. doi:10.1080/00313831.2020.1869823

Thurmann-Moe, A. C., Melby-Lervåg, M., & Lervåg, A. (2021). Single Case Design i evaluering av spesialpedagogiske tiltak: En gjennomgang av aktuelle analysemetoder for vurdering av effekt. *Spesialpedagogikk* 01/21, ISSN 0332-8457

Thurmann-Moe, A. C., Melby-Lervåg, M., & Lervåg, A. (2021). The impact of articulatory consciousness training on reading and spelling literacy in students with severe dyslexia: an experimental single case study. *Annals of Dyslexia*. doi:10.1007/s11881-021-00225-1



## Innhold

Sammendrag.....	V
Artikkel 1-3 .....	VI
Forord.....	VII
<b>1 Innledning .....</b>	<b>1</b>
1.1 Formål med prosjektet.....	1
1.2 Plan over avhandlingen .....	4
<b>2 Lesing og lesevansker.....</b>	<b>6</b>
2. 1 Fonologisk bevissthet og fonologiske representasjoner .....	7
2. 2 Hypotesen om fonologisk svikt.....	8
2. 3 Dysleksi.....	10
2. 4 Kausale sammenhenger ved dysleksi.....	11
2. 4. 1 Sensoriske forhold .....	12
2. 4. 2 Forhold knyttet til talepersepsjon.....	12
2. 4. 3 Forhold knyttet til taleproduksjon .....	15
2. 5 Artikulatorisk bevissthet .....	17
2. 5. 1 Virkemidler i trening av artikulatorisk bevissthet.....	18
2. 6 Oppsummering .....	19
<b>3 Vurdering av kausalitet i intervensjonsstudier .....</b>	<b>20</b>
3. 1 Hvordan vurdere hva som virker? .....	20
3. 2 Effektbegrepet .....	21
3. 3 Ytre validitet og generalisering.....	21
3. 4 «Hva virker» - bevegelsen og SCD.....	22
3. 5 Predikert effekt og effekt for den enkelte .....	23
3. 6 Evaluering av spesialpedagogiske tiltak etter enkeltvedtak.....	23
3. 7 Oppsummering.....	24
<b>4 Teoretisk grunnlag for intervensjonene .....</b>	<b>25</b>
<b>5 Metodiske avveininger.....</b>	<b>28</b>
5. 1 Utvalg.....	28

<b>5. 2 Valg av design</b> .....	32
<b>5. 3 Aktuelt intervensjonsprogram</b> .....	36
<b>5. 4 Prosedyrer for datainnsamling</b> .....	37
<b>5. 5 Ytre validitet</b> .....	38
<b>5. 6 Indre Validitet</b> .....	39
5. 6. 1 Seleksjon .....	39
5. 6. 2 Frafall .....	39
5. 6. 3 Modning .....	40
5. 6. 4 Testeffekt.....	40
<b>5. 7 Begrepsvaliditet</b> .....	43
<b>5. 7. 1 Operasjonalisering av begreper</b> .....	44
<b>5. 7. 2 Validitet knyttet til måleverktøy</b> .....	46
<b>5. 7. 3 Implementeringsvaliditet</b> .....	48
<b>5. 7. 4 Diffusjon</b> .....	49
<b>5. 7. 5 «Performance effekt»</b> .....	49
<b>5. 8 Sosial validitet</b> .....	50
<b>5. 9 Vurdering av effekt</b> .....	51
<b>5. 9. 1 Effektmål</b> .....	51
5. 9. 2 Visuelle analyser .....	51
<b>5. 9. 3 Statistiske analyser</b> .....	52
<b>5. 10 Etske problemstillinger</b> .....	53
<b>6 Oppsummering av artiklene</b> .....	<b>55</b>
<b>7 Diskusjon</b> .....	<b>58</b>
<b>7. 1 Effekten av artikulatorisk bevissthetstrening</b> .....	58
<b>7. 1. 1 Resultater vurdert i lys av en flerfaktormodell for dysleksi</b> .....	58
<b>7. 1. 2 Utvikling av artikulatorisk bevissthet</b> .....	60
<b>7. 1. 3 Effekt av treningen på ordavkoding</b> .....	62
<b>7. 1. 4 Effekt av treningen på staveferdighet</b> .....	63
<b>7. 1. 5 Effekt av treningen på fonologisk bevissthet</b> .....	63
<b>7. 1. 6 Effekt av treningen på bokstavkunnskap</b> .....	63
<b>7. 1. 7 Effekt av artikulatorisk trening sammenliknet med fonologisk basert trening</b> .....	63



7. 1. 8 Effekt av artikulorisk trening for elever med spesielt svake fonologiske ferdigheter .....	65
7. 2 Single Case Design i evaluering av spesialpedagogiske tiltak .....	66
7. 2. 1 Hva virker og hvordan vurdere det? .....	67
7. 3 Avsluttende konklusjoner .....	69
<b>Litteratur .....</b>	<b>72</b>

Artikkel 1-3

Appendiks



# 1 Innledning

## 1.1 Formål med prosjektet

Lese- og skriveferdigheter er viktige for å lykkes med skolegang og videre i arbeidslivet. Vansker knyttet til utvikling av disse ferdighetene fører derfor ofte bekymring både hos lærere og foreldre og kan få alvorlige personlige konsekvenser for dem det gjelder. Forståelse av vanskene og ikke minst utvikling og utprøving av tiltak som kan virke forebyggende er derfor viktig. Lesevansker kan deles inn i vansker med å forstå lest tekst og vansker med selve den tekniske delen av lesingen, avkodingen (dysleksi) (Gough & Tunmer, 1986). Disse to komponentene er distinkte og har røtter i den tidlige språkutviklingen. Mens leseforståelse er relatert til flere sider ved språket, slik som språkforståelse og grammatikk, er dysleksi først og fremst knyttet til tidlig utvikling av det fonologiske systemet. I et longitudinelt perspektiv er også tidlige tegn på fonologiske vansker en prediktor for dysleksi (Melby-Lervåg, Lyster & Hulme, 2012).

Fonologisk svikt regnes som en kjernevanske ved dysleksi (Lyon, Shaywitz, & Shaywitz, 2003) og forklares ofte med henvisning til problemer knyttet til den auditive kanalen, noe som også har preget tiltakstenkningen. Det er generelt bred enighet om at undervisning som tydeliggjør det fonologiske grunnlaget for skriftspråket gjennom syntetiske leseopplæringsmetoder (phonics) gir best effekt for elever med dysleksi (Galuschka, Ise, Krick, & Schulte-Korne, 2014; McArthur et al., 2012; SBU, 2014). Generelt har også trening av fonologisk bevissthet vist seg å ha positiv effekt på den tidlige utviklingen av leseferdighet (Bus & Ijzendoorn, 1999; Ehri et al., 2001; Lundberg, Frost & Petersen, 1988). Denne typen trening følger samme grunnprinsipp som i syntetiske leseopplæringsmetoder, men foregår i muntlig form og er primært lyttebasert.

Til tross for at disse metodiske prinsippene for leseundervisning har stor oppslutning i praksisfeltet og i Norge også er nedfelt i læreplanen, er det imidlertid en del elever som viser seg mer eller mindre resistente mot denne typen opplæring og som til tross for systematisk undervisning viser liten progresjon.

Som nevnt regnes fonologisk svikt som en kjernevanske ved dysleksi (Landerl et al., 2013; Swan & Goswami, 1997). Samtidig er nyere forskning i økende grad opptatt av både heterogene årsakssammenhenger og individuelle forskjeller når det gjelder i hvilken grad fonologiske vansker faktisk fører til problemer med avkodning (Moll, Loff, & Snowling, 2013; Pennington, 2006; Ramus, Marshall, Rosen, & van der Lely, 2013; Snowling & Melby-Lervåg, 2016). Forståelsen av dysleksi som en heterogen vanske øker også behovet både for mer finmaskede verktøy for kartlegging og individuelt tilpassede tiltak for denne gruppen elever.

Et mulig supplement til lyttebasert opplæring er å lære elevene å bli bevisste på hvordan lydene formuleres og hvordan ordenes fonologiske oppbygging kan defineres gjennom artikulatoriske egenskaper ved hver enkelt språklyd (artikulatorisk bevissthet). En slik innfallsvinkel kan teoretisk begrunnes utfra sammenhengen mellom taleproduksjon og talepersepsjon og hvordan begge disse komponentene virker inn på utvikling av avkodingsferdigheter (Anthony et al., 2010; Hulme & Snowling, 2009; Liberman & Whalen, 2000). Tidligere studier har funnet at elever med dysleksi avviker fra typiske utviklede elever når det gjelder artikulatoriske ferdigheter både knyttet til hastighet og artikulatorisk bevissthet (Duranovic, Sehic, Ravid, & Schiff, 2013; Fawcett & Nicolson, 2002; Foy & Mann, 2012; Griffiths & Frith, 2002; Montgomery, 1981). Intervensjonsstudier som har brukt ulike former for artikulatorisk bevissthetstrening for å øke fonologisk bevissthet og avkodingsferdigheter har generelt gitt noe divergerende resultater, delvis påvirket av hvilken målgruppe som er undersøkt (Boyer & Ehri, 2011; Fälth, Gustafson, & Svensson, 2017; Trainin, Wilson, Murphy-Yagil, & Rankin-Erickson, 2014; Wise, Ring, & Olson, 1999).

I norsk sammenheng har artikulatorisk bevissthetstrening med bruk av materialet PAS (Piktografisk Artikulatorisk System) (Kausrud, 2003) vært brukt i leseundervisning av barn med språkvansker. Materialet brukes også i logopedisk behandling og en del skoler har også tatt i bruk PAS som støtte i bokstavinnlæringen i første klasse. Bortsett fra en casestudie av en gutt med språkvansker (Ottem & Kausrud, 2001) har effekten ved bruk av materialet ikke vært systematisk vurdert. Et mål for dette prosjektet har derfor vært å bidra til å belyse effekten av artikulatorisk bevissthetstrening i forhold til avkodingsferdigheter og fonologisk bevissthet gjennom systematisk utprøving med PAS materialet.

Utdanningssystemet etterspør i økende grad forskningsbasert kunnskap om «hva som virker». Dette reiser både vitenskapsteoretiske og praktiske problemstillinger knyttet til hvordan forskningsfunn kan omsettes til praktisk-pedagogiske tiltak både i et generelt perspektiv og for den enkelte elev. I lys av dette eksemplifiseres det i denne avhandlingen hvordan effekten av samme tiltak kan belyses gjennom to ulike forskningsmetodiske tilnærminger, henholdsvis en randomisert kontrollert undersøkelse (RCT Randomized Controlled Trial) og et Single Case design (SCD). Begge disse er forskningsdesign som kan frambringe kunnskap om hva som virker. Imidlertid er de også på enkelte områder å regne som kontraster og bygger på noe ulik logikk (de Bruin, 2017; Lobo, Moeyaert, Baraldi Cunha, & Babik, 2017; Onghena, Michiels, Jamshidi, Moeyaert, & Van Den Noortgate, 2018).

I begge de to empiriske studiene i denne avhandlingen er målgruppen for tiltaket elever med lesevansker knyttet til den fonologiske komponenten i språket. I begge tilfeller er også utprøving av tiltaket knyttet til konkrete utfordringer i praksisfeltet.

I den første studien var målgruppen elever på første klassetrinn med forsinket leseutvikling. I denne studien ble intervensjonen realisert gjennom et samarbeid med den sentrale skoleadministrasjonen i en kommune. Til tross for at den aktuelle kommunen hadde iverksatt tiltak for å styrke begynneropplæringen i lesing, var andelen elever som skåret under kritisk grense på den nasjonale leseprøven i første klasse høyere enn ønskelig. Ledelsen ved skolekontoret ønsket å vurdere hvorvidt det ville være hensiktsmessig å iverksette intensiv trening i små grupper rettet mot de elevene som hadde skåret under kritisk grense. Skolekontoret var også positiv til å prøve ut et tiltak med en annen metodisk innfallsvinkel enn den som allerede var i bruk i klasserommene.

I den andre studien var målgruppen for tiltaket elever med alvorlig grad av dysleksi og utvalget ble hentet fra ventelistene til det nasjonale spesialpedagogiske støttesystemet (Statped). Samtlige av disse elevene var henvist på grunn av stagnert leseutvikling og behov for nye tiltak.

Utvalget i studie 2 representerer også den andelen av norske skoleelever som mottar spesialundervisning etter enkeltvedtak hjemlet i grunnskolelovens §5. 1. [https://lovdata.no/dokument/NL/lov/1998-07-17-61/KAPITTEL\\_6#KAPITTEL\\_6](https://lovdata.no/dokument/NL/lov/1998-07-17-61/KAPITTEL_6#KAPITTEL_6)

I lys av erfaringene fra studie 2 drøfter denne avhandlingen derfor også på hvilken måte Single Case Design (SCD) kan bidra til å evaluere effekt av konkrete tiltak innenfor rammen av spesialundervisning.

Formålet med avhandlingen kan derfor sies å være tosidig: For det første å undersøke hvorvidt økt bevissthet om språklydenes artikulatoriske egenskaper kan fungere kompensere for elever som enten har dysleksi (studie 2) eller er i risikogruppen på grunn av forsinket leseutvikling og svak fonologisk bevissthet (studie 1). For det andre å drøfte funksjonalitet av SCD når det gjelder å evaluere konkrete spesialpedagogiske tiltak for enkeltelever. Avhandlingen består av tre artikler basert på de to studiene presentert ovenfor.

Den første artikkelen omhandler effekten av en randomisert kontrollert undersøkelse av artikulatorisk bevissthetstrening i et utvalg på 129 elever på 1. trinn med resultater under kritisk grense på den nasjonale prøven i lesing.

Den andre artikkelen er en metodisk gjennomgang av SCD i spesialpedagogisk virksomhet. Formålet med denne artikkelen er primært å drøfte designets egnethet i forhold til evaluering av spesialundervisning for enkeltelever innenfor det norske tiltakssystemet. Artikkelen eksemplifiserer også hvordan SCD konkret kan brukes på skolene. For å vise dette er det valgt ut fire elev eksempeler fra studie to. Både det begrensede formatet (4 elever) og målsettingen knyttet til evaluering av effekt hos enkeltdeltakere kan sies å representere en «klassisk» bruk av SCD. Potensielle muligheter for at bruk av SCD kan produsere mer generalisert kunnskap gjennom replikasjoner diskuteres også.

Den tredje artikkelen omhandler artikulatorisk bevissthetstrening i et utvalg på 13 elever mellom 10 og 15 år med alvorlig grad av dysleksi (-2st fra aldersgjennomsnittet på to nonord leseprøver). Her brukes en tilpasset variant av multiple baseline/probe design. I denne artikkelen fokuseres det både på elevenes individuelle utbytte og den gjennomsnittlige felles effekten av tiltaket.

## **1.2 Plan over avhandlingen**

Videre i denne introduksjonsdelen vil det i kapittel 2 først bli presentert teoretiske perspektiver på lesing og dysleksi, deretter presenteres grunnlaget for artikulatorisk bevissthetstrenings i forhold til elever med lese- og skrivevansker. I kapittel 3 diskuteres vurderingspraksis når det

gjelder spesialundervisning med utgangspunkt i «hva virker» bevegelsen og egenskaper ved SCD. I kapitel 4 presenteres det teoretiske grunnlaget for intervensjonene. I kapitel 5 redegjøres det for metodiske valg og avveininger i prosjektet. I kapitel 6 presenteres resultater fra studiene med utgangspunkt i de tre artiklene. I kapitel 7 diskuteres resultatene både med utgangspunkt i overordnede problemstillinger og de målte variablene i de to empiriske studiene.

## 2 Lesing og lesevansker

Helt siden The simple View of Reading (SVR) (Gogh & Tunmer 1986) først ble foreslått og fram til i dag har denne modellen hatt stor innflytelse både for forståelsen av lesevansker og for utvikling av tiltak (Bishop & Snowling 2004; Catts, Adlof, Hogan, & Weismer, 2005; Ramus, Marshall, Rosen, & van der Lely, 2013). SVR modellen forklarer leseforståelse som produktet av ferdigheter i ordavkodning og språkforståelse og betegner i utgangspunktet disse to som likeverdige komponenter. Empiriske studier har dokumentert at så mye som 95-99% av variasjonen når det gjelder leseforståelse hos elever i barneskolealder kan forklares utfra denne modellen (Hjetland et al., 2018; Lervåg, Hulme, & Melby-Lervåg, 2018). Modellen har også vist seg å ha prediksjonsverdi i et longitudinelt perspektiv (Elwér, Keenan, Olson, Byrne, & Samuelsson, 2013; Torppa et al., 2007).

Nyere studier har vist komponentenes innflytelse på leseproduktet ikke er konstant, men endres i takt med barnas utvikling (Cain, Catts, Hogan, & Lomax, 2015; Catts & Vaughn, 2018; Lervåg et al., 2018). I et longitudinelt perspektiv har avkodingskomponenten mest å si for leseforståelse de første årene, mens språkforståelse får mer å si etter hvert som barna blir eldre og tekstene blir vanskeligere (Hjetland, Brinchmann, Scherer, & Melby-Lervåg, 2017; Kim, Wagner, & Lopez, 2012). I tråd med dette har nyere studier nyansert forståelsen av avkodingskomponenten i SVR modellen og vist at det er relevant å se leseflyt, det vil si automatisert ordavkodning, som et produkt av lesenøyaktighet (accuracy) og lesehastighet (speed) (Juil, Poulsen, & Elbro, 2014; Poulsen & Elbro, 2018; Protopapas, Simos, Sideridis, & Mouzaki, 2012). I en studie fra Danmark som fulgte et utvalg typisk utviklede elever fra barnehagen og til andre klasse viste Juul, Poulson og Elbro (2014) også at utvikling av lesehastighet først skjøt fart når elevene leste med ca. 70 % nøyaktighet.

Typisk leseutvikling kan forstås som en gradvis utvikling fra elementær fonologisk avkodning til automatisert ordavkodning (Ehri 1995; Hulme & Snowling 2009). Det er bred enighet om at avkodingskompetanse på avansert nivå kjennetegnes ved en fleksibel bruk av ulike strategier for avkodning (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Høien & Lundberg, 1992; Seidenberg 1993). Hvorvidt de ulike strategiene har røtter i kvalitativt ulike prosesser, slik det blant annet legges til grunn innenfor såkalte toveis modeller for lesing (dual



route) (Coltheart, et al, 2001; Høien & Lundberg, 1992), eller kan forstås som resultat av et samspill mellom flere kognitive faktorer (Seidenberg 1993) har imidlertid vært omdiskutert.

## 2. 1 Fonologisk bevissthet og fonologiske representasjoner

Fonologisk bevissthet er en viktig prediktor for utviklingen av avkodingskomponenten i SVR og angår barnets evne til å fokusere på språket i et metaperspektiv ved å rette oppmerksomheten mot mindre bestanddeler i språket (Ehri et al., 2001). Begrepet betegner et relativt bredt spekter av ferdigheter på ordnivå, setningsnivå, stavelsesnivå, rim og fonemisk nivå. Utviklingen av fonologisk bevissthet har vært antatt å gå fra helhet til del, der barnet først retter bevissthet mot større segmenter i språket, som setninger, stavelser og rimord og deretter gradvis utvikler fonemisk bevissthet (Hayward et al., 2017; Swan & Goswami, 1997).

Det er enighet om at fonologisk bevissthet, og da i særlig grad fonemisk bevissthet, er viktig for utvikling av avkodingsferdigheter. Imidlertid er spørsmålet om hva som konstituerer det fonologiske grunnlaget for avkoding, ofte betegnet som fonologiske representasjoner, ikke entydig besvart (Boada & Pennington 2006). Fonologiske representasjoner er et teoretisk konstrukt og bygger på en hypotese om at det foregår en mental lagring av detaljert informasjon om hver enkelt språklyds fonologiske egenskaper (Elbro & Jensen, 2005). Språklydenes identitet, både når det gjelder finstemte fonetiske og akustiske egenskaper, samt implisitte «regler» for den fonetiske grenseoppgangen mot andre språklyder, utgjør dermed hver enkelt språklyds fonologiske representasjon (Boada & Pennington 2006). Både kvalitative aspekter ved de fonologiske representasjonene og den mentale tilgangen til fonologisk informasjon i leksikon spiller en rolle i avkodingsprosessen (Anthony et al., 2010).

Undersøkelser tyder på at utvikling av kunnskap om fonologiske kjennetegn i morsmålet starter allerede tidlig i spedbarnsalderen (Hoonhorst et al., 2009; Hämäläinen, Salminen, & Leppänen, 2013; Lohvansuu, Hämäläinen, Ervast, Lyytinen, & Leppänen, 2018) og at det så skjer en gradvis spesifisering og omstrukturering av representasjonene i takt med den generelle utviklingen (Horlyck, Reid, & Burnham, 2012; Swan & Goswami, 1997). Spesifisering av de fonologiske representasjonene i fonemisk retning innebærer at barnet blir i stand til å rette oppmerksomheten mot mindre detaljer ved språket, herunder også mer subtile kontraster mellom lydene som ikke nødvendigvis er relevante for forståelsen av talt språk. Denne utviklingen påvirkes av barnets tidlige erfaringer med skriftspråket, og kan også

stimuleres pedagogisk gjennom tilrettelegging av oppgaver der barna trenes i å identifisere og manipulere fonemer med utgangspunkt i talespråket (Ehri et al., 2001; Lundberg, Frost, & Petersen, 1988).

Organisering av fonologiske representasjoner i retning av mer spesifikke fonemiske kategorier er et nødvendig utgangspunkt for typisk leseutvikling i alfabetiske skriftspråk. Studier har imidlertid vist at forholdet ikke kun bør forstås lineært og at lesing og fonemisk bevissthet påvirker hverandre gjensidig (Perfetti, Beck, Bell, Hughes, & Perfetti, 1987 ; Hulme, Snowling, Caravolas & Carrol 2005 ; Clayton, West, Sears, Hulme& Lervåg 2020.) Svak fonemisk bevissthet i utvalg med dysleksi kan derfor også sees som en konsekvens av lite leseerfaring (Ziegler & Goswami, 2005).

## **2. 2 Hypotesen om fonologisk svikt**

I henhold til hypotesen om fonologisk svikt (Swan & Goswami, 1997) er årsaken til de fonologiske problemene man ser hos elever med dysleksi forårsaket av at de fonologiske representasjonene er underspesifiserte og dermed gir et usikkert grunnlag for å knytte bokstavtegn til den enkelte lyd.

Elbro og kollegaer (Elbro & Jensen, 2005; Elbro, Petersen, & Borstrom 1998) har i flere studier undersøkt fonologiske representasjoner i utvalg av førskolebarn og skoleelever med og uten dysleksi. De fonologiske representasjonenes har i disse studiene blitt målt gjennom oppgaver der forsøkspersonene «samhandler» med en hånddokka (testleder) gjennom å se på bilder av objekter. «Håndokka» uttaler ordene på bildene feil (fonologisk underspesifisert) og barna blir så bedt om å demonstrere riktig uttale av ordene sakte og nøyaktig «for å lære dokka riktig uttale». Disse studiene har dokumentert at barn med svake fonologiske ferdigheter har større problemer enn barn med typisk fonologisk utvikling når det gjelder riktig uttale. Grundigere analyse av hvordan barna korrigerer egne uttalefeil viste også at elevene med dysleksi syntes å ha mindre distinkte representasjoner ved at de oftere enn kontrollpersoner hoppet over de fonemene som er minst framtrædende i uttale av ordene. Nøyaktigheten i barnas uttale målt i førskolealderen var i denne studien en unik prediktor for senere leseproblemer, også når fonologisk bevissthet, bokstavkunnskap og artikulasjonshastighet ble tatt i betraktning (Elbro et al., 1998).

Liknende resultater er også rapportert i studier der uttale av enkeltfonemer målt i førskolealderen har vist seg å både predikere fonologisk bevissthet og senere fonologisk avkoding (Mann & Foy 2007; Thomas & Sénéchal 2004).

Hypotesen om fonologisk svikt innebærer også en enkel kausal forklaringsmodell for dysleksi ved at uspesifiserte fonologiske representasjoner antas å føre til svak fonologisk bevissthet og påfølgende avkodingsproblemer (Figur 2.1).

### Figur 2.1

*Enkel kausal modell for dysleksi basert på hypotesen om fonologisk svikt*



Som supplement til denne hypotesen er nyere forskning i økende grad opptatt av heterogene og sammensatte årsaksfaktorer til dysleksi (Noordenbos & Serniclaes, 2015; Snowling & Melby-Lervåg, 2016). Dette innebærer for det første at den opplevde lesevansken kan forsterkes og modereres av generelle kognitive og miljømessige faktorer som er eksterne til det fonologiske problemet (Pennington & Bishop 2009; Catts & Vough 2018; Hulme & Snowling 2009; Snowling & Melby-Lervåg 2016). For det andre innebærer det at også at selve det fonologiske grunnproblemet kan ha ulikt opphav knyttet til svikt i ulike deler av den kausale kjeden (Hulme & Snowling 2009).

## 2.3 Dysleksi

Dysleksi er en spesifikk vanske med lesing og skriving som både fører til sakte avkoding og til unøyaktig avkoding og staving (Lyon et al., 2003). Dysleksi er også er hyppig komorbid med språklydsvansker, ADHD og utviklingsmessige språkvansker (Bishop & Snowling, 2004; Catts, Adlof, Hogan, & Weismer, 2005; Pennington, 2006; Ramus et al., 2013). Årsaken til dysleksi er ikke kjent, men det synes klart at både arv og miljømessige forhold spiller inn (Byrne et al., 2006; Christopher et al., 2013; Olson et al., 2011).

Forekomsten av skolelever som har dysleksi anslås i noen undersøkelser å være ca 7% av befolkningen (Hulme & Snowling, 2016) mens andre rapporter angir at forekomsten er mellom 3-5 % (SBU, 2014). Siden avkodingsvansker er normalfordelt kan man også definere dysleksi ut fra avvik fra et standardisert gjennomsnitt for alderen. SBU (2014) bruker i sin utvelgelse av publikasjoner - 1,25 standardavviksenheter, som tilsvarer ca 10 persentil, mens det i andre studier benyttes et avvik fra gjennomsnittet på 2 st og 1,5st (Hulme & Snowling 2009).

Symptomer på dysleksi varierer med alder og synes også delvis å være påvirket av språkmiljø. Forekomst av dysleksi antas å være større i språkkulturer med høy frekvens av inkonsistent stavemåte, som for eksempel i engelsk, enn i språkkulturer med transparent ortografi, som i Finland (Ziegler, Perry, Ma-Wyatt, Ladner, & Schulte-Korne, 2003). I tråd med dette har noen undersøkelser konkludert med at symptomer på dysleksi i land med transparente og semitransparente ortografier i mindre grad knyttes til feillesinger og i større grad til lesehastighet (Landerl et al., 2013). Andre undersøkelser har imidlertid vist at unøyaktig lesing og staving hos elever med dysleksi, om enn i mindre grad, også forekommer i språkkulturer med transparente ortografier (Caravolas, Volín, & Caravolas, 2001; Pennala et al., 2010; Ziegler et al., 2003).

I medisinske diagnosemanualer som ICD 10 fra Verdens helseorganisasjon (<https://www.who.int/classifications/icd/en/>) og det amerikanske DSM systemet (Diagnostic and Statistical Manual of Mental Disorders) (American Psychiatric Association (APA) 2013) er dysleksi som begrep ikke skilt ut som egen vanske. Vansker med lesing inngår i stedet som en underkategori under merkelappen «spesifikke lærevansker» (Specific Learning Disabilities) (APA 2013) Et problem med denne definisjonen er at den ikke skiller klart mellom avkodingsvansker og leseforståelsesvansker. Denne begrensningen gjør definisjonen

lite egnet som utgangspunkt for forskning rundt ulike årsaksfaktorer når det gjelder lesevanskene (Hulme & Snowling, 2009; Snowling & Hulme, 2012).

Dysleksibegrepet har vært omdiskutert og nyere studier har også stilt spørsmålsteget ved hvorvidt det er hensiktsmessig å definere dysleksi kategorisk og foreslått at man som alternativ ser fonologiske ferdigheter langs et kontinuum, der de med svakest utgangspunkt er i risiko for utvikling av avkodingsvansker (Wagner et al., 2019). I et slikt perspektiv vil graden av opplevd dysleksi kunne modereres av forebyggende faktorer og forsterkes ved forekomst av andre vansker (Moll, Loff, & Snowling, 2013). I nyere definisjoner av dysleksi tas det avstand fra en kategorisk forståelse. Vanskene defineres heller ikke utelukkende ut fra fonologiske problemer, men inkluderer også problemer med verbalt minne og verbal prosesseringshastighet. I tillegg defineres alvorlighetsgraden av vanskene i lys av manglende respons på godt funderte tiltak (Rose 2009).

## **2. 4 Kausale sammenhenger ved dysleksi**

Hypotesen om at uklare og lite distinkte fonologiske representasjoner kan være en årsaksfaktor ved dysleksi, reiser spørsmål knyttet til hvordan de fonologiske representasjonene er konstituert. Mer spesifikt dreier det seg om hvordan persepsjonsprosessen fra talt språk til mental representasjon av det fonologiske systemet foregår og hvordan talepersepsjon og taleproduksjon henger sammen. Forklaringer og hypoteser rundt dette inkluderer både såkalt «bottom up» perspektiver der fokus rettes mot sensoriske forhold ved talepersepsjon (Cumming, Wilson, & Goswami, 2015; Hämäläinen et al., 2013; Tallal, 1980) og «top down» perspektiver der talepersepsjon relateres til generelle kognitive og leksikalske forhold (Boada & Pennington, 2006; McBride-Chang, 1996; Noordenbos & Serniclaes, 2015; Ramus et al., 2013).

Hvorvidt problemer med uttale av språklyder er en årsak til eller en konsekvens av uklare fonologiske representasjoner kan diskuteres (Hulme & Snowling 2009; Mann & Foy 2011). På hvilken måte ulike former for sensorisk sensitivitet og persepsjon av ulike dimensjoner ved det talte språket henger sammen med leseferdighet er heller ikke fullt ut forstått, men flere studier konkluderer med heterogene og sammensatte årsaksfaktorer både på kognitivt, sensorisk og perseptuelt nivå (Hulme & Snowling 2009; Moll et al., 2003; Pennington & Bishop, 2009; Pennalla et al 2010; Snowling & Melby-Lervåg 2016).

### **2. 4. 1 Sensoriske forhold**

Hypotesen om en generell auditiv svikt som årsaksfaktor for dysleksi knyttes særlig til arbeidet til Paula Tallal. Tallal (1980) hevdet at de observerte lesevanskene hos elever med dysleksi kunne knyttes til problemer med å oppfatte de raskt skiftende auditive signalene i talt språk. I henhold til hypotesen fører dette til at det etableres upresise fonologiske representasjoner som gir et dårlig grunnlag for nøyaktig avkoding. Flere senere studier har bekreftet antakelsen om auditive diskriminasjonsvansker hos elever med dysleksi, men lagt mindre vekt på temporale forhold og mer vekt på auditiv sensitivitet knyttet til oppfattelse av rytmiske forhold som frekvens og stigningsgrad (f. eks Cumming et al., 2015). I en gjennomgang av studier som sammenliknet leseferdighet med ulike mål på auditiv sensitivitet ved hjelp av nevrofysiologiske målemetoder, Event Related Potentials (ERP) fant Hämäläinen et al (2012) at det både hos voksne og barn med dysleksi forekom atypiske mønstre i auditiv prosessering knyttet til oppfatning av tonegang, frekvens, stigningsgrad og varigheten av lydsignaler sammenliknet med typisk utviklede lesere. Denne studien omhandler imidlertid ikke sammenhengen mellom disse atypiske mønstrene og leseferdighet i et kausalt perspektiv.

En innvending mot å trekke for sikre konklusjoner fra denne typen studier har vært usikkerhet knyttet til målemetodenes validitet ved at også andre faktorer enn de som er relatert til auditiv diskriminering spiller inn. Særlig gjelder dette forhold knyttet til svakt fonologisk korttidsminne som ofte forekommer i utvalg med dysleksi (McBride-Chang, 1995) og graden av leseerfaring (Goswami, 2014).

### **2. 4. 2 Forhold knyttet til talepersepsjon**

Andre studier har nedtonet betydningen av en eventuell generell auditiv diskriminasjonssvikt og i stedet fokusert på avvikende forhold knyttet til det leksikalske og subleksikalske grunnlaget for talepersepsjon. I henhold til den såkalte feilkategoriseringshypotesen oppfattes de typiske kjennetegnene ved dysleksi, for eksempel sakte prosessering og forveksling av lyder, som en konsekvens av at utviklingen av fonemiske kategorier for språklydene er forsinket eller atypisk (Noordenbos & Serniclaes, 2015). Persepsjon på sensorisk nivå er antatt å være som hos typisk utviklede lesere, men kategoriseringen av språklydene kjennetegnes ved å i større grad fokusere på variasjoner innenfor hver fonemiske kategori (såkalte allofoner) enn på variasjoner mellom fonemene (Bogliotti, Serniclaes, Messaoud-Galusi, & Sprenger-Charolles, 2008). Allofonisk

kategorisering kan eksemplifiseres gjennom at det i talespråket forekommer ulike varianter av samme fonem på grunn av dialektvariasjoner. Begrepet betegner også varianter av samme fonem på mer subtilt nivå, for eksempel som en følge av koartikulasjon. Allofonisk kategorisering av språklyder er funksjonell som utgangspunkt for forståelse og uttale av talespråket og kjennetegner også den tidlige typiske språkutviklingen. I henhold til hypotesen er den perseptuelle kategoriseringen ved dysleksi a-typisk og kjennetegnet ved at den har stagnert på et allofonisk nivå. Dette får ikke innvirkning på generell språklig funksjon eller talespråk, men fører til problemer med å tilegne seg det grunnleggende fonemiske prinsippet for skriftspråk. Empirisk støtte til hypotesen finnes i resultater fra laboratorie-eksperimenter som har dokumentert a - typiske mønstre i kategorisering av lyder på subtilt nivå hos forsøkspersoner med dysleksi sammenliknet med typiske lesere (McBride-Chang, 1996; Serniclaes, Sprenger-Charolles, Carre, & Demonet, 2001).

Kategorisk talepersepsjon måles ved at forsøkspersoner lytter til en serie av innspilte enkeltlyder der Voice Onset Time (VOT) er manipulert slik at samme lyd framkommer med litt ulik «valør» langs et kontinuum. Målinger av kategorisering av språklyder har vært demonstrert med ulike stimuli relatert til relevante fonemiske kategorier i den aktuelle ortografien. Kontinuum basert på distinksjoner mellom stemte og ustemte versjoner av samme lyd (som regel b-p) har vært brukt i franske og engelske utvalg (Serniclaes et al., 2001; Snowling, Lervåg, Nash & Hulme, 2019) og lydmessige kontraster relatert til språklydens artikulasjonssted har vært brukt i nederlandske utvalg (Blomert, Mitterer, & Paffen, 2004). I Finland har også perseptuell kategorisering utfra språklydens varighet blitt målt med gjentatte lytteoppgaver av samme fonem i kort og lang utgave (f. eks. Pennala et al., 2010).

Flere studier der talepersepsjon måles gjennom lytteoppgaver basert på subtile kontraster mellom språklydene, konkluderer med sikre holdepunkter for korrelasjoner mellom a-typisk, det vil si ikke-fonemisk, kategorisering av språklydene og forekomst av dysleksi (se Noordenbos & Serniclaes (2015) for en metaanalytisk gjennomgang). Nyere studier har imidlertid nyansert forståelsen av en direkte kausal sammenheng mellom avvikende resultater knyttet til perseptuell kategorisering og lesing og det antydes at sammenhengen mellom lesing og perseptuell kategorisering delvis kan forklares via forhold knyttet til oppmerksomhet (Calcus, Lorenzi, Collet, Colin, & Kolinsky, 2016; Snowling et al 2019).

Noen studier har også funnet sterkere sammenhenger mellom a-typisk kategorisering og lesing i utvalg med forekomst av utviklingsmessige språkvansker (DLD) enn i utvalg med ren dysleksi, noe som kan indikere at årsaken til a-typisk kategorisering ikke primært er relatert til svake fonologiske ferdigheter, men er å finne i mer generelle språklige forhold (Robertson, Joannis, Desroches, & Ng, 2009; Snowling et al., 2019). Inkonsistente funn av denne typen kan delvis skyldes komorbide vansker, i og med at både ADHD og Utviklingsmessige språkvansker ofte er komorbide med dysleksi (Roberston et al 2009), men kan mulig også forklares med heterogene årsaker på subtilt sensorisk nivå i dyslektiske utvalg (Zhang & Mc Bride Chang 2010).

Det har også blitt reist innvendinger mot validiteten av denne typen eksperimentelle kategoriseringsoppgaver og blitt hevdet at mål på diskriminasjon av subtile kontraster mellom isolerte enkeltlyder ikke nødvendigvis er et relevant mål på ferdigheter knyttet til segmentering av fonemer i rekkefølge i naturlige lesesituasjoner (Boada & Pennington 2006). Blomert et al (2004) fant for eksempel at selv om barn med dysleksi hadde samme kategoriske grenser som kontrollelever langs et k/t kontinuum, avvek de fra typisk utviklede elever når det gjaldt i hvilken grad den fonologiske konteksten for enkeltfonemer spilte en rolle for diskriminasjon. Elevene med dysleksi ble i denne studien mer påvirket av forhold knyttet til koartikulasjon, det vil si at deres oppfattelse av den aktuelle lyden i større grad enn hos typisk utviklede lesere ble farget av hvilken lyd som kom før eller etter det aktuelle fonemet.

Forholdet mellom sensorisk sensitivitet, talepersepsjon og lesing er komplisert og, som vist, er resultatene fra de ulike studiene divergerende og delvis også påvirket av hvilket skriftspråk som undersøkes. McBride-Chang og kollegaer har i flere studier prøvd ut ulike modeller for sammenhengen mellom sensoriske forhold, persepsjon og lesing i flerspråklige utvalg. (McBride-Chang, 1996; McBride-Chang et al., 2008; Zhang & McBride-Chang, 2010; Zhang, McBride-Chang & Eccles 2014). Studiene omhandler hovedsakelig tospråklige utvalg fra Hong Kong (engelsk/kantonesisk), noe som gir unike muligheter til å undersøke ulike stier fra generelt auditivt nivå via talepersepsjon og fonologisk bevissthet til lesing i både alfabetisk og logografisk skriftkultur. Erfaring fra disse studiene indikerer at dysleksi delvis kan knyttes til vansker med å fange opp og oppfatte de perseptuelle nyansene som er relevante markører i det aktuelle skriftspråket. I mange alfabetiske skriftspråk vil dette først og fremst handle om nyanser på fonemisk nivå, f.eks. mellom stemte og ustemte versjoner av samme



lyd, i finsk kan det dreie seg om å diskriminere mellom lang og kort vokal (fonemisk lengde), mens det i kantonesisk kan være å oppfatte nyanser knyttet til prosodi og tonelag.

### 2. 4. 3 Forhold knyttet til taleproduksjon

Forekomst av artikulasjonsproblemer i relasjon til lesing er hyppig dokumentert i utvalg av barn med språklydsvansker i førskolealderen (Hayiou-Thomas, Carroll, Leavett, Hulme & Snowling, 2017; Preston, Hull, & Edwards 2013; Lewis, Avrich et al., 2011) og i generelle utvalg (Mann & Foy, 2007). Longitudinelle studier har vist at det spesielt er såkalt a-typiske språklydsvansker som har sammenheng med senere leseproblemer (Preston et al., *ibid*). Forekomst av språklydsvansker alene er imidlertid en usikker prediktor for dysleksi, men har større betydning dersom det forekommer sammen med andre risikofaktorer som svake generelle språkferdigheter og familiær risiko for dysleksi (Hayiou-Thomas et al., 2017; Hulme, Nash, Gooch, Lervåg, & Snowling, 2015; Lewis et al., 2011). Det synes som om språklydsvansker som er løst før den formelle leseopplæringen starter er av mindre betydning i et prediktivt perspektiv enn mer konsistente språklydsvansker ved skolestart (Hulme & Snowling, 2009). I tråd med den såkalte «illusory recovery» hypotesen (Scarborough & Dobrich 1990) har imidlertid studier av enkelt-case vist eksempler på at uttaleproblemer som var til stede i førskolealderen, men ble overvunnet i muntlig tale i løpet av tidlig skolealder, senere viste seg som stavefeil med samme typologiske mønster som de tidligere uttalevanskene (Snowling, Hulme, Wells, & Goulandris, 1992).

I en multifaktoriell forståelsesramme er problemer med taleproduksjon en latent risikofaktor for problemer i leseutviklingen og forekomst av språklydsvansker er også hyppig komorbid med dysleksi (Pennington & Bishop, 2009). På et mer subtilt nivå inngår taleproduksjon i de fonologiske representasjonene. Det har imidlertid vært omdiskutert hvorvidt a-typisk taleproduksjon reflekterer avvik i persepsjonsprosessen eller hvorvidt uttaleproblemene reflekterer et mer selvstendig problem (Hulme & Snowling 2009; Mann & Foy 2007).

I følge Hulme og Snowling (2009) kan de fonologiske vanskene ved dysleksi bedre defineres som problemer knyttet til taleproduksjon enn til problemer knyttet til talepersepsjon. Disse kan imidlertid være vanskelig å skille fra hverandre (*ibid*). I en studie med faktoranalytisk tilnærming, der et utvalg av førskolebarn (N= 175) ble undersøkt med et bredt spekter av ulike kognitive og språklige prøver, fant Anthony et al. (2010) at var det mulig å

skille ut tre faktorer som hver for seg hadde like høy ladning på begrepet fonologiske representasjoner. Disse tre faktorene; Taleproduksjon, talepersepsjon og tilgang til fonologiske enheter var imidlertid også interkorrelerte i modellen (ibid).

I henhold til motorisk teori om talepersepsjon (Liberman & Mattingly, 1985; Liberman, Shankweiler, & Liberman, 1989) er persepsjon av muntlig tale utgangspunktet for at det dannes fonologiske representasjoner. Dette sees imidlertid ikke som en isolert auditiv prosess, men som en sammensatt prosess der det foregår en perceptuell integrasjon av språklyden og den unike artikulatoriske gester som assosieres med denne språklyden. Språklydens fonemiske identitet kan dermed ikke skilles fra dens artikulatoriske identitet, de er to sider av samme sak.

Empirisk støtte til at observasjon av artikulatoriske gester bidrar i talepersepsjon er blant annet dokumentert i studier av voksne som skal lære et nytt språk. I en studie med spansktalende studenter fant Navarra & Soto-Faraco (2007) for eksempel at bilder av artikulatoriske gester assosiert med språklyder i det nye språket (katalansk) økte forsøkspersonens diskriminasjonsevne når det gjaldt ukjente lydkontraster. I en annen studie med voksne typisk utviklede lesere, fant Sato, Troille, Ménard, Cathiard, & Gracco (2013) at eksponering for bilder av munnbevegelser som ble presentert synkront med auditive signaler økte forsøkspersoners diskriminasjon av lydlike stavelser sammenliknet med oppsett som bare presenterte auditive signaler. I en ERP-studie i et mer klinisk utvalg med barn med fonologiske vansker (33 i risikoutvalg og 34 i kontrollutvalg) fant også Schaadt, Männel, Meer, Pannekamp, and Friederici (2016) at barn med fonologiske vansker tolket visuelle bilder av artikulatoriske gester tilknyttet språklydene på en annen måte enn barn uten fonologisk vansker, ved at bildene av gester hos disse førte til a-typisk aktivering assosiert med språklige sentre i hjernen. Hos elevene uten fonologiske vansker førte eksponering for bilder av artikulatoriske gester kun til aktivering av ikke språklige områder i hjernen. Studien viste også at tendensen til a-typisk aktivering av språklige områder var sterkest hos barn med mer alvorlige fonologiske vansker. Forfatterne konkluderer med at resultatene kan indikere at observasjon av artikulatoriske gester har en mulig kompensatorisk effekt i talepersepsjon hos barn med svake fonologiske ferdigheter.

Andre har vektlagt at artikulatoriske gester integrerer forholdet mellom talepersepsjon og taleproduksjon i det mentale leksikon (Ehri, 2014; Mc Bride Chang 1996; Sénéchal,

Ouellette, & Young, 2004; Thomas & Sénéchal, 2004). I følge Ehri (2014) har uttale av språklydene en sentral betydning som grunnlag både for utvikling av fonologiske representasjoner i fonemisk retning og som grunnlag for at det blir etablert skriftspråkrelevante mentale representasjoner på ortografisk nivå. Empirisk støtte til dette er blant annet vist i en randomisert studie som sammenliknet femteklassingers hukommelse for ukjente ords stavemåte etter tekstlesing (N=62)(Rosenthal og Ehri, 2011). Deltakerne var elever med svakt vokabular og studien viste at elevene som fikk beskjed om å uttale de ukjente ordene i teksten høyt mens de leste, husket ordenes stavemåte og betydningsinnhold signifikant bedre, både på kort og lang sikt, enn elevene i kontrollgruppen som kun markerte ordene med blyant.

Flere studier har vist at elever med dysleksi avviker fra kontrollelever når det gjelder artikulatoriske ferdigheter. Det gjelder både artikulatorisk bevissthet, ofte målt ved oppgaver der barnet skal reprodusere de nøyaktige artikulatoriske gestene til et oppgitt ord (Montgomery 1981, Griffiths & Frith 2002) og i oppgaver som måler artikulasjonshastighet, både når det gjelder artikulasjon av enkelte gester og sekvenser av gester (Fawcett & Nicholsson 2002; Duranovic & Sehic, 2013). Griffiths & Frith (2002) fant også at voksne med dysleksi hadde signifikant svakere artikulatorisk bevissthet enn kontrollpersoner uten dysleksi. Dette til tross for at leseproblemene i det dyslektiske utvalget delvis var overvunnet.

## **2.5 Artikulatorisk bevissthet**

Søkelys på artikulatorisk bevissthet i leseopplæringen kan begrunnes teoretisk på flere måter. En vinkling tar utgangspunkt i at artikulatoriske kjennetegn ved språklyden inngår som grunnlag for lagring av de fonologiske representasjonene i langtidsmindet (Ehri, 2014; Liberman & Mattingly 1985; McBride-Chang, 1996). I dette perspektivet antas det at aktivering av de artikulatoriske kjennetegnene ved lydene spiller en naturlig rolle når det gjelder prosessering av informasjon involvert i avkodingen. I et kompensatorisk perspektiv har det også vært antatt at artikulatorisk bevissthet kan fungere støttende/kompenserende for elever med svake fonologiske ferdigheter (f.eks. Wise et al., 1999, Torgesen et al 2001).

Studier av effekten av artikulatorisk bevissthetstrening har vært gjennomført både i utvalg med førskolebarn og i ulike utvalg med barn med forsinket leseutvikling. Studiene gir gjennomgående divergerende resultater. Noen studier finner at artikulatorisk trening gir samme effekt på utvikling av avkoding og fonologisk bevissthet som mer tradisjonelle

lyttebaserte innfallsvinkler, mens andre har funnet at artikulatorisk trening er spesielt gunstig for dem med svakest utgangspunkt. For gjennomgang av studier se artikkel 1 og 3.

### **2. 5. 1 Virkemidler i trening av artikulatorisk bevissthet**

Artikulatorisk støtte i leseopplæringen kan foregå på flere måter og med ulike typer materialer. Den aller enkleste formen innebærer at det ikke brukes noen spesielle materialer, men at det både i høytlesing og stillelesing legges vekt på tydelig formulering av lydene med sikte på både å lette tilgangen til de fonologiske kodene og bygge opp det fonologiske leksikon (Skjelfjord 1987a & 1987b b; Ehri 2014; Rosenthal & Ehri, 2011). En slik innfallsvinkel synes å ta utgangspunkt i forståelsen av at formulering av lydene er en naturlig del av prosessen med å lære å knytte språklydene til bokstavtegn (Lieberman, Shankweiler, & Lieberman, 1989). Artikulasjon eller lydformulering som naturlig mentalt redskap i lesing er også beskrevet av Skjelfjord (1987a og 1987b). Han observerte at barn i den første leseopplæringen spontant formulerte lydene inne i seg eller høyt som støtte i arbeid med segmenteringsoppgaver. Han observerte også at frekvensen av artikulasjon økte med oppgavens fonologiske kompleksitet (ibid).

Bruk av speil for å observere egne taleorganer er også et mye brukt virkemiddel for å opparbeide større bevissthet rundt de artikulatoriske egenskapene ved språklydene. Speil kan også supplere bruk av andre typer konkretiseringsmateriell (Studie 1&2; Castiglioni-Spalten & Ehri, 2003; Wolff, 2016).

Når det gjelder konkret undervisningsmateriell har de fleste studiene brukt kort med bilder/tegninger som viser munnens posisjon i ulike artikulasjonsstillinger. Slike kortbaserte systemer finnes i flere utgaver og har ulikt detaljnivå når det gjelder i hvilken grad delene ved det artikulatoriske apparatet er representert. Noen dekker alle aktuelle fonemer slik at de faktisk kan fungere som et selvstendig alfabet, mens andre kun dekker noen relevante lydmessige kontraster. Noen gjengir munnens posisjon helt naturalistisk (fotografi), mens andre viser tverrsnitt av taleapparatet og dermed også inkluderer de delene som ikke er umiddelbart synlige ved samtaler ansikt til ansikt. Felles for de fleste materialene er at de baseres på at barna kan gjenkjenne sitt eget taleapparat gjennom å se på bilder av andres munn. Materialene forutsetter dermed at barna mestrer å abstrahere fra et bilde eller tegning av en mer eller mindre naturalistisk munn til sitt eget taleapparat.

## 2. 6 Oppsummering

Dysleksi knyttes primært til svikt i den fonologiske komponenten i språket (Lyon et al., 2003). Det er også dokumentert at tiltak som tydeliggjør den fonologiske strukturen i ordene gjennom syntetiske leseopplæringsmetoder generelt gir best effekt for elever med dysleksi (Galuschka, Ise, Krick, & Schulte-Korne, 2014; SBU, 2014). I et multifaktorielt perspektiv er imidlertid både årsakene og symptomene til dysleksi sammensatte (Noordenbos & Serniclaes, 2015; Moll et al., 2013; Snowling & Melby-Lervåg, 2016). Dette øker behovet både for et videre spekter av innfallsvinkler når det gjelder tiltak for denne gruppen elever og individuell tilpasning av tiltak.

Taleproduksjon og talepersepsjon er tett sammenvevde prosesser (Hulme & Snowling, 2009; Liberman & Whalen, 2000) som begge bidrar til kvaliteter ved de fonologiske representasjonene (Anthony et al., 2010; Ramus et al., 2013). Den motoriske teorien om talepersepsjon (Liberman & Mattingly, 1985) har i særlig grad inspirert til utvikling av tiltak med fokus på trening i artikulatorisk bevissthet i leseopplæringen. Tidligere forskning på området har primært vært rettet mot yngre barn og gir generelt divergerende resultater.

### 3 Vurdering av kausalitet i intervensjonsstudier

Målet om å avdekke en eventuell kausal sammenheng mellom et tiltak og deltakernes atferd er utgangspunktet for alle intervensjonsstudier. Delvis handler det også om å trekke generaliserte slutninger om effekt slik at tiltaket kan anbefales for videre bruk i tilsvarende situasjoner og på tilsvarende type deltakere. I et praktisk pedagogisk perspektiv er det også viktig å kunne vurdere hvorvidt et igangsatt tiltak har effekt i den situasjonen det er igangsatt. I en norsk kontekst er dette ikke minst relevant for evaluering av spesialpedagogiske tiltak for enkeltelever.

#### 3.1 Hvordan vurdere hva som virker?

Spørsmål om hva som virker i praksis er sentralt i den såkalte «Hva virker» - bevegelsen (EBP Evidence Based Practice Movement) som internasjonalt, og også i Norge, har påvirket det utdanningsvitenskaplige forskningsfeltet siden 1990 tallet (Kvernbekk 2018). Et sentralt premiss for denne bevegelsen er at det legges til grunn at pedagogiske tiltak kan ha en generalisert effekt utover tid, rom og lokale forhold (Cartwright, 2013, Kvernbekk 2018). Dette kan eksemplifiseres ved at det i USA og også etter hvert i Europa har blitt etablert åpne databaser over tiltak «som virker». Den mest kjente av disse er What Works Clearinghouse (WWC) <https://ies.ed.gov/ncee/wwc/> som driftes av det amerikanske utdanningsdepartementet.

Hensikten med denne typen databaser er at aktører i praksisfeltet kan gå inn og hente informasjon om et etablert pedagogisk tiltak. Den åpne tilgangen gjør det også mulig å vurdere aktuelle effektstudiers «ranking» i henhold til gjeldende standarder for forskningskvalitet, før man eventuelt bestemmer seg for å iverksette tiltaket i egen kommune, skole eller klasserom.

Randomiserte kontrollerte studier regnes som mest troverdige når det gjelder å frambringe resultater som kan informere praksisfeltet om hva som virker og disse rangerer derfor høyest i EBPs evidenshierarki (WWC 2017; 2020). Gjennomføring av studier med bruk av RCT forutsetter imidlertid at man har relativt mange deltakere (Shadish, Cook, & Campbell, 2002). Dette gjør metoden mindre egnet når det gjelder forskning rettet mot marginaliserte grupper (Lobo, Moeyaert, Baraldi Cunha, & Babik, 2017).

Eksperimentelle Single Case design (Gast & Ledford, 2014; R. L. Tate et al., 2015; 2016) gir mulighet til å vurdere effekt av tiltak rettet mot enkeltelever eller små grupper av

elever og er blant annet brukt i effektstudier innenfor medisinsk rehabilitering (Krasny-Pacini & Evans, 2018). Det grunnleggende prinsippet for SCD er at hver deltaker tjener som sin egen kontroll ved at det foretas repeterte målinger i en baselinefase før tiltaket settes i verk og at resultatene fra baselinemålingene sammenliknes med resultatene etter at tiltaket er iverksatt. Innenfor spesialpedagogikk er SCD særlig brukt i forhold til marginaliserte vanskeområder som psykisk utviklingshemming og autisme.

### 3. 2 Effektbegrepet

Vurdering av effekt og kvantifisering av effektstørrelser er vesentlig i vurdering av resultatet i intervensjonsstudier (Shadish et al., 2002). Måling av effekt er også viktig for å kunne sammenlikne resultater på tvers av studier (Hedges, 2008). I gruppebaserte design knyttes effektbegrepet til statistiske analyser der både effektstørrelsen og sannsynligheten for at resultatet ikke skyldes tilfeldigheter (altså er signifikant) vektlegges (Shadish et al., 2002). I intervensjonsstudier med SCD har det vært mindre tradisjon for kvantifisering av effektstørrelser og vurderingen av effekt har tradisjonelt brukt en dikotom standard basert på visuell inspeksjon av rådata fra de repeterte målingene (Gast & Ledford, 2014).

Selv om det de siste årene har skjedd en endring i retning av mer bruk av statistiske metoder i SCD er det verdt å nevne at de historiske forskjellene i vurdering av effekt innenfor disse to forskningstradisjonene også reflekterer ulikheter i forståelsen av selve effektbegrepet fra et mer vitenskapsteoretisk perspektiv (Shadish, 2014). Dette kan eksemplifiseres ved at det i SCD også legges vekt på hvorvidt tiltaket *oppleves* som nyttig for den det gjelder og av nærstående personer rundt eleven (sosial validitet). En liten endring i atferd i positiv retning kan for eksempel oppleves som viktig og nyttig til tross for at effektstørrelsen er liten (Gast & Ledford, 2014; Onghena et al., 2018).

### 3. 3 Ytre validitet og generalisering

Generelt er målet for studier med SCD å prøve ut hvorvidt og under hvilke betingelser et tiltak virker for den enkelte deltaker. Siden målingene foregår kontinuerlig er det mulig å vurdere sammenhengen mellom tiltaket og deltakeres atferd på et detaljert nivå ved at målt atferd relateres til endringer i betingelsene gjennom kontrollert manipulering av tiltaket. Denne fleksibiliteten gir stor praktisk nytteverdi når det gjelder å tilpasse tiltak til enkeltdeltakeres behov, men fører samtidig til at SCD har lav ytre validitet (Onghena et al., 2018 ; Maggin, Cook, & Cook, 2018).

Generalisert effekt av SCD er avhengig av at resultatene repliseres. Indikasjon for replikasjon av tiltaket på nye deltakere knyttes generelt til sammenlikning på casenivå og det legges derfor stor vekt på nøyaktige casebeskrivelser og nøyaktig redegjørelse for tiltaket både når det gjelder materialer og gjennomføring (Tate et al., 2016).

Logikken når det gjelder replikasjon av effekt i SCD kan sammenliknes med det som gjelder for gruppebaserte design der den ytre validiteten styrkes ved bruk av metaanalyser (Shadish et al., 2002). I SCD angår replikasjonen både den indre validiteten ved at effekten må repliseres gjennom flere faser av studien og i et generalisert perspektiv ved at effekten må repliseres gjennom gjentatte (uavhengige)studier (Tate et al., 2016). En tommelfingerregel er at et tiltak vurdert med SCD kan regnes som evidensbasert dersom det er vist positiv effekt av samme tiltak/program i minst fem studier, med minst 20 deltakere til sammen utført av minst tre ulike forskerteam (3-5-20 regelen) (De Bruin, 2015; Gast & Ledford, 2014; Hitchcock, Kratochwill & Chezan 2015; Kratochwill et al., 2013; Maggin, Lane & Pustejovsky 2017; WWC 2017). Det stilles også krav til de enkeltstående studiene både når det gjelder det planlagte designet og når det gjelder kvaliteter ved de data som er samlet inn (artikkel 2; Kratochwill et al., 2013; Zimmerman et al., 2018)

### **3.4 «Hva virker» - bevegelsen og SCD**

Det har vært reist kritikk mot at til at standarder for vurdering av hva som virker primært har tatt utgangspunkt i kvalitetskrav innen gruppedesign metodologi og særlig konsekvensen dette har fått for metodeutviklingen i SCD i retning av økt fokus på generalisering (de Bruin, 2017; Ledford, Wolery & Gast, 2014; Maggin et al., 2017; Odom et al., 2005). Kritikere har også etterlyst større oppmerksomhet rundt fordelene med småskala forskning i databasene (de Bruin, 2017; Maggin et al., 2017).

Andre har svart på kravene til generalisering ved å utvikle teknikker og analysemetoder som likner de som brukes i gruppebaserte studier (Hedges, Pustejovsky, & Shadish, 2012; Pustejovsky, Hedges, & Shadish, 2014; Shadish et al., 2015). Siste årene har særlig utviklingen av flernivå regresjonsanalyser for SCD og utvikling av metoder som gjør det mulig å regne ut effektstørrelser på samme skala som for gruppe design (Cohens *d*) åpnet for at resultater fra SCD i prinsippet kan sammenliknes direkte med resultater for gruppedesign og også inngå i metaanalyser (Valentine, Tanner-Smith, Pustejovsky, & Lau, 2016).



Fra kritikere har det blitt advart mot for stor optimisme når det gjelder bruk av sofistikerte statistiske analysemetoder (Kratochwill & Levin, 2014). Imidlertid ser det ut til å være en økende bruk av statistiske analyser i SCD og fra 2020 er også utregning av både samlet deltakereffekt og bruk av designuavhengige effektmål innebygd i nylig publiserte standarder for SCD (WWC 2020).

### **3. 5 Predikert effekt og effekt for den enkelte**

«Hva virker» bevegelsen fokuserer som nevnt primært på å samle inn dokumentasjon om virkning av pedagogiske tiltak som grunnlag for å kunne velge tiltaket i nye kontekster. Vurderingen av effekt baseres i hovedsak på forskningsdesign som sammenlikner gjennomsnittseffekten mellom deltakere som har fått tiltaket og deltakere som ikke har fått tiltaket, fortrinnsvis i store utvalg (Shadish et al., 2002). Selv om konklusjoner om effekt fra denne typen design gir holdepunkter i et prediktivt perspektiv er det også feilkilder ved sammenlikning av gjennomsnitt (Onghena et al., 2018; de Bruin, 2017). Selv om effektmål som Cohens  $d$  tar høyde for individuell variasjon gjennom at effekten regnes ut med utgangspunkt i standardavviket i utvalget, vil vurdering på bakgrunn av den gjennomsnittlige effekten i et utvalg likevel kunne gi upresise vurderinger av effekten for hver enkelt deltaker (Onghena et al 2018). I spesialpedagogisk virksomhet er målsettingen ofte å utvikle tiltak som passer for enkeltelever. Selv om erfaringer basert på gjennomførte gruppestudier kan gi føringer for hva slags tiltak som vil være best egnet for enkeltelever vil spørsmål om hva som virker for den enkelte også måtte vurderes konkret i hvert enkelt tilfelle.

### **3. 6 Evaluering av spesialpedagogiske tiltak etter enkeltvedtak**

I en norsk kontekst er retten til spesialundervisning hjemlet i grunnskolelovens § 5. 1. Rettigheten har form av et juridisk enkeltvedtak og utløses etter faglig vurdering fra en ekstern kommunal instans (pedagogisk psykologisk tjeneste (PPT)). Den faglige vurderingen skal inneholde både en utredning av det spesialpedagogiske behovet og forslag til hvordan det spesialpedagogiske tiltaket skal utformes både med hensyn til målsetting, omfang, innhold og organisering. Videre knyttes enkeltvedtaket til lovpålagt årlig rapportering av hvorvidt målsettingen med tiltaket oppnås. Mye tyder imidlertid på at denne tiltakskjeden ikke fungerer optimalt og de siste årene har gjentatte offentlige utredninger stilt spørsmålstegn ved hvorvidt spesialundervisning virker etter intensjonene (artikkel 2).

### 3. 7 Oppsummering

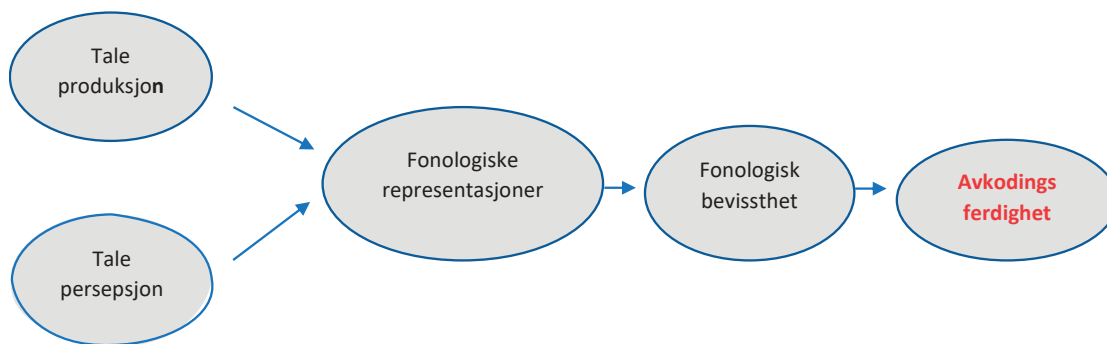
Ulike forskingsdesign er egnet til å besvare ulike problemstillinger i praksisfeltet. RCT og metaanalyser basert på denne typen design er best egnet når det gjelder å opplyse praksisfeltet om tiltak som har hatt en gjennomsnittlig effekt i liknende utvalg. SCD er fleksible når det gjelder tilpasning til praksisfeltet og er spesielt godt egnet til å vurdere effekt av konkrete tiltak for enkeltelever eller marginaliserte grupper. Denne typen design har lavere ytre og indre validitet enn randomiserte kontrollerte gruppestudier, men kan ha stor praktisk verdi (sosial validitet) når det gjelder å vurdere funksjonalitet av tiltak for den enkelte. Single Case design kan også bidra til forskningsbasert kunnskap gjennom at resultater fra enkeltstående design repliseres.

## 4 Teoretisk grunnlag for intervensjonene

Som diskutert i kapitel 2 er talepersepsjon og taleproduksjon tett sammenvevde prosesser som begge kan sies inngå som grunnlag for fonologiske representasjoner. En mulig kausal kjede for sammenhengen mellom disse er visualisert i figur 4.1.

**Figur 4.1**

*Mulig kausal modell for dysleksi med vekt på kvaliteter ved de fonologiske representasjonene*



I tråd med denne forståelsen prøves det i dette prosjektet ut en alternativ innfallsvinkel til trening av fonologisk bevissthet. Intervensjonen innebærer at det som erstatning/supplement til den auditivt baserte treningen som vanligvis brukes i leseopplæringen, også fokuseres på bevisstgjøring av de fonologiske enhetene gjennom tiltak rettet mot taleproduksjon. Konkret innebærer intervensjonen trening i å bruke artikulasjon som støtte i avkoding og staving.

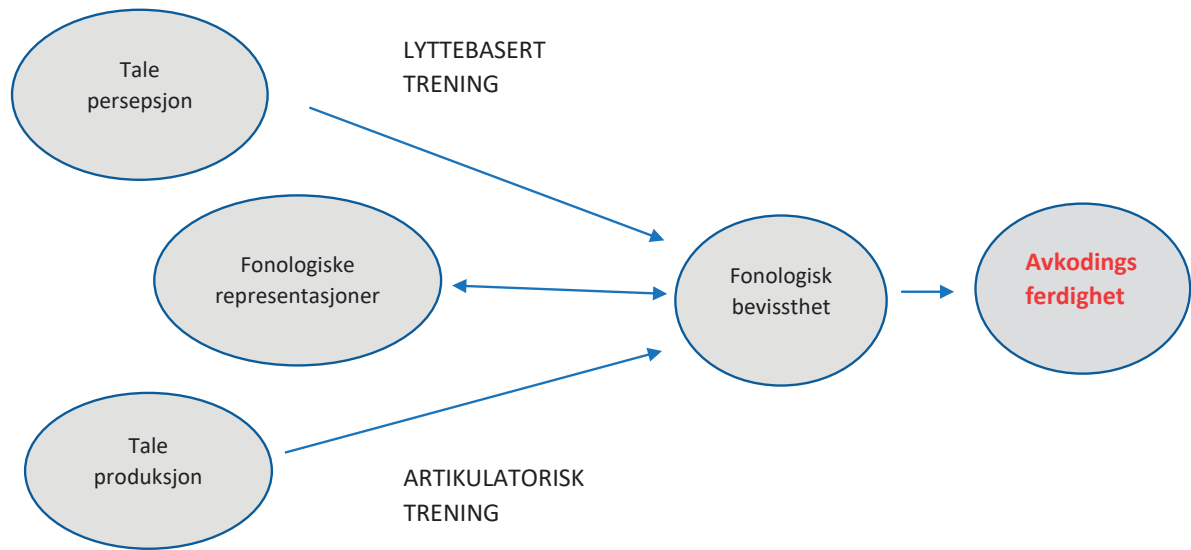
Hypotesen som ligger til grunn for treningen er at elever med svake fonologiske ferdigheter i mindre grad enn typisk utviklede lesere kan utnytte informasjon basert på en direkte kobling fra grafem til den fonologiske representasjonenes lydmessige identitet. For disse kan det antas at bevissthet om artikulatoriske kjennetegn ved lydene og aktivering av disse i møte med skriftegn kan lette tilgangen til de fonologiske representasjonene.

Konkret betyr det at dersom eleven ikke hører forskjell på b og p eller d og t, som i norsk er identiske lyder der forskjellene kun kan identifiseres på subtilt nivå utfra grad av stemthet, vil lyttebasert undervisning med sikte på automatisering av fonem/grafem forbindelsene ikke «treffe» fordi eleven ikke diskriminerer auditivt mellom lydene. I en slik situasjon kan det antas at artikulatorisk støtte vil kunne bidra til å tydeliggjøre det alfabetiske prinsippet. En slik mulig treningseffekt vil også kunne være relevant dersom man legger til grunn at de uspesifiserte fonologiske representasjoner delvis skyldes svak artikulatorisk bevissthet. I tillegg kan det antas at søkelys på artikulatoriske egenskaper ved språklydene i leseopplæringen kan bidra til utvikling av mer presise og distinkte fonologiske representasjoner, særlig for elever med auditive persepsjonsvansker (Elbro, Petersen, & Borstrom, 1998; Swan & Goswami, 1997).

Figur 4. 2 viser hypotesen som ligger til grunn for treningen i dette prosjektet i en forenklet versjon. I praksis er forholdet mellom de ulike faktorene i den kausale modellen mer sammensatt (f.eks. Rose, 2009). I dette prosjektet kan en mulig sti, som ikke er visualisert her, for eksempel være at taleproduksjon (og trening knyttet til dette) kan påvirke kjente prediktorer som bokstavkunnskap og benevningshastighet (Foy & Mann 2012, og dermed virke på avkodingsferdighet på flere (eller på andre måter) enn via fonologisk bevissthet. Siden verken benevningshastighet, verbalt minne eller morfologisk kunnskap blir målt i dette prosjektet innebærer det at prosjektet avgrenses til kun å omhandle en av flere teoretisk mulige veier til avkoding.

**Figur 4.2**

*Alternative innfallsvinkler til trening av fonologisk bevissthet*



## 5 Metodiske avveininger

I prosjektet brukes to ulike metodiske innfallsvinkler til vurdering av spørsmål rundt virkningen av to varianter av det samme pedagogiske tiltaket med samme målsetting. I dette kapitlet vil det blitt redegjort for metodiske avveininger og valg i de to gjennomførte studiene.

### 5.1 Utvalg

Studie 1 omhandler elever i første klasse med bekymringsfulle resultater på den obligatoriske nasjonale leseprøven. Utvalget i denne studien ble hentet fra databasen til skoleadministrasjonen i den aktuelle kommunen etter at leseprøven var gjennomført. Skoleadministrasjonen identifiserte først de elevene som var aktuelle. Konkret var kriteriet at elevene skulle ha skåret under kritisk grense på minst to av fire delprøver relatert til fonologiske ferdigheter. Disse var «Å finne lyder i ord», «Å trekke sammen lyder til ord», «Å stave ord» og «Å lese ord» (UDIR 2018). Deretter ble de skolene som hadde flere enn seks elever i denne kategorien valgt ut. Skoler med færre elever enn seks ble ekskludert da vi ønsket likeverdige kontroll og eksperiment grupper med mellom tre og fem deltakere. I alt 13 skoler fylte kriteriene, av disse var det en skole som hadde tilstrekkelig antall elever til at det ble laget fire grupper, altså to eksperimentgrupper og to kontrollgrupper. For skoler som hadde flere enn ti elever, men ikke nok til at det var aktuelt med to grupper ble det tilfeldig trukket hvilke elever som skulle delta. Det vil si at på hver skole er det tilfeldig hvilke av de elevene som fylte kriteriene som ble trukket til å delta i studien. Likeledes ble fordelingen av elever til enten intervensjon eller kontroll gruppe trukket tilfeldig. I begge disse prosessene ble det brukt et nettbasert program for PC ( <https://www.randomizer.org/> ). Det endelige utvalget består av 129 elever fra 13 skoler, fordelt i 14 eksperimentgrupper og 14 kontrollgrupper. For detaljer, se figur 5.1.

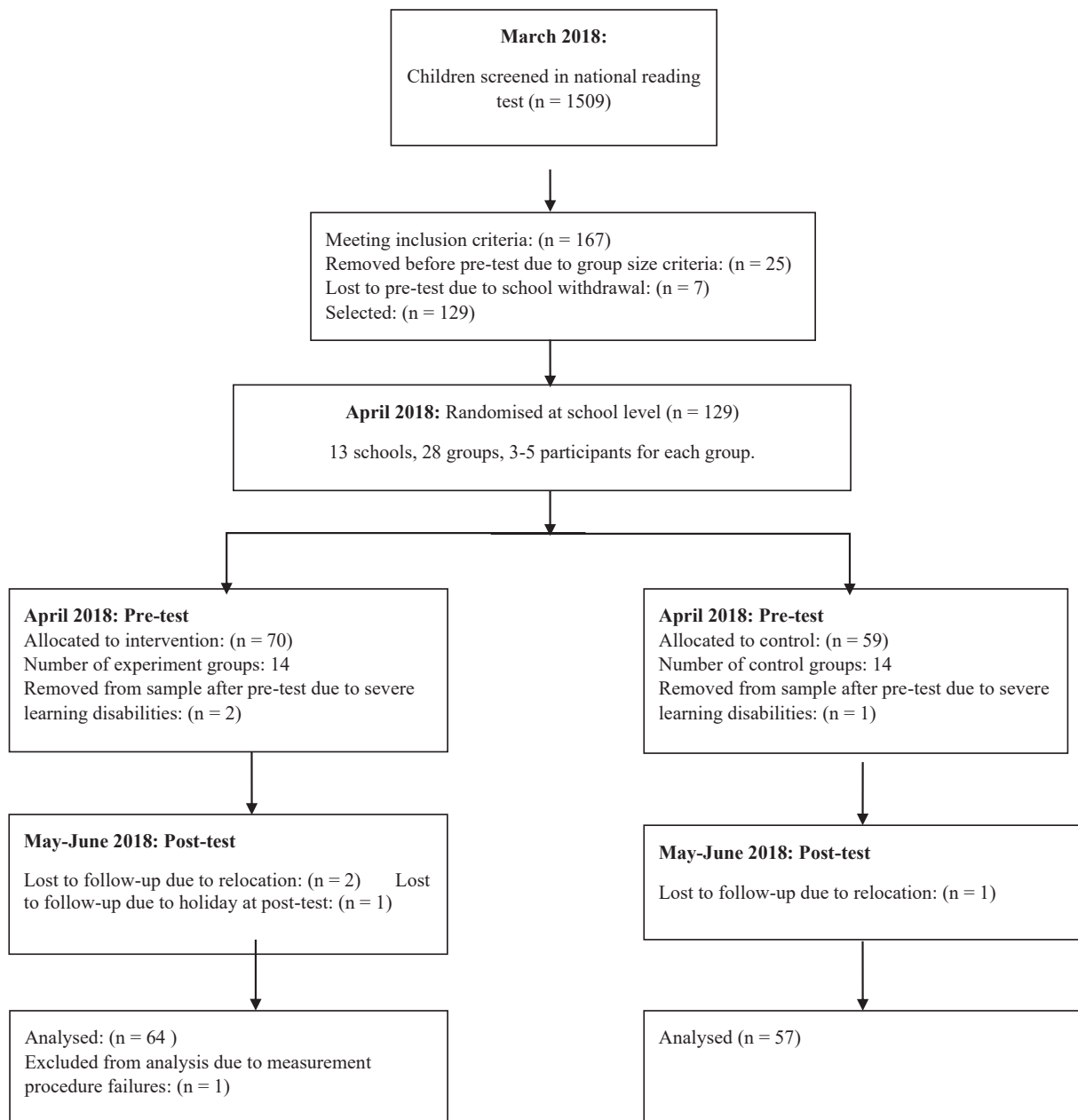
Studie 2 tar for seg virkningen av artikulatorisk trening i et utvalg av elever med alvorlig grad av dysleksi. Elevene ble rekruttert fra ventelisten i avdeling for språk og tale i Statped (nasjonalt spesialpedagogisk støttesystem).

For å realisere et utvalg bestående av elever med alvorlige fonologiske vansker ble inklusjonskriteriene for denne studien satt til resultater som tilsvarer – 2 standardavvik fra

aktuelt aldersgjennomsnitt på to non ord leseprøver fra STAS batteriet (Klinkenberg & Skaar 2001). Da det var ønskelig å vurdere tiltakets effekt på et utvalg med mest mulig «ren» dysleksi ble også elever med kjente komorbide vansker og redusert generelt evnenivå utelatt. For elevene som oppfylte disse primære kriteriene var det også nødvendig at elevenes hjemmeskoler var villige til å omdisponere ressurser for å realisere gjennomføring av tiltaket. For to elever som fylte de primære inklusjonskriteriene lot dette seg ikke gjøre. Utvalget består dermed av 13 elever som fylte begge inklusjonskriteriene.

En elev falt fra som følge av feil når det gjaldt målinger i baselinefasen og en elev benyttet seg av muligheten til å trekke seg og ble tatt ut etter eget valg. Øvrige elever er tatt med i analysene. Datagrunnlaget består dermed av 11 elever (figur 5. 2).

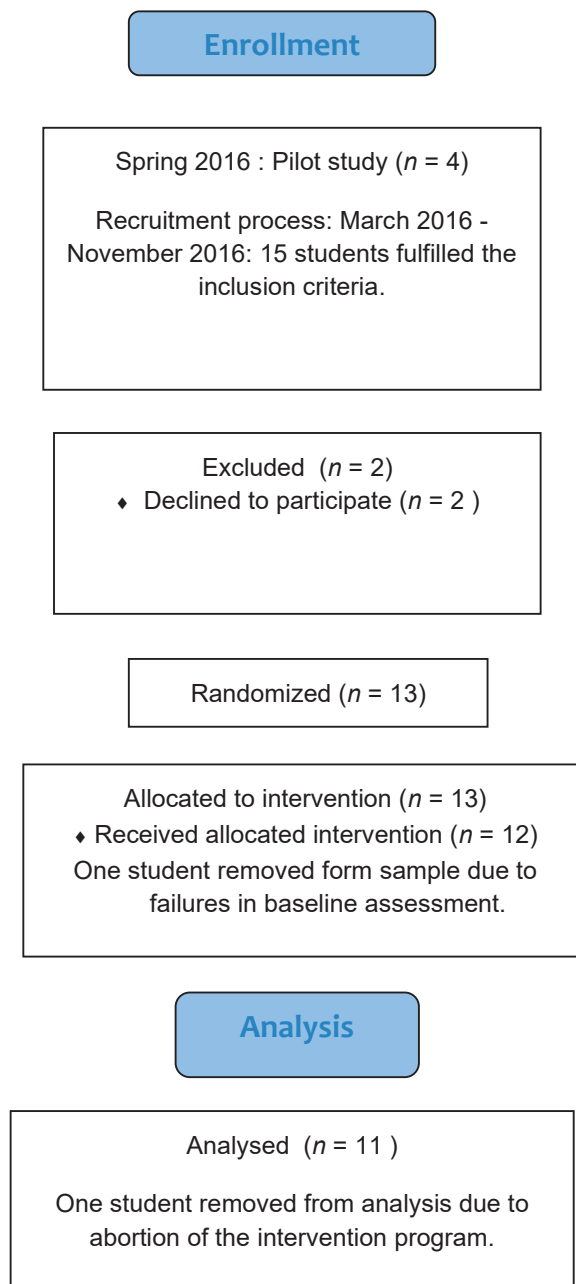
**Figur 5.1.**



Figur 5.1. Flytdiagram med oversikt over deltakere i studie 1.



**Figur 5.2.**



Figur 5.2. Oversikt over studie 2.

## 5. 2 Valg av design

*Studie 1:* I intervensjonsstudier regnes RCT design å være det sikreste alternativet når det gjelder å vurdere effekt av et tiltak fordi dette designet i større grad enn andre design kan utelukke alternative forklaringer (WWC 2017; Moher et al., 2010).

Det viktigste skillet mellom RCT design og andre eksperimentelle design ligger i randomiseringen. Når fordelingen av elever til intervensjon eller kontrollgruppe skjer tilfeldig vil påvirkning fra eksterne forhold også skje tilfeldig (Shadish Cook & Campbell 2002).

Generelt er det vanlig i denne typen design at utvalgene randomiseres etter kun ett kriterium (Deaton & Cartwright, 2017). Ideelt er det derfor ønskelig at alle andre faktorer enn de som angår intervensjonsprogrammet holdes konstante (Moher et al 2010). Siden den aktuelle studien (artikkel 1) foregikk i en naturlig situasjon er dette imidlertid vanskelig, for ikke å si umulig å realisere (Montgomery et al 2018). I dette tilfellet ble randomiseringen gjort på hver skole ved at de elevene som fylte inklusjonskriteriene ble tilfeldig trukket til å delta i enten eksperimentgruppe eller kontrollgruppe. Ved at randomiseringen ble gjort på hver skole sikres det at fordelingen til intervensjon- og kontrollgruppene ikke påvirkes av eventuelle ulikheter mellom skolene. For også å ta i betraktning at tilbudet til elevene i intervensjonsgruppene ikke vil være helt identisk på tvers av skoler og grupper, men kan variere noe mellom gruppene, ble det i analysene benyttet metoder tilpasset et såkalt «partial nested cluster design» (Lohr et.al, 2014; artikkel 1).

*Studie 2:* Målsettingen med studie 2 var tosidig. For det første: Siden deltakerne var henvist Statped med behov for utvikling av nye individuelt tilpassede tiltak var det primære målet å undersøke om det valgte tiltaket hadde effekt for hver enkelt elev. For å imøtekomme de individuelle behovene til de enkelte deltakerne ble det i etterkant av datainnsamlingen avholdt lokale møter med skoler, foreldre og pedagogisk psykologisk tjeneste der eventuelle endringer og justeringer av tiltaket ble diskutert.

For det andre var det også et mål å vurdere om artikulatork trening kunne ha en kompensatorisk effekt for elever med alvorlig grad av dysleksi i et mer generalisert perspektiv. Studien ble derfor designet med en standardisert prosedyre med flest mulig like elementer i den perioden datainnsamlingen foregikk. I lavfrekvente utvalg er det vanskelig å oppnå nok likeverdige deltakere til at det er relevant å benytte gruppebaserte forskningsdesign. I studie 2 ble det derfor valgt et eksperimentelt SCD (Tate et al., 2016).

SCD finnes i flere utgaver og formater (Gast & Ledford 2014) og varierer både med tanke på indre logikk og grad av eksperimentell kontroll (Tate et al 2016). Siden målet for studien var en varig endring av atferd var det nødvendig å bruke et design som ikke forutsatte reversering av tiltaket. I dette tilfellet ble det valgt å bruke et multiple baseline design (deltakere) (Baer, Wolf, & Risley, 1968). Videre ble det valgt en tilpasset versjon som også inkluderte elementer fra multiple probe design (Horner & Baer 1978). «Probevarianten» innebærer at de repeterte målingene ikke foregår kontinuerlig, men i planlagte bolker (probes) (Horner & Baer, 1978). Valget av probevarianten ble delvis gjort for å minske risiko for svekket validitet på grunn av testeffekt. Det ble antatt at lengre opphold mellom tidspunktene for kartleggingen både ville gjøre det mindre sannsynlig at elevene ville huske oppgavene og også redusere risikoen for at de ville gå lei. Det ble derfor valgt å organisere baselinemålingene i tråd med multiple probe strukturen. Likeledes ble denne strukturen valgt i postfasen av studien først og fremst fordi opphold mellom målingene ble antatt å kunne bidra til en sikrere vurdering når det gjaldt eventuell varig effekt av intervensjonen. I intervensjonsfasen ble det vurdert som verdifullt å følge progresjonen mer kontinuerlig gjennom ukentlige målinger. Som vist gjorde dette det også mulig å få et relativt presist mål på tidsaspektet med hensyn til elevenes progresjon når det gjaldt å automatisere bruk av PAS symbolene (artikkel 1 og 3).

Siden studie 2 ble gjennomført med deltakere fra forskjellige skoler kan designet karakteriseres som et såkalt «non-concurrent» multiple baseline design (Watson & Workman, 1981). Non-current design tillater at deltakerne kartlegges i forskjellige tidsperioder og er spesielt anvendelige for bruk i praktisk rådgivningsarbeid i skolen siden elever med samme behov ofte henvises på ulike tidspunkter (Watson & Workman, 1981; Winn et al., 2004). Non current multiple baseline design har av enkelte blitt hevdet å være mer sårbare for påvirkning fra uforutsette hendelser («historie») av betydning for utfallet enn det som gjelder for tradisjonelle (såkalt «concurrent») multiple baseline design (Carr, 2005; Gast, Lloyd & Ledford, 2014). Andre har imidlertid hevdet at sannsynligheten for at uforutsette hendelser av betydning for validiteten inntreffer gjentatte ganger i løpet av en intervensjonsfase i ulike miljøer generelt er lav (ibid.). Faktorer knyttet til antall målepunkter og antall deltakere hevdes også generelt å ha større betydning for den indre validiteten (Christ, 2007). Non-concurrent design er også mindre sårbare for såkalt «spillover» effekt av tiltaket til deltakere i ventefasen («diffusjon») (Winn, Skinner, Allin & Hawkins 2004).

I studie 2 var intervensjonsfasen like lang for alle deltakerne, men i henhold til multiple probe strukturen hadde elevene ulike individuelle mønstre med hensyn til hvor lenge baselinefasen og postfasen varte (Figur 5.3). Alle ble imidlertid kartlagt like mange ganger i alle faser. Målingene foregikk en gang pr uke. I henhold til den planlagte strukturen (figur 5.3) inngikk ni av deltakerne i parallelle «trappetrinnsdesign», der de ble kartlagt i samme tidsperiode – se figur 5.3. For disse ble aktuell «rute» og oppstart av intervensjonen i henhold til designet fordelt gjennom randomisering. De øvrige deltakerne (elev 11-14) ble rekruttert senere og fulgte samme «rute» som den første elevene i hvert «trappetrinnsdesign». I tråd med retningslinjene fra Tate et al. (2016) ble studien designet med fem målepunkter i baselinefasen, åtte målepunkter i intervensjonsfasen og fem målepunkter i postfasen. For en av elevene (elev 4) ble det kun gjennomført tre baseline målinger på grunn av sykdom hos testleder og elev og for elev 6 ble det kun gjennomført seks målinger i intervensjonsfasen av samme årsak.

For å styrke designet ble det også foretatt pre og post tester med delprøver fra den standardiserte lese og stavetesten STAS (Klinkenberg & Skaar 2001) (artikkel 3). Designet ble også forsøkt styrket ved at det ble lagt til et uavhengig kontrollmål (pegboard testen). Denne ble administrert ved samme tidspunkter som øvrige avhengige variabler og ble primært inkludert som kontroll for eventuell testeffekt (se nedenfor).

I et validitets perspektiv regnes ustabile data i baselinefasen som en trussel mot den indre validiteten i SCD og det anbefales generelt å fortsette målingene i baselinefasen inntil det er oppnådd stabilitet (f.eks. Kratochwill et al., 2013). I dette tilfellet var det imidlertid nødvendig med standardiserte prosedyrer både fordi de ukentlige målingene delvis ble foretatt av fagpersoner med begrenset kjennskap til designet og fordi ressursene til kartlegging av elevene måtte avtales og tilpasses i henhold til PPTs generelle arbeid.

**Figur 5.3**

*Kartleggingsplan*

ELEV/ UKE	39	40	41	42	43	44	45	46	47	48	49	50	51	52	1	2	3	4	5	6	7	8	9	1	11	12	13	14	15	16	17	18				
7																																				
8																																				
9																																				
10																																				

Figur 5.3 Oversikt over kartleggingsplan med ukenummer for fire elever med i alt 18 kartleggingstidspunkter pr elev. Markerte horisontale felt i oransje angir skoleferiene, henholdsvis uke 40, 51, 52 og 8. Ulike nyanser av grått angir de ulike fasene av studien. Lys grått = baselinefase, grått = intervensjonsfase, mørk grått = postfase.

### 5.3 Aktuelt intervensjonsprogram

Undervisningsmaterialet som er brukt i dette prosjektet avviker på noen måter fra det som er brukt i tidligere studier av artikulatorisk bevissthetstrening. Som nevnt har tidligere studier brukt bilder av munn i ulike posisjoner. PAS (Piktografisk Artikulatorisk System) bruker piktografiske symboler. PAS ble utviklet som en «fonologisk versjon» av BLISS systemet, som er et hjelpemiddel for alternativ kommunikasjon beregnet på mennesker med store språk og talevansker (Kausrud, 2003).

PAS inneholder piktografiske elementer som viser til ulike deler av taleapparatet (figur 5. 4). Disse elementene settes så sammen utfra språklydenes artikulatoriske og akustiske egenskaper og hvert PAS symbol utgjør dermed en presis artikulatorisk «oppskrift» på språklyden. Ved å følge oppskriften nøyaktig produseres lyden. I artikulatorisk bevissthetstrening brukes symbolene også som utgangspunkt for samtale om lydenes egenskaper og for å konkretisere kontraster mellom lydene. PAS symbolene kan også brukes til å skrive og lese enkle ord og en PAS font kan lastes ned via <https://www.statped.no/laringsressurs/sprak-og-tale/PAS--laremidde-for-utvikling-av-bevissthet-om-spraklyder/>

**Figur 5. 4** PAS symboler



Figur 5. 4 Eksempel på PAS symboler i staving av ordet «ROSE» [² ru:sø ]. Blå symboler (konsonanter) med flere elementer som viser til akustiske egenskaper og artikulasjonssted ved uttale av lyden, rød symboler (vokaler) viser til form og størrelse på munnens åpning ved uttale av lyden.

I dette prosjektet er det utviklet to undervisningsopplegg, et for smågruppeundervisning (studie 1) og et for individuell opplæring (studie 2). Mange av elementene i programmene er like, men tilpasninger ble gjort særlig med tanke på elevenes alder.

I studie 1 ble det hovedsakelig brukt en lekbasert innfallsvinkel blant annet ved å benytte kjente prinsipper fra Lotto og Memory spill. I studie 2 var rammen rundt lese- og skriveaktivitetene inspirert fra metoden Helhetslesing (Frost, 2003; Godøy & Monsrud, 2011). Helhetslesing er en pedagogisk ramme for dialogisk basert leseundervisning der det arbeides med utvalgte og tilpassede tekster eller tekstutdrag. Helhetslesing rettes primært mot utvikling av fonologiske strategier i lesing og ser leseutvikling som en stadiebasert utviklingsprosess (Ehri 1995; Frost, 2003; Spear-Swerling & Sternberg, 1996). Målet er å øke elevenes metaspråklige kompetanse og gjennom samhandling og dialog i læringsprosessen overføre hensiktsmessige strategier for ordavkodning. Programmet som er utviklet i dette prosjektet bruker samme pedagogiske ramme som i Helhetslesing, men her ble PAS brukt som utgangspunkt for dialog og som middel til å tydeliggjøre de aktuelle avkodings- og stavestrategiene. PAS kortene ble også brukt for å demonstrere kontraster mellom lydene med sikte på å unngå systematiske forvekslingsfeil som var hyppig forekommende hos et flertall av elevene i studie 2.

#### **5. 4 Prosedyrer for datainnsamling**

*Studie 1:* I studie en ble alle elevene testet to ganger. Pretestene ble foretatt uken før tiltaket startet opp og testingen foregikk ved besøk på skolene. Deltakerne på hver skole var organisert med individuelle ID nummer og rekkefølgen var tilfeldig bestemt slik at det ikke framgikk hvorvidt eleven hørte til eksperimentgruppe eller kontrollgruppe. Testingen ble utført av undertegnede, supplert med kollegaer fra Statped og timebetalte forskningsassistenter som var rekruttert fra universitetet. Kollegaene og forskningsassistentene fikk opplæring i små grupper (max tre) i forkant av datainnsamlingen. Testingen tok i gjennomsnitt 30 minutter pr elev. Post test ble gjennomført i uken etter at intervensjonene var ferdige og fulgte samme prosedyre. Også her var elevenes status i kontroll eller eksperimentgruppe blindet fram til siste deltest som omhandlet kjennskap til de artikulatoriske symbolene (PAS) som kun var brukt i intervensjonsgruppene (artikkel 1).

*Studie 2:* I studie to ble alle elevene kartlagt 18 ganger. Datainnsamlingen ble gjennomført enten av undertegnede eller av ansatte i den pedagogisk psykologiske

rådgivningstjenesten (PPT) i aktuell kommune. Disse fikk opplæring i prosedyrene for datainnsamlingen i forkant av oppstart. Opplæringen ble hovedsakelig gitt individuelt. Gjennomføringen av de repeterte målingene fulgte den forhåndsoppsatte planen for studien (figur 5. 3.), men konkret tidspunkt for de ukentlige målingene ble bestemt lokalt. Hver kartlegging tok i gjennomsnitt 20 minutter.

## 5. 5 Ytre validitet

Ytre validitet angår i hvilken grad resultater fra de gjennomførte studiene kan generaliseres til også å gjelde populasjonen utvalgene er hentet fra (Shadish et al., 2002). I dette tilfellet vil det si elever i første klasse som skåret under kritisk grense på den nasjonale leseprøven (studie 1) og elever mellom 10 og 15 år med alvorlig grad av dysleksi (studie 2).

I studie 1 ble utvalget hentet fra en populasjon på 1509 elever i første klasse som gjennomførte den nasjonale prøven i lesing i en utvalgt kommune. Det aktuelle utvalget i studie 1 ble tilfeldig trukket fra den andelen av elevene som hadde skåret under kritisk grense på relevante delprøver knyttet til fonologisk bevissthet og fonologisk avkoding.

Det må antas at utvalget er representativt for denne populasjonen i aktuell kommune, mens mulig generaliseringseffekt av resultatene til en større populasjon av elever på første trinn i landet for øvrig er mer usikkert. Dette har blant annet sammenheng med at designet for studien lot kontrollgruppene motta «business as usual», det vil si ordinær tilpasset undervisning. Den eventuelle overføringseffekten til andre kommuner vil derfor også være avhengig av i hvilken grad den undervisningen som ble gitt i kontrollgruppene kan regnes som representativ for ordinær tilrettelegging i andre kommuner. Siden det både nasjonalt, på kommunenivå og i det enkelte klasserom gis rom for stor variasjon i tilretteleggingen og stor grad av metodefrihet for lærere i den begynnende leseopplæringen er det sannsynlig at «business as usual» ikke er det samme overalt.

I studie to ble utvalget hentet fra ventelistene til Statped i et gitt tidsrom på ca. 8 måneder. Alle henviste elever som fylte kriteriene for deltakelse i løpet av dette tidsrommet fikk tilbud om å delta i studien. Siden deltakelse i studien også forutsatte at elevenes hjemmeskoler måtte oppfylle de gitte kravene til omlegging av spesialundervisningen for å realisere intervensjonen, er utvalget ikke tilfeldig trukket. SCD har generelt lav ytre validitet og eventuell generalisert effekt er avhengig av at funn repliseres (se avsnitt 3.3.). Denne studien har et større antall deltakere enn det som er vanlig for SCD (Shadish & Sullivan,



2011). Det er likevel vanskelig å generalisere resultatene til andre elever med alvorlig dysleksi og resultatene er derfor i hovedsak å regne som eksempler fra praksis (artikkel 3).

## **5. 6 Indre Validitet**

Begge studier er gjennomført i naturlige settinger, det vil si på skolen. Dette er forskningsmessig utfordrende fordi man ikke har mulighet til å oppnå fullstendig eksperimentell kontroll med eksterne faktorer som potensielt kan føre til endringer hos dem som undersøkes. Slike forhold er umulig å ha kontroll med når man arbeider i en praktisk kontekst, men det er mulig å iverksette tiltak for å sikre best mulig kontroll.

I det følgende vil jeg gå inn på generelle og designspesifikke trusler mot indre validitet i de to gjennomførte studiene og redegjøre for hvordan disse truslene er forsøkt møtt i planlegging og gjennomføring av studiene.

### **5. 6. 1 Seleksjon**

Denne trusselen angår primært gruppebaserte design og knyttes til hvorvidt deltakerne som sammenliknes med hverandre også er ulike ved oppstart av eksperimentet slik at den eventuelle ulikheten mellom gruppene man ser etter intervensjonen ikke nødvendigvis knyttes til effekt av tiltaket, men også kan reflektere eksisterende ulikheter allerede før tiltaket startet. Den mest anerkjente metoden for å unngå denne typen feil i intervensjonsstudier er å randomisere deltakere til hver gruppe. Det er som nevnt gjort i studie 1, der elevene som skåret innenfor inklusjonskriteriene på den nasjonale leseprøven ble tilfeldig fordelt til enten eksperimentgruppe eller kontroll gruppe på hver skole.

I studie to ble elevene rekruttert fra ventelistene til Statped utfra fastsatte kriterier (se over). De representerer derfor ikke et tilfeldig og randomisert utvalg.

### **5. 6. 2 Frafall**

Frafall av deltakere fra pre til posttest svekker validitet av studien. Dette særlig dersom frafallet ikke skyldes tilfeldigheter. I begge de aktuelle studiene i dette prosjektet har det skjedd frafall av deltakere fra pre til post test (baselinefase til postfase i studie 2). I forbindelse med studie 1 ble det lagt stor vekt på å sørge for at så mange som mulig av elevene som hadde deltatt i studien også gjennomgikk posttest. Elever som var borte fra skolen på den aktuelle testdagen fikk derfor nye muligheter ved at det ble foretatt gjentatte besøk på skolene. Noen elever var imidlertid ikke mulig å fange opp, enten fordi de flyttet fra kommunen i løpet av intervensjonsperioden eller fordi de var på langvarig ferie ved posttest. I

studie 2 var det en deltaker som falt fra på grunn av feil ved målingene i baselinefasen, en annen avbrøt intervensjonen før de planlagte åtte ukene var gått. I tillegg var det en av deltakerne som byttet bosted i løpet av prosjektperioden. Denne eleven var under barnevernets omsorg og ble flyttet til nytt varig bosted hos en ny fosterfamilie. Han fortsatte imidlertid på gammel skole for å fullføre den planlagte intervensjonsperioden og etterkartleggingen knyttet til dette prosjektet. Likevel kan det antas at hans forutsetninger for utbytte av intervensjonen ble svekket og resultatene forstyrret av endringer i hjemmesituasjonen. Denne antakelsen bekreftes også av lærers evaluering.

Denne typen trusler mot validiteten betegnes ofte som historie og omhandler det faktum at uforutsette hendelser som man ikke kan kontrollere kan påvirke resultatet av et eksperiment (Shadish, et al., 2002; Gast 2014).

### **5. 6. 3 Modning**

Andre aktuelle trusler mot indre validitet er knyttet til at det uavhengig av intervensjonen kan skje en modning av elevens ferdigheter i løpet av intervensjonsperioden som kan påvirke resultatet av intervensjonen. Denne typen trusler er spesielt relevant dersom deltakerne er unge barn i utvikling eller dersom intervensjonen varer i over 4 til 6 måneder (Gast 2014). I studie 1 var elevene i en utviklingsfase når det gjaldt de ferdighetene som ble målt før og etter intervensjonen. Siden elevene fra hver skole som nevnt ble randomisert i enten eksperiment gruppe eller kontrollgruppe sikrer dette at en eventuell modning av ferdigheter i perioden ikke er systematisk forskjellige mellom elevene i de to gruppene.

I studie to var elevene eldre og hadde i henhold til rapporter fra skolene en lang historie med liten utvikling av lese- og staveferdigheter. Sannsynligheten for en spontan modning i tidsrommet for utprøvingen av tiltaket i dette prosjektet er dermed mindre sannsynlig. En eventuell modning av ferdigheter er i studie 2 også tatt i betraktning gjennom at det ble foretatt fem repeterte målinger av alle avhengige variabler før intervensjonen startet (baseline).

### **5. 6. 4 Testeffekt**

Testeffekt er i prinsippet en utfordring for alle studier med repeterte målinger. Problemstillingen er likevel i mindre grad aktuell for studie 1 der elevene bare ble testet to ganger og der en eventuell testeffekt ikke ville påvirket sammenlikningen mellom

eksperiment- og kontrollgruppene på grunn av randomiseringen. I studie 2 der alle elevene ble testet 18 ganger er dette imidlertid en aktuell problemstilling (Gast 2014).

Repeterte målinger representerer en trussel mot den indre validiteten av primært to årsaker assosiert til å enten overestimere eller underestimere effekt av tiltaket. For det første risiko for at deltakerne lærer seg oppgavene (praksiseffekt) og for det andre risiko for at de går trett av å utføre samme oppgave gjentatte ganger og derfor mister motivasjon for å prestere. I studie 2 ble risikoen for testeffekt forsøkt redusert og kontrollert på flere måter.

Multiple probe design minimerer antallet måletidspunkter uten å svekke validiteten av de repeterte målingene (Horner & Baer, 1978) og er derfor antatt å redusere risiko for testeffekt. Dette ble valgt i baseline og postfasen av studien. Det ble videre laget et spørreskjema til testleder som vurderte elevens motivasjon i forhold til testingen. Lærerne krysset her av for hvorvidt eleven verbalt hadde gitt uttrykk for motvilje mot å være med ut og hvorvidt de hadde nektet å bli med. Positiv innstilling til testingen ble rapportert på samme måte. For enkelte deltakere ble det ved enkelte anledninger rapportert verbale innsigelser mot å forlate klasserommet for å bli testet. Dette ble av testleder knyttet til anledninger der klassen holdt på med felles sosiale aktiviteter. Generelt ble det ikke rapportert uvilje mot å bli testet gjentatte ganger og for noen elever ble det også rapportert en uttrykt positiv holdning. Det antas på bakgrunn av dette at elevenes skårer i begrenset grad ble påvirket av at de gikk lei av å bli kartlagt. For å redusere risiko for praksiseffekt av de repeterte målingene ble det utviklet unike, men likeverdige ordlister for stave- og leseoppgavene. Disse alternerte fra gang til gang etter en randomisert rekkefølge for hver deltaker (se avsnitt 5.7.2.).

Både valget av probe løsningen for kartlegging og alternering av rekkefølgen av testord i ordlistene må antas å ha redusert sjansen for at elevene skulle huske oppgavene. I tillegg til dette ble det valgt å også inkludere et eksternt kontrollmål, en såkalt ikke-ekvivalent avhengig variabel (Shadish et al., 2002). Det ble valgt en finmotorisk prøve med tidsbegrensning (pegboard-testen). Målet med pegboard-testen var å styrke validiteten av lese- og stavetestene ved å kontrollere for testeffekt av de repeterte målingene med et nøytralt måleverktøy som ikke var relatert til treningen elevene fikk verken i intervensjonen eller i andre skolefag. Det var også et poeng at testen skulle være enkel å gjennomføre og ikke ta lang tid.

Materialet i pegboard-testen besto av et perforert Brett med hull og stifter, såkalt stiftmosaikk. <https://www.trignonor.no/leker/byggesett/plastbyggesett-og->

stiftmosaikk/stiftmosaikk-250-stifter-4097-p0000007906. Materialet inngår ofte i basen av formingsmateriell og leker i barnehager og aktivitetsskoler og er, ifølge produsenten, beregnet på barn fra 3 år og oppover. Alle testledere fikk utdelt en pose med ensfargede stifter og et rektangulær perforert brett (154 hull) med dimensjonen 19 x 24 cm.

Ved hvert testpunkt ble brettet plassert foran eleven med kortsiden opp og stiftene ble lagt vedsiden av brettets langside. For høyrehendte ble stiftene lagt på den høyre siden og for venstrehendte ble stiftene lagt på venstre side av brettet. Oppgaven var å plassere stiftene på brettet i en vertikal rekke så fort som mulig med en tidsfrist på 30 sekunder. Ingen ytterligere instruksjoner ble gitt. For hvert testtidspunkt ble denne testen administrert to ganger.

I nevropsykologiske utredninger brukes Pegboard-tester som et mål på psykomotorisk tempo, koordineringsevne og hånddominans. Pegboard-tester brukes også for å kunne vurdere endringer hos individer med ulike psykiatriske og nevrologiske sykdommer som for eksempel HIV, schizofreni, multiple sklerose og afasi. Et sentralt tema er derfor hvorvidt denne typen målinger er stabile fra testtidspunkt til testtidspunkt eller hvorvidt det forekommer treningseffekter ved gjentatt testing. Logikken er at dersom man antar at treningseffekten er 0 og at resultatet av de repeterte målingene dermed er 100% stabile over tid vil alle endringer kunne tilskrives andre faktorer, det være seg endret kognitiv status som følge av en progredierende sykdom eller som en effekt av trening for eksempel i forbindelse med intervensjoner.

Undersøkelser av test-retest reliabiliteten til de klinisk mest brukte pegboard-testene, Purdue pegboard-test og Grooved pegboard-test viser at en viss treningseffekt som følge av repeterte mål forekommer ved bruk av oppgaver av denne typen. (Barr, 2002; McCaffrey, Ortega, & Haase, 1993). Enkelte studier antyder også den generelle treningseffekten av motoriske tester er større enn for innlærte ferdigheter som for eksempel ordlesing (Levine, Miller, Becker, Selnes, & Cohen, 2004). Selv om pegboard-materialet brukt i dette prosjektet ikke er identisk med materialet i de standardiserte utgavene må det antas at de erfaringene som er gjort med de nevnte standardtypene delvis også vil være aktuelle for testen som ble brukt her.

I studie to viste målingene at både pegboard-testen og de øvrige avhengige variablene hadde en viss progresjon i løpet av baselinefasen, det vil si i løpet av de fem første målingene. Dette indikerer noe testeffekt. For pegboard-testen var den prosentvise økningen i skåre størst

i starten av målingene (baselinefasen) og mindre i intervensjonsfasen og postfasen. For de øvrige avhengige variablene var den prosentvise økningen større i intervensjons og postfasen. Imidlertid ble det oppnådd statistisk signifikant effekt på pegboard-testen på tvers av effektmål (artikkel 3).

Selv om artikkel 3 konkluderer med at resultatene fra pegboard-testen svekker grunnlaget for konklusjonene når det gjelder vurdering av effekten på de øvrige avhengige variablene, indikerer resultatene samtidig at fem målinger i baselinefasen trolig fanget opp den største delen av progresjonen som følge av testeffekt. Dette kan tyde på at testeffekt som følge av repeterte målinger er størst ved de første målingene og så flater ut. Artikkel 3 konkluderer derfor også med at testeffekt trolig ikke forklarer hele endringen fra baseline til intervensjon/post fasen for variablene knyttet til lesing og staving.

Inklusjon av en ikke ekvivalent avhengig variabel for å kontrollere for testeffekt er, så vidt kjent, ikke tidligere brukt i SCD. Tidligere studier har brukt kontrollmål av samme type som de som inngår i treningen, for eksempel ved at lister med utrente ord fungerer som kontrollmål for effekten på trente ord (Krasny-Pacini & Evans, 2018). Problemer knyttet til bruk av repeterte målinger og testeffekt vil bli nærmere diskutert i avslutningsdelen i denne avhandlingen (kapitel 7).

## **5. 7 Begrepsvaliditet**

En viktig forutsetning for å trekke konklusjoner etter eksperimentelle intervensjonsstudier er at det som faktisk måles har sammenheng med det man ønsker å vurdere. Begrepsvaliditet angår i hvilken grad de teoretiske fenomenene man undersøker er operasjonalisert på en måte som dekker innholdet i begrepene (Shadish et al., 2002). I dette prosjektet handler det konkret om hvorvidt de målemetodene og måleverktøyene som er brukt for å vurdere leseferdighet, fonologisk bevissthet, bokstavkunnskap, staving og artikulatorkunnskap faktisk måler disse ferdighetene. Begrepsvaliditeten angår også hvorvidt utvalget av deltakere er operasjonalisert i samsvar med problemstillingene som undersøkes (se avsnitt om utvalg over), i hvilken grad den praktiske gjennomføringen av intervensjonen er utført i samsvar med det som er planlagt og hvorvidt testprosedyrene følger retningslinjene for hvordan de operasjonaliserte variablene måles (ibid.).

### 5. 7. 1 Operasjonalisering av begreper

For å øke begrepsvaliditeten ble det i konstruksjon av måleverktøyene brukt samme type oppgaver som det som vanligvis brukes til måling av denne typen ferdigheter. Dette innebærer derfor en anerkjent operasjonalisering av begrepene. I begge studier har det også blitt lagt vekt på å måle samme teoretiske konstrukt på flere måter.

I studie 1 innebærer det at fonologisk bevissthet ble målt med tre sett av oppgaver som er antatt å reflektere ulike sider ved begrepet. Dette var henholdsvis identifikasjon av første lyd, fonemreduksjon og såkalt «spoonerism», det vil si oppgaver der barnet skal lage nye ord ved å bytte ut første lyd i målordet (f.eks. hva blir kake med b først?). Disse tre settene (12 oppgaver i hvert sett) ble gjennomført med utgangspunkt i tre kategorier med ord. Dette var henholdsvis ord som ble aktivt brukt i intervensjonen (trente ord), utrente ord og pseudo ord (totalt 36 oppgaver).

Også leseferdighet ble vurdert fra ulike vinkler. Her ble det brukt to prøver for lesing av ordinære ord; en liste med trente ord og en med utrente ord og to non ord lister. For non ordene ble det brukt to oversatte, men ikke standardiserte versjoner av TOWRE (Test Of Word Reading Efficiency) (Torgesen, Wagner & Rashotte 1999). I tillegg ble leseferdighet også målt gjennom en egenprodusert test som målte lesenøyaktighet.

Også i studie 2 ble det lagt vekt på å måle leseferdighet fra uke til uke på en måte som sikret validiteten. Dette ble lagt til grunn både i forbindelse med utviklingen av ordlistene (se nedenfor) og ved at alle målinger ble repetert to ganger ved hvert måletidspunkt.

Staving i dette prosjektet er avgrenset og operasjonalisert til staving av pseudoord med ulik vanskegrad. Staveferdighet ble bare målt i studie 2.

Artikulatorisk bevissthet ble i begge studiene målt med den samme transposisjonstesten. Begrepsvaliditeten av dette verktøyet er blant annet knyttet til hvorvidt innholdet i oppgavene som ble brukt er representativt for begrepet artikulatorisk bevissthet. Testen vurderer i hvilken grad symbolene transponeres til riktige alfabetiske skrifttegn. Skåringen av testen er todelt, slik at elevene fikk poeng både for korrekt transponert ord og dessuten ble kreditert for antall symboler som ble riktig transponert, selv om ikke ordet var fullstendig stavet eller eventuelt ble feilstavet. Transposisjonstesten er ikke et eksakt mål på hvor mange av symbolene som faktisk er lært, siden noen symboler var oftere representert

enn andre, men kan forstås som et mer generelt mål på effektivitet i å tolke de artikulatoriske symbolene i ulike fonologiske kontekster

I et validitetsperspektiv er det klart at ulike andre faktorer enn de som dekkes av begrepet artikulatorisk bevissthet kan ha spilt inn på resultatene. For det første forutsetter oppgavene at elevene er kjent med de alfabetiske skrifttegnene. Denne faktoren er spesielt aktuell for studie 1 der elevene som deltok var elever på første trinn under kritisk grense på den nasjonale prøven i lesing. Selv om posttesten i denne studien ble foretatt helt i slutten av første skoleår kan det ikke utelukkes at usikkerhet knyttet til bokstavkunnskap kan ha virket inn. Imidlertid viste måling av bokstav kunnskap ved post test i samme studie nærmest takeffekt når bokstavene ble presentert isolert (artikkel 1).

Det som imidlertid ikke ble kontrollert for og som kan ha spilt en rolle er elevenes skrivemotoriske ferdigheter. Siden testen hadde en tidsbegrensning er det mulig at noen av elevene kan ha brukt mye tid på bokstavutformingen og at dette kan ha påvirket resultatet. Sett i ettertid ville en transposisjonstest basert på høytlesing i stedet for staving kunnet eliminere feilkilden knyttet til skrivemotorikk. I intervensjonsprogrammet i studie 1 var det imidlertid et poeng å ha med noen aktiviteter der elevene kunne arbeide individuelt og da ble transposisjonsoppgavene vurdert som greie trenings- og repetisjons oppgaver. I intervensjonsprogrammet ble disse introdusert de siste ukene av intervensjonen da det ble antatt at elevene hadde automatisert bruk av symbolene. Elevene hadde dermed erfaring i selve oppgavetypen ved posttest og var vant med å transponere ord fra PAS til alfabetisk skrift. I henhold til opplysninger fra lærernes evalueringsskjema ble også disse individuelle oppgavene vurdert å ha god funksjonalitet. Selv om problemet med skrivemotorikk må sies å svekke begrepsvaliditeten av transposisjonstesten i studie 1, kan den likevel sies å være et valid mål på det direkte utbyttet av intervensjonen ved å måle «sluttproduktet» av treningen.

En annen relevant faktor som kan ha virket inn på resultatene er oppmerksomhet. Siden PAS symbolene er detaljerte og forutsetter at man fastholder flere momenter knyttet til de aktuelle artikulatoriske kjennetegnene på samme tid er det mulig at elever med svakt visuelt minnespenn kan ha hatt problemer med å tolke og dermed «avkode» PAS symbolene, uavhengig av om de har forstått at lydene kan representeres ved artikulatoriske egenskaper. En liknende feilkilde kan også gjøres gjeldende for elever med generelle

oppmerksomhetsvansker i retning av ADHD/ADD eller elever med oppmerksomhetsproblemer knyttet generell umodenhet.

Når det gjelder elevene i studie 1 ble elever med kjente lærevansker ekskludert fra studien. Det ble imidlertid ikke innhentet informasjon om generell kognitive status hos de elevene som ble trukket ut til å delta. Det er sannsynlig at noen av elevene som ble trukket ut til å delta var umodne elever uten konsistente fonologiske vansker (artikkel 1). For noen av disse kan det ikke utelukkes at generelt svak oppmerksomhet kan ha spilt inn både når det gjelder utbyttet av intervensjonen og for prestasjoner på transposisjonsoppgavene. Når dette er sagt bør det også nevnes at informasjon fra lærerne som deltok i studie 1 ikke tyder på at elevene har hatt problemer med å forholde seg til kompleksiteten i symbolene. Det generelle inntrykket er derimot at elevene relativt raskt har «knekt koden» når det gjelder logikken i PAS systemet med hensyn hvordan de ulike egenskapene ved lyden blir representert i symbolene.

Når det gjelder elevene som var med i studie 2 er de momentene som er diskutert over mindre aktuelle. Disse elevene var ved oppstart av studien mellom 10 og 14 år og ingen var rapportert å ha skrivemotoriske problemer. Ingen av elevene i denne studien var heller diagnostisert med vansker i AHD/ADD spekteret.

### **5. 7. 2 Validitet knyttet til måleverktøy**

Både studie 1 og 2 bruker delvis måleverktøy som er utviklet for dette prosjektet. I forbindelse med begge studiene ble det foretatt pilotering av måleverktøyene i forkant av oppstart noe som i begge tilfeller førte til justeringer. I begge tilfeller gjaldt justeringene reduksjon av antall oppgaver for å redusere tidsbruk.

Valg av egenproduserte måleverktøy er gjort utfra flere hensyn. For det første, i studie 1 var valg av egenprodusert materiale gjort av design hensyn. Når det gjaldt vurdering av fonologisk bevissthet var det ønskelig å vurdere effekten av tiltaket både på de ordene som ble brukt i intervensjonen, på ord som var matchet til de trente ordene og på pseudo ord som var matchet til de øvrige utfra fonologisk kompleksitet. Når det gjaldt ordlesing, ble det brukt en tilsvarende innfallsvinkel. De to ordlesingsoppgavene var ord brukt i treningen og ord som var matchet til disse. For vurdering av eventuell overføringseffekt ble det brukt to oversatte utgaver av TOWRE for måling av pseudoord lesing.



Når det gjaldt studie 2 var valg av egenprodusert materiale motivert ut fra behov for å redusere trusler mot den indre validiteten av studien. En viktig designspesifikk utfordring med repeterte målinger er som nevnt risiko for at eleven blir for godt kjent med måleinstrumentet og dermed husker oppgavene fra gang til gang. For å redusere risikoen for dette, og samtidig sikre reliabiliteten rundt de repeterte målingene, ble det laget likeverdige, men unike versjoner av ordlistene som ble brukt i lese- og diktatprøvene.

Dette ble gjort ved at det først ble tatt utgangspunkt i databasen over de mest høyfrekvente norske ordene (Norsk Frekvensordliste [http://www. tekstlab. uio. no/frekvensordlister/](http://www.tekstlab.uio.no/frekvensordlister/)). Deretter ble det plukket ut ord med regulær stavemåte. Disse ble så organisert etter ordlengde og fonologisk kompleksitet og delt inn i ni ordkategorier, hver bestående av 22 ord. Ordene var unike, men hadde lik fonologisk struktur og samme kompleksitet (artikkel 3). For hver leseprøve ble det tilfeldig trukket ut seks ord fra hver av de ni kategoriene (til sammen 54 ord) Samme strategi ble brukt for konstruksjon av prøvene for pseudoord og for diktatprøvene (artikkel 3). For irregulære ord ble en liknende strategi benyttet, men her ble ordene valgt fra eksisterende lesetester og organisert i tre pooler etter kompleksitet og antall stavelser. I den enkleste kategorien besto poolen av 22 ord og de to andre besto av 44 ord hver.

I alt ble det konstruert 36 versjoner av hver av de tre leseprøvene og 18 versjoner av diktatprøven. Når det gjaldt ordleseprøvene (ett minutt høytlesing) gjennomførte elevene to likeverdige versjoner for hver ordkategori (regulære ord, pseudoord og irregulære ord) ved hver kartlegging. For diktatprøven ble det gjennomført en prøve ved hver kartlegging. Rekkefølgen av versjonene av testene ble randomisert for hver enkelt elev, slik at alle elever ble testet med alle versjoner, men til ulikt tidspunkt og ikke i samme rekkefølge. Reliabiliteten til leseprøvene ble vurdert ved at skårene fra de to settene som ble administrert på samme tidspunkt ble sammenliknet gjennom en korrelasjonsanalyse (Pearsons  $r$ ). Resultatene viste en gjennomsnittlig korrelasjon på henholdsvis .968 for de regulære ordene, .898 for nonordene og .883 for de irregulære ordene.

Diktatprøver gjennomføres vanligvis uten tidsbegrensning, men av hensyn til tidsaspektet knyttet til de ukentlige kartleggingene ble det her satt en tidsfrist på 4 minutter. For elever som gjorde diktaten ferdig før tidsfristen på fire minutter ble det gitt tidsbonus basert på antall passerte og antall feil i den aktuelle prøven. Disse grepene ble gjort for å gjøre

de ukentlige kartleggingene overkommelige rent tidsmessig. Likevel kan det i et validitetsperspektiv innvendes at tidsaspektet kan ha virket inn og at prøven derfor også måler hastighet som egen dimensjon. Det kan heller ikke utelukkes at enkelte av deltakerne kan ha blitt påvirket negativt av tidtakingen og dermed gjorde flere feil enn om tiden ikke hadde blitt målt.

Målingen av i hvilken grad elevene lærte å bruke PAS symbolene til å representere fonemer ble målt med samme måleverktøy (transposisjon testen) i begge studier. Denne testen ble konstruert gjennom at det ble trukket nonord med ulik vanskegrad fra den allerede omtalte «poolen» av non ord i ni kategorier. For dette formålet ble kun ord fra de enkleste tre nivåene brukt. Det vil si to-lyds ord med strukturen VK eller KV (i alt 12), tre-lyds ord med ett konsonantkluster, VKK (i alt seks) og fire-lyds ord med KVKV struktur (i alt 6). I studie 2 ble det konstruert i alt 13 likeverdige versjoner (en test for hvert målepunkt) av denne testen og rekkefølgen av versjonene ble randomisert for hver deltaker.

Ved målingene som ble gjort i studie 1, ble en av versjonene valgt ut slik at alle elevene ble testet med nøyaktig samme test. For begge studier var instruksjonen til elevene at de skulle transponere («oversette») ordene som var skrevet i PAS font til ordinær alfabetisk skrift i løpet av fire minutter. Elevene arbeidet individuelt med oppgavene.

### **5. 7. 3 Implementeringsvaliditet**

For å kunne si noe om et tiltak har virket må det sikres at tiltaket utføres som planlagt og i tråd med retningslinjene i programmet. Noe av årsaken til at implementeringsvaliditet har blitt et viktig fokusområde henger sammen med ideen om at kvalitetsmessig utført forskning kan garantere for en generell effekt av et tiltak. Det vil si at det legges til grunn at intervensjonsprogrammet, dersom det gjennomføres i samsvar med forhåndsbestemte prosedyrer, er antatt å virke på samme måte i nye situasjoner (Cartwright, 2013; Kvernbekk 2018).

Dette prosjektet har hatt som mål å vurdere effekt av intervensjonene, men har samtidig delvis hatt en pragmatisk innretning der selve intervensjonsprogrammet og de oppsatte aktivitetene har hatt form av en ramme der lærerne som deltok også har hatt frihet til å tilpasse undervisningen. I lys av dette kan en viss effekt av lærerinnsats eller mangel på dette ikke utelukkes.

Implementeringsvaliditeten i de to utførte studiene er vurdert ved hjelp av loggføring fra lærer og lydopptak fra undervisningstimene. Gjennom disse kan man få et innblikk i både hvorvidt det oppsatte programmet følges som planlagt og man får også et inntrykk av kvaliteten på samhandlingen mellom elev og lærer i undervisningssituasjonen. I denne studien er implementeringsvaliditeten kun vurdert utfra i hvilken grad de oppsatte aktivitetene ble fulgt som planlagt. En nærmere kvalitativ vurdering og koding av aspekter knyttet til samhandlingen mellom lærere og elever både i de 14 gruppene som deltok i studie 1 og for lærer/elev dyadene i studie 2 ville trolig kunnet bidra med verdifull informasjon om effekt og mangel på effekt på et mer detaljert nivå. Dette har imidlertid ligget utenfor rammene for dette prosjektet.

#### **5. 7. 4 Diffusjon**

Når deltakere i samme eksperiment ferdes i samme miljø kan det foregå en «lekkasje» av metoder brukt i intervensjonsgruppen til kontrollgruppen, såkalt diffusjon (Shadish et al., 2002). I gruppestudien (artikkel 1) ble en eventuell diffusjonseffekt kontrollert for ved posttest ved at alle elevene, både i kontrollgruppene og eksperimentgruppene, ble vist eksempler på de spesielle PAS symbolene som ble brukt i intervensjonen og spurt om de kjente til disse. Resultatene viste 100 % samsvar mellom gruppetilhørighet og elevenes svar, noe som gir et sikkert holdepunkt for at elevene i kontrollgruppen ikke ble undervist med symbolene.

Når det gjelder studie 2 var deltakerne elever ved forskjellige skoler og en diffusjonseffekt av tiltaket mellom deltakerne er dermed ikke relevant (Winn et al., 2004).

#### **5. 7. 5 «Performance effekt»**

Performance effekt gjelder for undersøkelser der deltakerne er klar over at de deltar i et eksperiment og påvirkes av dette slik at den målte atferden i undersøkelsen ikke er representativ for typisk atferd (Shadish et al., 2002). I dette prosjektet er denne risikoen aktuell for begge studiene.

Selv om barna som deltok i studie 1 var unge, og trolig ikke hadde en bevisst holdning til at de var trukket ut til å delta i et eksperiment, er det sannsynlig at de forsto at de deltok i noe utenom det vanlige. Dette ikke minst fordi materialet som ble brukt i intervensjonen var nytt for alle deltakerne. En mulig performance effekt er også knyttet til lærerne som deltok. Siden lærerne som deltok i studie 1 ikke i utgangspunktet hadde valgt dette selv, men ble

pålagt å utføre intervensjonen fra skolens ledelse, kan det antas at lærernes innsats og engasjement i undervisningen kan ha vært annerledes enn vanlig. Dette kan ha påvirket utfallet i både positiv og negativ retning avhengig av i hvilken grad lærerne var positive eller negative til å delta i studien (artikkel 1).

I studie 2 var lærerne gjennomgående positive til å delta i intervensjonen og elevene var i høy grad klar over at de var med på noe nytt. En viss tilpasning av atferd i positiv retning som følge av at de var klar over at de deltok i et eksperiment kan derfor ikke utelukkes.

Målingsfeil som skyldes testleders forventninger, er også en aktuell problematikk i begge studier. Et mulig scenario er at et ønske hos testleder i retning av å oppnå positive resultater kan føre til at en til tross for standardiserte prosedyrer ubevisst gir mer støtte til elever i eksperiment gruppen enn i kontroll gruppen. Denne typen trusler ble i studie 1 forsøkt motvirket gjennom at elevenes personlige identifikasjonsnummer som ble brukt ved pre og posttest var tilfeldig tildelt før fordelingen til de to gruppene (intervensjon/kontroll) ble foretatt. Testleder var ved posttest dermed ikke klar over hvorvidt eleven var i eksperimentgruppe eller kontrollgruppe før siste deloppgave som omhandlet kjennskap til de artikulatoriske symbolene ble presentert.

I studie 2 var sammenhengen mellom kartleggingen og tiltaket «blindet» (Tate et al., 2016) ved at testleder ikke var den samme personen som den som utførte tiltaket. Opplæringen i bruk av materialet ble kun gitt til de lærerne som skulle arbeide sammen med eleven og testleder hadde derfor ikke kunnskap om materialet som ble brukt. Kontroll for eventuelle målingsfeil ble ellers i begge studier gjennomført ved at et representativt utvalg av lydopptak fra kartleggingene ble gjennomgått på nytt. For begge studier viste denne revurderingen et tilfredsstillende resultat (artikkel 1 &3).

## **5. 8 Sosial validitet**

SCD studier er i mindre grad enn RCT opptatt av å avdekke generalisert effekt, men kan sies å ha et mer klinisk perspektiv der fokus primært er å sikre konklusjoner angående hva som virker for en konkret elev (Lobo et al., 2017; Onghena et al., 2018; Maggin et al., 2018). Det legges derfor også vekt på sosial validitet, det vil si i hvilken grad signifikante personer rundt eleven opplever intervensjonen som positiv i henhold til både prosedyrer, målsettinger og effekt (Gast 2014). I studie 2 ble den sosiale validiteten undersøkt med et detaljert

spørreskjema til lærerne i etterkant av studien. Spørreskjema omhandler både elevenes utbytte og intervensjonsprogrammets opplevde funksjonalitet i forhold til ulike delmål.

## 5.9 Vurdering av effekt

### 5.9.1 Effektmål

Effektmålene som er brukt i dette prosjektet er delvis designspesifikke. Innenfor gruppemetodologi er Cohens  $d$  (Cohen 1988) det mest brukte effektmålet. Målet er basert på avvik fra et antatt normalfordelt gjennomsnitt og angis i standardavviks enheter. Cohen (1988) foreslo at effektstørrelser på 0.25 skulle regnes som liten, 0.50 som middels og 0.80 som stor. Denne skalaen ble brukt for å vurdere effekt av tiltaket i studie 1 og for vurdering av samlet deltakereffekt i studie 2.

For SCD har vurdering av effekt tradisjonelt brukt inspeksjon av grafisk framstilling av resultater fra de repeterte målingene som utgangspunkt for å vurdere hvorvidt det har skjedd en endring som kan assosieres med tidspunkt for når tiltaket ble satt i verk (Gast & Spriggs 2014). Nyere SCD bruker imidlertid også i økende grad både parametriske og ikke parametriske kvantitative mål for å vurdere individuell effekt. De siste årene har det som et tillegg til disse også blitt lansert statistiske metoder for utregning av samlet deltakereffekt med samme metriske system som Cohens  $d$  (Hedges, Pustejovsky, & Shadish, 2012; Pustejovsky et al., 2014; Valentine, Tanner-Smith, Pustejovsky, & Lau, 2016).

Både de statistiske løsningene og de tradisjonelle visuelle metodene har svakheter blant annet knyttet til validitet (Pustejovsky, 2019; artikkel 2), og i de fleste nyere publikasjoner anbefales det å bruke flere måleverktøy for å sikre konklusjonene (Shadish et al., 2015; Wolfe, Dickenson, Miller & Mc Grath, 2019).

### 5.9.2 Visuelle analyser

De visuelle analysene omhandler vurderinger av data både i hver fase og ved sammenlikning mellom fasene og inkluderer inspeksjon av nivåendringer, eventuelle trender, overlapp mellom data, konsistens i datamønstrene, variasjon i data, og eventuell umiddelbar effekt (Kratochwill et al., 2013; Lane & Gast, 2014). Nyere artikler anbefaler en multimetodisk tilnærming for SCD der det brukes både visuelle analyser og statistiske analyser (Lobo et al., 2017; Wolfe et al., 2019). En «arbeidsdeling» mellom disse to har også blitt foreslått ved at man først bruker visuelle analyser for å vurdere kvaliteten på innsamlede data og så bruker kvantitative effektmål for å vurdere størrelsen av effekten dersom kvaliteten

tilsier at det er grunnlag for å vurdere effekt (Lobo et al., 2017). I studie 2 ble det utarbeidet grafikk for alle elever og variabler ved hjelp av Microsoft Excel og visuelle analyser ble brukt for å vurdere stabilitet, trender og nivåendringer mellom fasene. Visuelle analyser ble også brukt som utgangspunkt for å spesifisere modeller for de statistiske analysene av gjennomsnittlig intervensjonseffekt ( artikkel 3).

### 5. 9. 3 Statistiske analyser

*Studie 1:* I studie 1 ble effekt fra intervensjonen vurdert gjennom flernivå regresjons analyser. I analysene ble resultater for hver avhengig variabel ved posttest satt opp i regresjonsmodeller med pretest og en dummy variabel (intervensjon/kontroll) som uavhengig variabel. For alle analyser ble det brukt en Restricted Maximum Likelihood estimator. På de neste nivåene i analysene ble intraklasse korrelasjoner vurdert ved å inkludere både variasjon mellom skoler og variasjoner mellom intervensjonsgrupper. Det ble på disse nivåene brukt modeller tilpasset et «partial nested RCT design» (Lohr et al., 2014). Disse modellene tar høyde for at deltakerne internt i hver av intervensjonsgruppene vil kunne være likere hverandre enn det de er i forhold til elevene i intervensjonsgruppene fra andre skoler. Der resultatene viste at dette var tilfelle ble dette korrigert for i analysene (artikkel 1).

*Studie 2:* I tråd med en multimetodisk innfallsvinkel ble det i studie 2 valgt å bruke visuelle analyser og statistiske analyser parallelt. Utregning av effektstørrelser i SCD gjøres hovedsakelig med utgangspunkt i to forskjellige prinsipper som også representerer ulike statistiske metoder. I den første, basert på parametrisk statistikk, er grunnlaget for utregningen sammenlikning av gjennomsnittsskåren i baselinefasen og intervensjonsfasen (Hedges et al., 2012). I den andre, basert på ikke-parametrisk statistikk, vurderes effektstørrelsen ved å sammenlikne graden av overlapp av datapunkter mellom fasene (Parker, Vannest, & Davis, 2011).

Det ble valgt å bruke et parametrisk effektmål, Standard Mean Difference (SMD) (Busk & Serlin, 1992), et ikke-parametrisk effektmål, Tau\_U (Parker, Vannest, Davis, & Sauber, 2011) og et effektmål for vurdering av samlet deltakereffekt, Between Case- Standard Mean Difference (BC-SMD) (Pustejovsky et al., 2014; Valentine et al., 2016). Disse effektmålene ble valgt fordi de ble antatt å belyse data fra litt ulike vinkler samtidig som de også gir mulighet for å regne ut konfidensintervall og dermed vurdere hvorvidt effekttestimatet

er statistisk signifikant. For utregning av individuell effekt i studie to ble det brukt tilgjengelige nettbaserte effektkalkulatorer (artikkel 2 og 3).

Når det gjaldt målet for samlet deltakereffekt, BC – SMD er dette et relativt nytt analyseverktøy spesielt utviklet for å kunne sammenlikne resultater fra SCD i metaanalyser (Valentine, Tanner-Smith, Pustejovsky, & Lau, 2016). BC-SMD bruker samme skala som Cohens  $d$  og dette gjør at det i prinsippet er mulig å sammenlikne resultater fra SCD med resultater fra gruppestudier på samme skala (ibid.). Utregningen av effektstørrelse foregår ved bruk av en online-kalkulator og baseres på flernivå regresjonsanalyser med mulighet for å spesifisere individuell variasjon på det første nivået og felles trekk ved data på det andre nivået i modellene. Det ble i alle utregningene brukt en Restricted Maximum Likelihood estimator. Videre ble det valgt å spesifisere modellene for utregning av effekt utfra kriterier foreslått av Wolfe et al.(2019). Deres studie vurderer sammenhengen mellom ulike effektmål i SCD og konkluderer med at BC-SMD brukt med de foreslåtte kriteriene for modellspesifikasjon korrelerer godt med øvrige effektmål. For nærmere beskrivelser se ellers artikkel 3.

I artikkel 2 som omhandler Single Case Design anvendt i evaluering av spesialpedagogiske tiltak er det i tillegg til de omtalte effektmålene også inkludert to supplerende effektmål, Log Response Ratio (LRR) og Percent of nonoverlapping data (PND). Disse ble inkludert fordi de antas å kunne være spesielt anvendelige for praksisfeltet. Med LRR kan effekttestimatet konverteres til et mål på prosentvis økning/reduksjon. Dette kan intuitivt være enklere å forstå enn effekttestimater og antas å kunne ha praktiske fordeler når effektstørrelsen skal drøftes også med personer uten spesiell forhåndskunnskap (for eksempel foreldre og lærere) (Pustejovsky, 2018b). PND har vært kritisert som effektmål (artikkel 2), men kvantifiserer også endringene mellom baseline og intervensjonsfasen på en lett forståelig måte ved å vise til graden av overlappende datapunkter i de to fasene og ble derfor tatt med.

## **5. 10 Etske problemstillinger**

Det er generelt etiske problemer knyttet til rekruttering av elever til utprøving av intensive spesialpedagogiske tiltak. Dette fordi elevene som rekrutteres bruker tid på et tiltak man ikke vet om oppleves nyttig og at det i tillegg påløper tid for elevene til kartlegging før og etter intervensjonen.

I begge studiene ble skriftlig informert samtykke om deltakelse i studiene innhentet fra foresatte. Foresatte gir tillatelse på vegne av umyndige barn i tråd med gjeldende retningslinjer fra Norsk Senter for forskningsdata. Det ble forutsatt at foresatte informerte og diskuterte deltakelse med barnet før de ga sitt samtykke. I informasjonsskrivene er det også poengtert at deltakere når som helst kan trekke seg fra studien uten å gi begrunnelse. For studie 2 ble det i tillegg understreket at det å trekke seg fra deltakelse i studien ikke ville få innvirkning på bistand fra Statped i henhold til henvisningen. En av elevene i studie 2 benyttet seg av muligheten til å trekke seg. For denne eleven ble tiltaket avsluttet tidligere enn planlagt og resultatene ble ikke inkludert i analysene.

I forbindelse med studie 2 er det også noen etiske utfordringer knyttet til valg av design siden kravet om repeterte målinger i SCD førte til at elevene som deltok i denne studien gjennomførte i alt 18 kartlegginger i løpet av prosjektperioden og dermed brukte mye tid på dette. Elevens motivasjon for å delta i kartleggingsprosedyrene gjennom alle faser av studien ble evaluert gjennom et spørreskjema til samtlige testledere i etterkant av tiltaksfasen. Svarene her indikerer at ingen av elevene viste motstand mot å delta i kartleggingen verken gjennom verbale ytringer eller ved å fysisk nekte å delta. Det kom imidlertid fram at det for noen av elevene enkelte ganger kunne være lite lystbetont å måtte forlate klasserommet, særlig dersom resten av klassen var involvert i sosiale aktiviteter.

I forbindelse med studie 1 er det også noen designspesifikke etiske utfordringer knyttet til randomisering av elevene i to grupper der den ene gruppen får tiltaket og den andre ikke. Denne skjevfordelingen både av pedagogiske ressurser og av mulige fordeler knyttet til deltakelse i intervensjonsprogrammet innebærer noen etiske dilemmaer, gitt at elevene har likeverdige vansker og likt behov for bistand. Disse dilemmaene ble diskutert med skolekontoret i den aktuelle kommunen før studien ble påbegynt. Skolekontoret ønsket å ha mulighet til å også gi elevene i kontrollgruppene noe ekstra hjelp. I og med at mye av skolens ressurser var bundet til utprøving med PAS materialet, måtte tidsrammen for intervensjonsperioden reduseres i forhold til det opprinnelig planlagte (og ønskede) omfanget. Dette for å gi skolene mulighet til å sette i verk tilpassede tiltak for elevene i kontrollgruppen før sommerferien.



## 6 Oppsummering av artiklene

### **Artikkel 1: Effects of Articulatory Consciousness Training in First Graders with a Reading Delay: A Randomised Control Trial**

I artikkel 1 vurderes effekten av en fem ukers intervensjon rettet mot elever i første klasse med forsinket leseutvikling (N= 129). Intervensjonen ble organisert som smågruppe undervisning 4 timer pr uke med 3-5 deltakere pr gruppe. Lokale lærere fikk opplæring gjennom et gruppebasert kurs på 3 timer. Materiell i form av PAS kort og oppgaver ble utdelt og gruppene fulgte et forhåndsbestemt program med konkrete planer for hver uke. Elevene som var trukket ut til kontrollgruppene mottok ingen spesiell tilrettelegging, men deltok i den ordinære klasse og smågruppeundervisningen som vanligvis ble tilbudt på skolen (såkalt «business as usual»).

Forskningsspørsmålene angikk:

- Hvorvidt barna faktisk lærte å automatisere de artikulatiske symbolene (PAS).
- Hvorvidt barna i intervensjonsgruppene gjennomsnittlig oppnådde bedre skårer på mål for fonologisk bevissthet, avkoding av ordinære ord, avkoding av non ord, lesenøyaktighet og bokstavkunnskap sammenliknet med kontrollgruppene.

Resultatene viste at barna i intervensjonsgruppene lærte å bruke PAS systemet, men det var ikke mulig å identifisere spesiell effekt av treningen på de øvrige post testene sammenliknet med elevene i kontrollgruppene. Tendensen var at alle elevene, både de som fikk treningen og de som ikke fikk treningen, hadde en positiv utvikling fra pre til posttest. Artikkel 1 konkluderer derfor med at gruppebasert undervisning, med det valgte treningsprogrammet og i det valgte omfanget, ikke ga bedre effekt på fonologisk bevissthet og avkodingsferdighet enn det tilbudet den aktuelle kommunen selv allerede hadde etablert og som ble gitt til elevene i kontrollgruppene.

## **Artikkel 2: Single Case Design i evaluering av spesialpedagogiske tiltak: En gjennomgang av aktuelle analysemetoder for vurdering av effekt.**

Artikkel 2 presenterer egenskaper og metodiske utfordringer ved SCD og drøfter designets funksjonalitet når det gjelder evaluering av individuelle spesialpedagogiske tiltak i en norsk kontekst. Som eksempel brukes del-data fra studie 2 der hovedvekt er lagt på den praktiske gjennomføringen av studien, herunder samarbeid mellom pedagogisk psykologisk tjeneste og skolene, operasjonalisering av avhengig variabel, ulike mønstre i innsamlede data på bakgrunn av visuelle analyser og de ulike effektmålenes egnethet.

Artikkelen konkluderer med at SCD kan være godt egnet til vurdering av spesialpedagogiske tiltak og at selv enkle tofase design kan bidra til valide vurderinger av hva som virker for en konkret elev. Det diskuteres også hvordan pedagogisk psykologisk tjeneste gjennom å sikre økt eksperimentell kontroll og gjennom å replisere resultatene også kan bidra til utvikling av forskningsbasert praksis.

## **Artikkel 3: The Impact of Articulatory Consciousness Training on Reading and Spelling Literacy in Students with Severe Dyslexia: An Experimental Single Case Study.**

Artikkel 3 vurderer effekten av artikulatorisk bevissthetstrening i et utvalg av elever mellom 10-15 år (N=13) med alvorlig og vedvarende dysleksi. Utvalget er hentet fra ventelisten til Statped og for samtlige elever var henvisningsgrunnen behov for nye retningslinjer i spesialundervisningen. Intervensjonen foregikk over åtte uker og ble organisert som individuell trening 4 timer pr uke. Lokale lærere fikk individuell veiledning i forkant (1, 5 timer) og fulgte et forhåndsbestemt program med innebygd progresjon fra uke til uke. De fikk også utdelt intervensjonsmateriell som omfattet PAS kort og plansjer. Effekten av tiltaket ble målt gjennom repeterte målinger i tre faser hvorav de fem første målingene for hver elev (baseline) fungerte som en kontrollfase.

Forskningsspørsmålene var knyttet til:

- Hvorvidt treningen førte til positiv endring i form av økte råskårer på ordlesingstester for ord med regulær stavemåte, non ord og ord med inkonsistent stavemåte.

- Hvorvidt treningen førte til økt lesenøyaktighet for ord med regulær stavemåte, non ord og ord med inkonsistent stavemåte.
- Hvorvidt treningen førte til bedre staveferdigheter.

Resultatene ble målt både for hver enkelt deltaker og som en samlet effekt og viste at majoriteten av elevene har hatt positiv effekt av tiltaket. Effekten varierer imidlertid mellom deltakerne, både når det gjelder effektstørrelse og i forhold til målte variabler. Mest konsistent effekt på tvers av deltakere sees på målet for non ord lesing og non ord staving, mens effekten på ord med irregulær stavemåte var gjennomgående mindre. Designet inkorporerte også et kontrollmål for å vurdere eventuell testeffekt av de repeterte målingene. Også dette ga signifikant effekt om enn i et mindre omfang enn for flertallet av de øvrige avhengige variablene.

## 7 Diskusjon

I dette kapitlet vil resultatene fra de to empiriske studiene bli diskutert både i lys av forskningsspørsmålene i artikkel 1 og 3, begrensninger ved studiene og de mer overordnede problemstillingene som er redegjort for i denne introduksjonsdelen (kapittel 2). Videre vil konklusjonene fra artikkel 2 også drøftes i et bredere perspektiv.

### 7.1 Effekten av artikulatork bevissthetstrening

Prosjektet har hatt som målsetting å vurdere om, og eventuelt i hvilken grad, søkelys på artikulatork bevisstgjøring i opplæringen av elever med dysleksi (studie 2) eller med forsinket leseutvikling (studie 1) kan bidra til å styrke skriftspråklige ferdigheter.

Den grundige dokumentasjonen av at opplæring rettet mot det fonologiske systemet gir best effekt for elever med dysleksi (Galuschka et al., 2014; SBU, 2014), gjør det nødvendig å lete etter relevante pedagogiske innfallsvinkler for å møte behovene til de elevene som i mindre grad nyttiggjør seg tradisjonell lyttebasert leseopplæring. I dette prosjektet har artikulatork bevissthetstrening vært prøvd ut som et middel til å realisere leseopplæring rettet mot det fonologiske systemet. Hovedhypotesen har vært at trening i artikulatork bevissthet antas å kunne bidra til at de fonologiske representasjonene får en mer presis identitet, samtidig som artikulasjon også antas å kunne bidra til å lette tilgangen til de fonologiske representasjonene i avkodingsprosessen. Resultatene fra de to gjennomførte studiene vil nedenfor kort bli drøftet i lys av en flerfaktormodell for dysleksi før resultater knyttet til de målte variablene diskuteres punktvis.

#### 7.1.1 Resultater vurdert i lys av en flerfaktormodell for dysleksi

Som nevnt er nyere forskning i økende grad opptatt av multifaktorielle årsakssammenhenger når det gjelder lesevansker (Pennington, 2006; Snowling, 2008; Snowling, Gallagher, & Frith, 2003). Et viktig moment i denne sammenhengen er forholdet mellom latente risiko faktorer knyttet til familiær forekomst, såkalte endofenotyper og den faktiske forekomsten av lesevansker hos barn med arvelig risiko (Eklund, Torppa, Aro, Leppänen, & Lyytinen, 2015; Moll, et al., 2013; Pennala et al., 2010; Melby-Lervåg et al., 2012; Snowling, 2008; Snowling & Melby-Lervåg, 2016).

Flere longitudinelle studier har konkludert med at fonologisk bevissthet alene er en usikker prediktor for dysleksi. Moll et al. (2013) fant for eksempel at både fonologisk

bevissthet og nonord repetisjon framsto som såkalte endofenotyper for dysleksi, mens svak fonologisk bevissthet alene ikke var en sikker prediktor. I en større longitudinell studie fra Finland fant Pennala et al (2010) at perseptuell sensitivitet for varighet av språklyder var signifikant svakere i delen av utvalget med familiær risiko, men at kun de barna som også hadde flere kognitive risikofaktorer som svakt verbalt korttidsminne, dårlig benevningshastighet og lav verbal IQ utviklet dysleksi. Andre undersøkelser har i tillegg til fonologisk bevissthet pekt på benevningshastighet, bokstavkunnskap og vokabular som viktige samtidige prediktorer (Catts and Vaughn, 2018; SBU 2014).

Longitudinelle studier tyder altså på at den opplevde lesevansken kan forsterkes og modereres av generelle kognitive og miljømessige faktorer som er eksterne til det fonologiske problemet. For det andre kan også det fonologiske grunnproblemet ha ulikt opphav knyttet til svikt i ulike deler av den kausale kjeden (Hulme & Snowling 2009). En slik multifaktoriell forståelse når det gjelder årsakenene til dysleksi innebærer også at det er sannsynlig at ulike teoribaserte treningsopplegg vil ha ulik effekt avhengig både av individuell vanskeprofil og eksterne forhold.

I dette prosjektet er en av flere teoretisk sett mulige årsakskjeder knyttet til det fonologiske problemet valgt ut som utgangspunkt for trening i to utvalg: Et relativt heterogent utvalg av elever i første klasse med forsinket leseutvikling og et mer homogent utvalg av eldre elever med store fonologiske vansker, høy grad av familiær risiko for dysleksi, samt fravær av risikofaktorer knyttet til nonverbalt evnenivå og kjente komorbide vansker.

Resultatene fra studiene tyder på at forekomst av svak fonologisk bevissthet i første klasse ikke er en klar indikator for å innføre artikulatork trening som alternativ til lyttebasert opplæring (artikkel 1). I utvalg med en klarere definert vanskeprofil og svært svake fonologiske ferdigheter indikerer resultatene fra studie 2 at artikulatork trening kan gi effekt i form av positiv endring i retning av mer nøyaktig avkoding. Disse resultatene synes å være mest tydelige på lesing av non ord, lesing av ord med regulær stavemåte og staving av non ord og i mindre grad i forhold til ord med irregulær stavemåte. Imidlertid er et viktig funn i studie 2 at det, også innenfor dette begrensede utvalget på kun 11 elever, finnes individuelle forskjeller med hensyn til vanskeprofil som mulig kan ha påvirket i hvilken grad deltakerne har hatt utbytte av treningen (artikkel 3).

Resultatene antyder at de elevene som på forhånd hadde mest unøyaktig avkoding også var de som profitterte mest på treningen. Designet er begrenset når det gjelder muligheten til å trekke generaliserte slutninger og det kan selvsagt også være andre årsaker til ulikt utbytte. Divergerende resultater kan imidlertid også tolkes i lys av heterogene årsakssammenhenger på individuelt nivå.

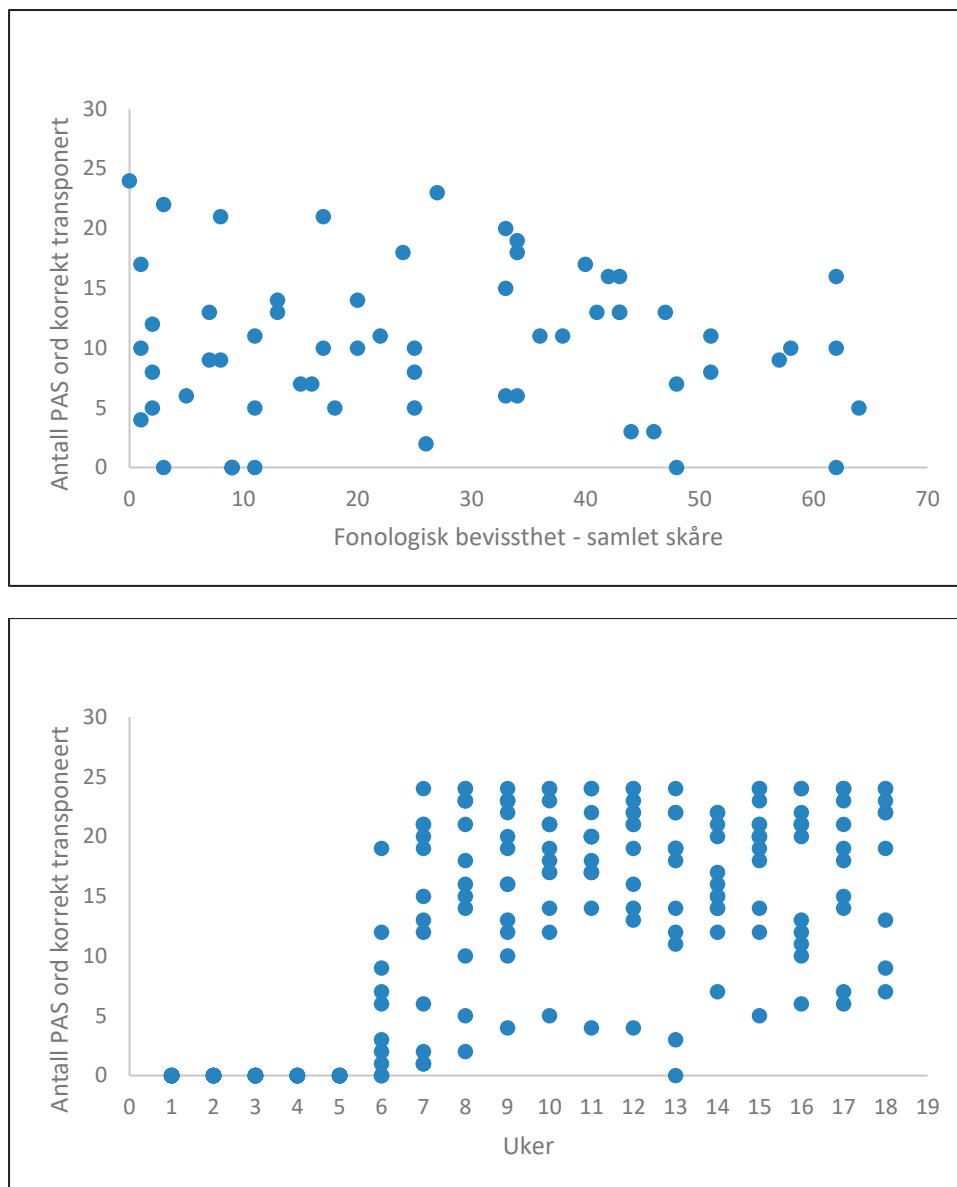
### **7. 1. 2 Utvikling av artikulatortisk bevissthet**

Artikulatortisk bevissthet ble i begge studiene målt med transposisjonstesten der elevene oversatte nonord skrevet med artikulatortiske symboler til ordinær skrift på tid. Resultatene fra studiene viser at elevene i studie 2 nesten nådde takeffekt på denne testen etter ca. to ukers opplæring. Når det gjaldt elevene i første klasse varierte resultatene i stor grad etter fem ukers opplæring (figur 7.1).

Studiene kan ikke gi et klart svar angående årsaken til forskjellene. En årsak kan selvsagt være knyttet til testens validitet (se over), kanskje særlig med tanke på at kreditering på denne testen krever skrivemotoriske ferdigheter som kan ha vært svakt i utvalget i første klasse. Det kan være at elevenes alder har spilt inn, det kan være at den individuelle opplæringen som ble gitt i studie 2 er mer egnet for opplæring i PAS systemet enn den gruppebaserte opplæringen som ble gitt i studie 1. En hypotese kan også være at selve systemet appellerer mer til elever med svært svake fonologiske ferdigheter enn til elever med mindre alvorlige fonologiske vansker. I lys av det siste er det interessant at flere av de svakeste elevene i første klasse hadde god utvikling på transposisjonstesten (figur 7. 1). Gitt at dette var elever som hadde et svært svakt utgangspunkt i både lesing og fonologisk bevissthet før intervensjonen, er det imponerende at flere av dem oppnådde såpass gode resultater på transposisjonstesten. Skåre på denne krever som nevnt både avkoding av ukjente non ord skrevet i PAS, samt å gjengi disse i alfabetisk skrift ved å skrive selv. Dette kan tyde på at noen av elevene opplevde ordavkoding basert på artikulatortisk informasjon som enklere enn ordavkoding basert på ordinær alfabetisk skrift. Denne tendensen kan imidlertid ikke gjøres gjeldende for utvalget som helhet og gjaldt heller ikke for samtlige av de svakeste elevene (artikkel 1).

**Figur 7.1**

*Resultater fra transposisjonstesten*



Figur 7.1 Scatterplot som viser skårer på transposisjonstesten (Max skåre = 24) Øverste panel: Resultater fra studie 1. Oppnådd skåre på transposisjonstesten etter fem ukers trening sett i forhold til resultater fra samlet skåre fra tre deltester i fonologisk bevissthet ved pretest ( $n = 64$ ). Nederste panel: Resultater fra studie 2. Oppnådd skåre på transposisjonstesten for alle elever ( $N = 11$ ) gjennom 18 uker (intervensjonen startet i uke 6).

### 7. 1. 3 Effekt av treningen på ordavkoding

Både i studie 1 og i studie 2 ble elevene kartlagt i forhold til lesing av ord med ulik struktur. I studie 1 ble det brukt egenproduserte ordlister for lesing av ord med regulær stavemåte, mens det ble brukt to oversatte versjoner fra TOWRE for nonord lesing (ett minutt høytlesing). I tillegg ble elevene i studie 1 også testet med en prøve i lesenøyaktighet av non ord, uten tidsfrist. Resultatene fra studie 1 viste at effekten av treningen ikke ga bedre resultater enn den undervisningen skolene vanligvis tilbød når det gjaldt bedret leseferdighet, verken når det gjaldt lesing av regulære ord eller non ord. Det var heller ingen signifikant effekt i forhold til lesenøyaktighet.

I studie 2 er leseferdighet målt med ett minutt høytlesing av regulære ord, irregulære ord og nonord. I studie 2 er ordlistene for de ukentlige målingene egenproduserte, mens det ble brukt standardiserte versjoner fra STAS batteriet i pre og posttest. Her har alle elevene, med unntak av en, vist framgang når det gjaldt lesing av regulære ord og nonord. For de irregulære ordene er det gjennomgående lavere effektstørrelser og færre som oppnår signifikante resultater. Best individuell effekt sees på målet for nonord, noe som også bekreftes av resultatene fra den standardiserte posttesten (artikkel 3). For flertallet av elevene så man også at nøyaktigheten i avkodingen økte for både regulære ord og non ord (artikkel 3).

Generelt må konklusjoner om effekt av tiltaket i studie 2 ta høyde for en mulig testeffekt av de repterte målingene (artikkel 3). Med dette forbeholdet viser resultatet at elevene har hatt best framgang på nonordlesing. Nonordlesing assosieres ofte med bruk av fonologisk lesestrategi (Pennington & Bishop, 2009). Resultatene indikerer dermed at treningen hadde best effekt på utvikling av denne lesestrategien. Dette er konsistent med resultater fra andre studier som har brukt en artikulatork tilnærming (Torgesen et al., 2001; Joly-Pottuz, Mercier, Leynaud, & Habib, 2008).

Resultatet for irregulære ord var i tråd med hypotesen før intervensjonen. Siden tiltaket som helhet var rettet inn mot utvikling og forbedring av fonologiske lesestrategier var det forventet mindre effekt på lesing av ord med irregulær stavemåte.



#### **7. 1. 4 Effekt av treningen på staveferdighet**

Staving ble kun vurdert i studie 2 og resultatene målt med de individuelle effektmålene viser at ca. halvparten av utvalget oppnår resultater som ikke er statistisk signifikante. Av de elevene som hadde signifikante resultater er effektstørrelsene i området moderat til stor effekt (artikkel 3).

#### **7. 1. 5 Effekt av treningen på fonologisk bevissthet**

Effekten av treningen på fonologisk bevissthet ble kun målt i studie 1. Her viser resultatene at det ikke er oppnådd signifikant effekt av treningen i forhold til fonologisk bevissthet sammenliknet med undervisningen som ble gitt elevene i kontrollgruppen (artikkel 1).

#### **7. 1. 6 Effekt av treningen på bokstavkunnskap**

Effekten av treningen i forhold til bokstavkunnskap ble kun målt i studie 1. Resultatene viser at det ikke ble oppnådd signifikant bedre effekt på dette målet for elevene i intervensjonsgruppene sammenliknet med kontrollgruppene. For utvalget som helhet var det også delvis takeffekt på dette målet (artikkel 1).

#### **7. 1. 7 Effekt av artikulatork trening sammenliknet med fonologisk basert trening**

De fleste tidligere studiene på området (Castiglioni-Spalten & Ehri, 2003; Torgesen et al., 2001; Trainin, Wilson, Murphy-Yagil, & Rankin-Erickson, 2014) har sammenliknet intervensjoner med artikulatork trening med intervensjoner med fonologisk (primært lyttebasert) trening. De to studiene som er gjennomført her har ikke hatt det som utgangspunkt. I begge studiene ble den aktuelle intervensjonen sammenliknet med det tilbudet elevene ville fått som del av den vanlige tilretteleggingen dersom de ikke hadde vært deltakere i studiene (såkalt «business as usual»).

Siden lyttebasert trening er en viktig del av opplæringen, både i den grunnleggende leseopplæringen og i spesialundervisning for elever med lesevansker kan resultatene fra begge studier likevel delvis også tolkes i lys av denne problemstillingen.

I studie 1 var designet lagt opp som en sammenlikning mellom en gruppeintervensjon med artikulatork trening og vanlig undervisning i klasserommene (business as usual). Pilotstudien i et utvalg klasserom i samme kommune avdekket at den vanlige undervisningen i klasserommene i stor grad inneholdt lyttebasert fonologisk trening. Spørreskjema til lærerne

i etterkant av intervensjonen viste også at mange av elevene i kontrollgruppen hadde hatt fonologisk basert trening i liten gruppe eller individuelt i perioden intervensjonen foregikk (artikkel 1).

Elevene i studie 2 var henvist til Statped grunnet dårlig effekt av igangsatte spesialpedagogiske tiltak og opplevelsen av at «alt er prøvd, men ingenting virker» var utbredt. Studien inkluderer ikke detaljerte opplysninger om spesialundervisningen gjennom hele elevens skolehistorie. Data inneholder kun informasjon om innretning og innhold i spesialundervisning i det semesteret intervensjonen ble gjennomført, altså i tidsrommet før intervensjonen startet (ca 5-9 uker). I henhold til opplysningene fra lærerne hadde 8 av 11 elever (2 har ikke besvart) en eller annen form for lyttebasert trening i denne perioden.

Erfaringene fra studie 1 tyder på at den artikulatorkasert baserte treningen elevene fikk i intervensjonsgruppene ikke ga bedre effekt på noen av de målte ferdighetene enn den treningen elevene i kontrollgruppene fikk gjennom de ordinære tiltakene skolene hadde til rådighet. I artikkel 1 diskuteres det mulige årsaker til dette også i relasjon til begrensninger knyttet til designet i den gjennomførte studien. Konklusjonen er likevel at artikulatorkasert trening i dette omfanget (fem uker) ikke hadde effekt på fonologisk bevissthet og lesing, alternativt at det hadde like god effekt som lyttebasert trening i fonologisk bevissthet. Den siste konklusjonen støtter resultater fra tidligere studier der designet har inkludert to likeverdige intervensjonsgrupper (Torgesen et al., 2010; Wise, Ring, & Olson, 1999).

Når det gjelder de eldre elevene med dysleksi (studie 2) kan det se ut som de fleste elevene hadde bedre effekt av treningen i intervensjonsperioden sammenliknet med den treningen de hadde før intervensjonen startet. For flertallet av elevene var tidsbruken også stort sett den samme under intervensjonen sammenliknet med tidsbruk i tidligere spesialundervisning. Resultatene er imidlertid ikke entydige og noen av elevene hadde begrenset utvikling i løpet av intervensjonsperioden. Et forbehold når det gjelder resultatene er også at resultatene fra pegboard-testen indikerte av selve målemetoden med de repeterte testene kan ha hatt effekt på utfallet (artikkel 3). Et viktig moment i dette utvalget er at det også må tas i betraktning at elevene opplevde det som positivt å prøve noe nytt. Alle lærerne i studie 2 rapporterte at elevene likte å arbeide med PAS symbolene og at de opplevde stor grad av mestring i å lære seg å representere fonemene med artikulatorkaserte symboler. Elevenes resultater fra transposisjonstesten (figur 7.1), bekrefter at flertallet av elevene hadde rask

progresjon og oppnådde tilnærmet takeffekt på denne testen etter ca. 2 ukers trening. Den målte effekten på lesing og staving kan derfor også delvis forstås som et resultat av økt motivasjon og arbeidsinnsats i timene som en følge av at opplegget hadde nyhetens interesse.

### **7. 1. 8 Effekt av artikulatorisk trening for elever med spesielt svake fonologiske ferdigheter**

Prosjektet har som helhet ønsket å vurdere effekten av artikulatorisk trening for elever med svake fonologiske ferdigheter. Den eventuelle effekten er knyttet til hvorvidt det å rette søkelys på de artikulatoriske kjennetegnene ved lydene kan bidra til at det blir enklere å differensiere mellom lyder med like egenskaper. Denne hypotesen har som utgangspunkt at nettopp dette er et kjerneproblem for elever med svake fonologiske ferdigheter.

I utvalget i første klasse (artikkel 1) var det som nevnt ikke bedre effekt av den artikulatoriske treningen enn av den ordinære treningen som foregikk i klasserommene. En av begrensningene ved denne studien er at utvalget viste seg å være langt mer heterogent enn antatt. Til tross for at elevene var trukket ut fordi de hadde skåret under kritisk grense på relevante delprøver fra den nasjonale leseprøven som gjennomføres i full klasse, viste resultatene fra pretest at mange av elevene i utvalget skåret godt på prøver i fonologisk bevissthet og avkodning når disse ble administrert som individuell kartlegging. I artikkel 1 diskuteres det hvorvidt disse ulikhetene i prestasjoner på oppgaver med relativt likt innhold kan ha sammenheng med at selve testsituasjonen ved gruppeprøver og individuelle prøver er ulike. Artikkel 1 diskuterer også hvorvidt utvalget i studie 1 mulig inkluderer såkalt «falske positive», det vil si elever med alderstypiske fonologiske ferdigheter som skåret svakt på gruppetesten av andre årsaker. Av designmessige årsaker, blant annet knyttet til utgangspunktet for randomiseringen, lot det ikke gjøre å foreta meningsfulle statistiske analyser på mindre deler av utvalget som i større grad viste svake fonologiske ferdigheter også ved individuell kartlegging.

I utvalget av elever med dysleksi hadde samtlige definerte fonologiske vansker. Disse framkom ved pretest både i form av lav lesehastighet, unøyaktig avkodning og mange feil på nonorddiktaten fra STAS (artikkel 3). I dette utvalget er også elevenes generelle læreforutsetninger dokumentert ved at det ble gjennomført delprøver fra WISC IV ved pretest. Elevene har i henhold til denne en relativt lik evneprofil med aldersadekvate skårer innenfor nonverbale ferdigheter og semantisk baserte språkferdigheter, og gjennomgående

svakere resultater på delprøven tallminne (artikkel 3). Koherente opplysninger viste at flertallet i utvalget med unntak av to elever hadde forekomst av dysleksi i nær familie, noe som tyder på arvelig disposisjon. For seks av elevene ble det i det retrospektive spørreskjemaet til foreldrene også rapportert at det hadde vært bekymring knyttet til avvikende språkutvikling i førskolealderen. Utvalget i studie 2 kan derfor sies å bestå av elever med tydelige tegn på fonologisk svikt, selv om det også var betydelige forskjeller mellom elevene både i forhold til alvorlighetsgrad og symptomer

Elevene i denne studien lærte raskt å bruke de artikulatoriske symbolene til å representere fonemer. Man så også en gradvis bedring i resultater på de ukentlige målingene ved en generell økning av antall korrekt leste ord og ved at lesenøyaktigheten økte. Konklusjoner fra studien som helhet må ta forbehold når det gjelder resultatene siden det også framkom effekt på den avhengige variabelen (pegboard-testen) som ikke var relatert til intervensjonen. Selv om dette indikerer en viss testeffekt som følge av de repeterte målingene konkluderer imidlertid artikkel 3 med at dette ikke forklarer hele endringen fra baseline til intervensjon/postfasen. I dette utvalget kan det dermed se ut til at trening i artikulatorisk bevissthet, i alle fall i noen grad og for flertallet av deltakerne, påvirket elevenes leseferdighet i retning av mer nøyaktig fonologisk avkoding.

## **7.2 Single Case Design i evaluering av spesialpedagogiske tiltak**

En av problemstillingene i dette prosjektet har vært å vurdere funksjonaliteten av SCD i relasjon til et avgrenset område, nærmere bestemt vurdering av effekten av spesialundervisning etter enkeltvedtak. Eksisterende vurderingspraksis baseres hovedsakelig på lærers subjektive vurderinger supplert med resultater fra standardiserte kartleggingsprøver for klassetrinnet. I det følgende vil både de praktiske mulighetene for å kunne bruke SCD i vurdering av spesialpedagogiske tiltak og ulike modeller for bruk bli diskutert.

Erfaringer fra studie 2 (artikkel 2) viser at det er praktisk mulig for PPT/skole å gjennomføre repeterte målinger av en elev både før, under og etter at et spesialpedagogisk tiltak er iverksatt. Eksemplet viser også at skoler og lærere kan være villige til å samarbeide om å prøve ut tiltak etter fastsatte retningslinjer og at skolene er i stand til å omdisponere ressurser for å realisere utprøving av nye tiltak i spesialundervisningen. Metoder for vurdering av effekt fra SCD er også lett tilgjengelige gjennom programvare som for eksempel Microsoft

Excel og nettbaserte kalkulatorer for de ulike effektmålene. Infrastrukturen for et godt samarbeid rundt evaluering av spesialpedagogiske tiltak ved bruk av SCD er derfor til stede.

I artikkel 2 skisseres det også en to trinns modell for bruk av SCD i skolene. Det argumenteres for at selv enkle tofase design, altså design som ikke fyller kravene til å være eksperimentelle i henhold til gjeldende standarder (Tate et al., 2016), likevel kan være funksjonelle som evalueringsverktøy når det gjelder tiltak for enkeltelever (Shadish, 2014). Videre diskuteres det hvordan særlig pedagogisk psykologisk tjeneste kan bruke SCD til å forske i egen praksis. For et slikt formål må den eksperimentelle kontrollen økes ved at effekten repliseres gjennom flere enn to faser (ABAB design), eventuelt gjennom replisering i flere situasjoner (skolefag) (multiple baseline design) (Tate et al., 2016). Brukt slik og ved å også replisere resultatene på flere elever, vil SCD i skolen kunne bidra til forskningsbasert kunnskap i en praksisnær kontekst. Siden repetert testing, som vist, kan innebære en viss praksiseffekt, bør vurdering av effekt på bakgrunn av gjentatte mål av den samme ferdigheten ta denne risikoen med i betraktning. En måte å gjøre det på kan være å bruke å bruke eksterne kontrollmål slik det er gjort i denne studien. Andre alternativer kan være å unngå å bruke repetert testing som avhengig variabel og i stedet operasjonalisere avhengig variabel utfra elevens atferd i arbeid med konkrete undervisningsoppgaver.

### **7. 2. 1 Hva virker og hvordan vurdere det?**

I denne avhandlingen har det blitt lagt vekt på å vise at spørsmål om hva som virker i praksis kan besvares fra ulike perspektiver. Dette er eksemplifisert gjennom to studier som begge er relatert til konkrete problemstillinger om hva som virker i praksis.

I studie 1 som angikk en relativt stor gruppe elever og der målsettingen var å vurdere et nytt tiltak som kommunen eventuelt ville kunne gjenta året etter dersom det viste seg å ha effekt, var det relevant å velge et randomisert kontrollert gruppedesign som reflekterte populasjonens og utvalgets heterogenitet. Det skoleadministrasjonen ønsket svar på var hvor sannsynlig det var at gjennomføring av det valgte tiltaket i den avgrensede perioden som var til rådighet ville øke deltakernes leseferdigheter i større grad enn dersom tiltaket ikke hadde blitt gjennomført. Studien ga et klart negativt svar på dette spørsmålet (artikkel 1). Designet som ble valgt og de valgte analysemetodene var egnet til å besvare den enkle, men relevante problemstillingen som var utgangspunktet for studien. Likevel vil det i en slik kontrollert gruppestudie også kunne være relevant informasjon som går tapt og som også delvis skjules

av at vurderingene tar utgangspunkt i gjennomsnitt på gruppenivå (Lobo et al., 2017; Onghena et al. 2018). Selv om konklusjonen fra studie 1 basert på de gjennomsnittlige resultatene konkluderer med at tiltaket ikke hadde effekt, er det også noen elever i dette utvalget som beviselig hadde god effekt av tiltaket. Dette fanges imidlertid ikke opp av designet fordi vurderingen baseres på gjennomsnittet.

I studie 2 var utvalget av elever begrenset til 13 elever som alle hadde alvorlig grad av dysleksi, men som også var ulike både med tanke på historikk, alder, bosted osv. Også her var oppdraget å vurdere om det valgte tiltaket virket. Valg av SCD ble gjort utfra elevenes sammensatte vanskeprofil og at det ble vurdert som både urealistisk og uetisk å rekruttere et kontrollutvalg som kunne matche disse 13 elevene. Hver av elevene i studie 2 ble derfor primært sammenliknet med seg selv gjennom at målingene i baselinefasen fungerte som kontroll. Som nevnt kan denne typen design være godt egnet for å vurdere hvorvidt et tiltak virker innenfor den konkrete rammen det er ment å fungere innenfor (artikkel 2). SCD kan også gi svar på mange praktiske spørsmål knyttet til kausalitet ved at sammenhengen mellom intervensjonen og tiltaket måles gjentatte ganger over tid. Slike spørsmål kan angå hvor lang tid man må trene før man ser effekt, hvorvidt pauser i treningen fører til nedgang i skåre samt hvor lenge tiltaket bør vare for å oppnå varig endring av målt atferd.

Selv om repeterte kartlegginger gir mulighet til å få fram detaljer knyttet til kausalitet er det også feilkilder knyttet til repeterte målinger, særlig dersom det som måles innebærer en gradvis tilegnelse eller endring av akademiske ferdigheter, som for eksempel lesing. Dette har som nevnt spesielt sammenheng med at mulig praksiseffekt knyttet til målingene kan svekke den indre validiteten (artikkel 3). Dette betyr at selv om det skjer en endring i positiv retning i den perioden intervensjonen varer er det ikke sikkert at denne endringen kun kan tilskrives effekt av det igangsatte tiltaket, siden resultatene også påvirkes av at elevene har blitt kjent med rutinene for kartleggingen og oppgavene de testes med. Dersom målet er å vurdere effekten av et tiltak isolert sett er dette en viktig begrensning.

I en praktisk kontekst gir imidlertid SCD mulighet til å ikke bare å vurdere om et tiltak virker, men også til å bringe klarhet i detaljer rundt de betingelsene som sammen fører til at tiltaket virker (Cartwright, 2013; Onghena et al., 2018). Dette gjør designet spesielt egnet for vurdering av effekt av tiltak knyttet til tiltak for enkeltelever (artikkel 2). Siden betingelsene langt på vei vil være lokalt definert og spesielle for hver situasjon er resultatene vanskelige å

generalisere og det er derfor kun gjennom replikasjoner at effekten av et tiltak med SCD kan regnes som gyldig i et mer generalisert perspektiv (Maggin et al., 2018). Det er som nevnt utarbeidet kriterier for hvordan generalisert effekt av SCD skal vurderes, for eksempel gjennom «3-5-20 regelen» (Zimmerman et al., 2018). Den videre utviklingen av retningslinjer for vurdering av samlet deltakereffekt fra enkeltdesign innenfor SCD, som for eksempel BC-SMD, kan være lovende med tanke på å utarbeide enklere metoder for å vurdere generalisert effekt av SCD (WWC 2020). Bruk av dette effektmålet er også eksemplifisert i artikkel 3.

Når det gjelder å vurdere virkningen av pedagogiske tiltak har RCT og SCD både ulikt perspektiv og hver sine primære bruksområder (Odom et al., 2005). Disse kan i praksis også komplette hverandre:

- SCD kan brukes til pilotering før man eventuelt vurderer å prøve ut et tiltak med et større og mer kostnadskrevenende RCT design.
- RCT kan gjennom å vurdere gjennomsnittseffekten og sannsynligheten for at tiltak virker i et representativt utvalg inspirere til videre bruk av tiltaket i nye utvalg.
- SCD kan brukes for å vurdere hvorvidt et igangsatt tiltak faktisk virker for den enkelte og kan gjennom replikasjoner også bidra til dokumentasjon i et mer generelt perspektiv.

### 7.3 Avsluttende konklusjoner

En konkret pedagogisk lærdom fra dette prosjektet er at det er praktisk mulig å gjennomføre SCD i evaluering av spesialpedagogiske tiltak etter enkeltvedtak. Dette bør kunne være et verdifullt tilskudd til å øke kvaliteten når det gjelder å vurdere utbyttet av igangsatte spesialpedagogiske tiltak for enkeltelever i skolen.

Når det gjelder SCD som verktøy for å frambringe forskningsbasert kunnskap fra praksisfeltet bør videre studier undersøke hvordan avhengige variabler best kan operasjonaliseres. Dette særlig med tanke på å sikre validiteten både når det gjelder å unngå mulig testeffekt av repeterte målinger og samtidig oppnå best mulig eksperimentell kontroll med andre typer ytre påvirkning. I studie 2 ble det valgt et relativt stringent mål på avkodingsferdighet (ett minutt ordlesing av ord med ulik struktur). Dette valget var primært inspirert fra tidligere gruppestudier med pre-post design innenfor lesefeltet. Denne

målemetoden ble i studie 2 tilpasset det valgte designet ved at ordlistene ble konstruert spesielt med tanke på å motvirke praksiseffekt av repeterte målinger. Til tross for disse tiltakene for å sikre validiteten indikerer resultatene på pegboard-testen at det likevel forekom noe praksiseffekt og at dette svekker grunnlaget for å vurdere effekten (artikkel 3).

Mulig effekt av repeterte målinger er en utfordring, særlig i SCD som bruker testmål av lærte ferdigheter som avhengig variabel. Erfaringene fra studie 2 viser også at denne utfordringen er vanskelig å imøtekomme selv om man aktivt forsøker å motvirke testeffekt gjennom å konstruere likeverdige oppgaver og alternere disse i randomisert rekkefølge.

En alternativ innfallsvinkel kunne vært å bruke mindre stringente mål og i stedet foretatt en mer kvalitativ vurdering av endringene gjennom de tre fasene. Repeterte mål mer direkte relatert til anvendt effekt av treningen på konkrete områder ville trolig vært mindre sårbare for praksiseffekt knyttet til repetert testing, men samtidig mulig vært mer sårbare for andre trusler mot indre validitet. Alternative mål som for eksempel antall ord lest korrekt i tekster med likeverdige vanskegrad og omfang av korrekt stavede ord ved tekstproduksjon ville gitt direkte informasjon om hvordan treningen virket inn på elevens funksjon i praktiske lese og skrive oppgaver. Datainnsamlingen ville også vært enklere da man ikke hadde behovd å gjennomføre standardiserte testprosedyrer, men i stedet kunne basert vurderingene på ukentlige innsamlede elevarbeider. Samtidig ville en slik innfallsvinkel mulig også hatt feilkilder for eksempel knyttet til at bakgrunnskunnskap i forhold til innholdet i tekstene mulig kunnet spille inn på motivasjonen og dermed også på utfallet av målingene.

Videre studier bør vurdere både den praktiske nytteverdien og validiteten ved bruk av denne typen enklere mål når det gjelder SCD studier der målet er varig endring av akademiske ferdigheter og veie dette opp imot risikoen for testeffekt ved mer stringente mål som for eksempel repetert ordlesing.

Når det gjelder bruk av artikulatorkisk bevissthetstrening som virkemiddel for å styrke fonologisk bevissthet og avkodingsferdighet gir prosjektet som helhet divergerende resultater. I studie 1 ga, som nevnt, ikke denne typen trening bedre resultater enn den tilretteleggingen kommunen allerede hadde iverksatt for å imøtekomme behovet til elever i første klasse med forsinket leseutvikling. Under forutsetning av at elevene i kontrollgruppene også fikk godt tilpasset fonologisk basert undervisning (artikkel 1), bekrefter resultatene tidligere studier i liknende utvalg som har funnet at trening med fokus på artikulatorkisk bevissthet er like



effektiv som trening som bruker mer tradisjonelle lyttebaserte tilnærminger når det gjelder utvikling av fonologisk bevissthet og grunnleggende avkodingsferdigheter (Castiglioni-Spalten & Ehri, 2003; Torgesen et al., 2010).

I studie 2 har et flertall av elevene hatt godt utbytte av intervensjonen. Selv om denne tolkingen må ta forbehold om at en mulig testeffekt kan ha spilt inn på resultatene, synes dette ikke å være hele forklaringen på de endringene som ble funnet (artikkel 3). Med nevnte forbehold kan resultatene sies å bekrefte resultater fra tidligere studier som har funnet at elever med spesielt svake fonologiske ferdigheter synes å profitere mest på tiltak knyttet til artikulatorisk bevissthet i leseopplæringen (Fälth et al. 2017; Joly-Pottuz et al., 2008; Rosenthal & Ehri, 2011; Trainin et al., 2014).

Samtidig viste resultatene fra studie 2 at ikke alle elevene hadde like godt utbytte av opplæringen. Resultatene viste også at effekten av treningen syntes å være bedre på mål knyttet til fonologisk avkoding og staving enn for avkoding av irregulære ord (artikkel 3). Dette kan indikere at elever med en unøyaktig fonologisk avkodingsstrategi kan ha mer å hente med denne typen tilrettelegging sammenliknet med elever med dysleksi der vanskene primært er knyttet til lav hastighet i avkodingen. Siden utvalget i studie 2 er begrenset til kun 11 deltakere er det imidlertid ikke mulig å trekke generaliserte slutninger rundt sammenhengen mellom individuelle forutsetninger og utbytte av artikulatorisk støtte i leseopplæringen. I tråd med en multifaktoriell forståelsesramme for dysleksi bør derfor videre studier fokusere på sammenhengen mellom effekt av tiltak med artikulatorisk bevissthetstrening og individuelle vanskeprofiler og årsakssammenhenger i et bredere format.

## Litteratur

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. Arlington: American Psychiatric Association; 2013.

Anthony, J. , Williams, J. , Aghara, R. , Dunkelberger, M. , Novak, B. , & Mukherjee, A. (2010). Assessment of individual differences in phonological representation. *An Interdisciplinary Journal*, 23 (8), 969-994. doi:10. 1007/s11145-009-9185-7

Baer, D. M. , Wolf, M. M. , & Risley, T. R. (1968). SOME CURRENT DIMENSIONS OF APPLIED BEHAVIOR ANALYSIS 1. *Journal of Applied Behavior Analysis*, 1 (1), 91-97. doi:10. 1901/jaba. 1968. 1-91

Barr, W. B. (2002). Neuropsychological testing for assessment of treatment effects: methodologic issues. *CNS Spectr*, 7 (4), 300-302, 304-306. doi:10. 1017/s1092852900017715

Bishop, D. V. M. , & Snowling, M. J (2004). Developmental Dyslexia and Specific Language Impairment: Same or Different? *Psychological Bulletin*, 130 (6), 858-886. doi:10. 1037/0033-2909. 130. 6. 858

Blomert, L. , Mitterer, H. , & Paffen, C (2004). In Search of the Auditory, Phonetic, And/or Phonological Problems in Dyslexia: Context Effects in Speech Perception. *Journal of Speech, Language, and Hearing Research*, 47 (5), 1030. doi:10. 1044/1092-4388 (2004/077)

Boada, R. , & Pennington, B. F (2006). Deficient implicit phonological representations in children with dyslexia. *Journal of Experimental Child Psychology*, 95 (3), 153-193. doi:10. 1016/j. jecp. 2006. 04. 003

Bogliotti, C. , Serniclaes, W. , Messaoud-Galusi, S. , & Sprenger-Charolles, L (2008). Discrimination of speech sounds by children with dyslexia: Comparisons with chronological age and reading level controls. *Journal of Experimental Child Psychology*, 101 (2), 137-155. doi:http://dx. doi. org/10. 1016/j. jecp. 2008. 03. 006

Boyer, N. , & Ehri, L. C. (2011). Contribution of Phonemic Segmentation Instruction With Letters and Articulation Pictures to Word Reading and Spelling in Beginners. *Scientific Studies of Reading*, 15 (5), 440-470. doi:10. 1080/10888438. 2010. 520778

Bus, A. G. , & Ijzendoorn, M. H. V (1999). Phonological Awareness and Early Reading: A Meta-Analysis of Experimental Training Studies. *Journal of Educational Psychology*, 91 (3), 403-414. doi:10. 1037/0022-0663. 91. 3. 403

Busk, P. L. , & Serlin, R. C. (1992). Meta-analysis for single-case research. In *Single-case research design and analysis: New directions for psychology and education*. (pp. 187-212). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Byrne, B. , Olson, R. K. , Samuelsson, S. , Wadsworth, S. , Corley, R. , DeFries, J. C. , & Willcutt, E (2006). Genetic and Environmental Influences on Early Literacy. *Journal of Research in Reading*, 29 (1), 33-49. doi:10. 1111/j. 1467-9817. 2006. 00291. x

Cain, K. , Catts, H. , Hogan, T. , & Lomax, R (2015). *Learning to read: Should we keep things simple?* (Vol. 50).

Calcus, A. , Lorenzi, C. , Collet, G. , Colin, C. , & Kolinsky, R (2016). Is There a Relationship Between Speech Identification in Noise and Categorical Perception in Children With Dyslexia? *Journal of Speech, Language and Hearing Research (Online)*, 59 (4), 835-852. doi:10. 1044/2016\_JSLHR-H-15-0076

Caravolas, M. , Volín, J. , & Caravolas, M (2001). Phonological spelling errors among dyslexic children learning a transparent orthography: the case of Czech. *Dyslexia (Chichester, England)*, 7 (4), 229-245. doi:10. 1002/dys. 206

Carr, J. E. (2005). Recommendations for reporting multiple-baseline designs across participants. *Behavioral Interventions*, 20 (3), 219-224. doi:10. 1002/bin. 191

Cartwright, N. (2013). Knowing what we are talking about: why evidence doesn't always travel. *Evidence & Policy: A Journal of Research, Debate and Practice*, 9 (1), 97-112. Retrieved from <http://www. ingentaconnect. com/content/tpp/ep/2013/00000009/00000001/art00008>

Castiglioni-Spalten, M. L. , & Ehri, L. C. (2003). Phonemic Awareness Instruction: Contribution of Articulatory Segmentation to Novice Beginners' Reading and Spelling. *Scientific Studies of Reading*, 7 (1), 25-52. doi:10. 1207/S1532799XSSR0701\_03

Catts, H. W. , & Vaughn, S (2018). The Simple View of Reading: Advancements and False Impressions. *Remedial and Special Education*, 39 (5), 317-323. doi:10.

1177/0741932518767563

Catts, H. W. , Adlof, S. M. , Hogan, T. P. , & Weismer, S. E (2005). Are Specific Language Impairment and Dyslexia Distinct Disorders? *Journal of Speech, Language, and Hearing Research*, 48 (6), 1378. doi:10. 1044/1092-4388 (2005/096)

Christ, T. J. (2007). Experimental control and threats to internal validity of concurrent and nonconcurrent multiple baseline designs. *Psychology in the Schools*, 44 (5), 451-459. doi:10. 1002/pits. 20237

Christopher, M. E. , Hulslander, J. , Byrne, B. , Samuelsson, S. , Keenan, J. M. , Pennington, B. , . . . Olson, R. K (2013). Modeling the Etiology of Individual Differences in Early Reading Development: Evidence for Strong Genetic Influences. *Scientific Studies of Reading*, 17 (5), 350-368. doi:10. 1080/10888438. 2012. 729119

Clayton, F. J. , West, G. , Sears, C. , Hulme, C. , & Lervåg, A. (2020). A Longitudinal Study of Early Reading Development: Letter-Sound Knowledge, Phoneme Awareness and RAN, but Not Letter-Sound Integration, Predict Variations in Reading Development. *Scientific Studies of Reading*, 24 (2), 91-107. doi:10. 1080/10888438. 2019. 1622546

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed. ed.). Hillsdale, N. J: Laurence Erlbaum.

Coltheart, M. , Rastle, K. , Perry, C. , Langdon, R. , & Ziegler, J (2001). DRC: A Dual Route Cascaded Model of Visual Word Recognition and Reading Aloud. *Psychological Review*, 108 (1), 204-256. doi:10. 1037/0033-295X. 108. 1. 204

Cumming, R. , Wilson, A. , & Goswami, U (2015). Basic auditory processing and sensitivity to prosodic structure in children with specific language impairments: a new look at a perceptual hypothesis. *Frontiers in psychology*, 6, 972. doi:10. 3389/fpsyg. 2015. 00972

de Bruin, C. L. (2017). Conceptualizing effectiveness in disability research. *International Journal of Research & Method in Education*, 40 (2), 113-136. doi:10. 1080/1743727X. 2015. 1033391

Deaton, A. , & Cartwright, N (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, 210, 2-21. doi:10. 1016/j. socscimed. 2017. 12. 005

Duranovic, M. , & Sehic, S (2013). The Speed of Articulatory Movements Involved in Speech Production in Children with Dyslexia. *Journal of Learning Disabilities*, 46 (3), 278-286. doi:10. 1177/0022219411419014

Ehri, L. C (1995). Phases of Development in Learning to Read Words by Sight. *Journal of Research in Reading*, 18 (2), 116-125. doi:10. 1111/j. 1467-9817. 1995. tb00077. x

Ehri, L. C. (2014). Orthographic Mapping in the Acquisition of Sight Word Reading, Spelling Memory, and Vocabulary Learning. *Scientific Studies of Reading*, 18 (1), 5-21. Retrieved from <https://doi.org/10.1080/10888438.2013.819356>. doi:10.1080/10888438.2013.819356

Ehri, L. C. , Nunes, S. R. , Willows, D. M. , Schuster, B. V. , Yaghoub-Zadeh, Z. , & Shanahan, T (2001). Phonemic Awareness Instruction Helps Children Learn To Read: Evidence from the National Reading Panel's Meta-Analysis. *Reading Research Quarterly*, 36 (3), 250-287. doi:10. 1598/RRQ. 36. 3. 2

Eklund, K. , Torppa, M. , Aro, M. , Leppänen, P. H. T. , & Lyytinen, H (2015). Literacy Skill Development of Children With Familial Risk for Dyslexia Through Grades 2, 3, and 8. *Journal of Educational Psychology*, 107 (1), 126-140. doi:10. 1037/a0037121

Elbro, C. , & Jensen, M. N (2005). Quality of phonological representations, verbal learning, and phoneme awareness in dyslexic and normal readers. *Scandinavian Journal of Psychology*, 46 (4), 375-384. doi:10. 1111/j. 1467-9450. 2005. 00468. x

Elbro, C. , Petersen, D. K. , & Borstrom, I (1998). Predicting Dyslexia from Kindergarten: The Importance of Distinctness of Phonological Representations of Lexical Items. *Reading Research Quarterly*, 33 (1), 36-60. doi:10. 1598/RRQ. 33. 1. 3

Elwér, Å. , Keenan, J. , Olson, R. , Byrne, B. , & Samuelsson, S (2013). Longitudinal stability and predictors of poor oral comprehenders and poor decoders. *Journal of experimental child psychology (Print)*, 115 (3), 497-516. doi:10. 1016/j. jecp. 2012. 12. 001

- Fawcett, A. J. , & Nicolson, R. I. (2002). Children with dyslexia are slow to articulate a single speech gesture. *Dyslexia*, 8 (4), 189-203. doi:10. 1002/dys. 222
- Foy, J. , & Mann, V. (2012). Speech production deficits in early readers: predictors of risk. *An Interdisciplinary Journal*, 25 (4), 799-830. doi:10. 1007/s11145-011-9300-4
- Frost, J (2003). *Prinsipper for god leseopplæring : innføring i den første lese- og skriveopplæringen*. Oslo: Cappelen akademisk forl.
- Fälth, L. , Gustafson, S. , & Svensson, I. (2017). Phonological Awareness Training with Articulation Promotes Early Reading Development. *Education*, 137 (3), 261-276.
- Galuschka, K. , Ise, E. , Krick, K. , & Schulte-Korne, G (2014). Effectiveness of Treatment Approaches for Children and Adolescents with Reading Disabilities: A Meta-Analysis of Randomized Controlled Trials (Research Article). *PLoS ONE*, 9 (2), e89900. doi:10. 1371/journal. pone. 0089900
- Gast, D. (2014). General Factors in Measurement and Evaluation. In Gast, D. L. , & Ledford, J. R. *Single case research methodology : applications in special education and behavioral sciences (2nd ed. ed.)*. New York: Routledge. 85-105.
- Gast, D. L. , & Ledford, J. R. (2014). *Single case research methodology : applications in special education and behavioral sciences (2nd ed. ed.)*. New York: Routledge.
- Godøy, O. R. , & Monsrud, M. -B (2011). *Spesialpedagogisk leseopplæring : en veileder ([Rev. utg. ]. ed.)*. Oslo: Bredtvet kompetansesenter.
- Goswami, U (2014). Sensory theories of developmental dyslexia: three challenges for research. *Nature Reviews Neuroscience*, 16 (1). doi:10. 1038/nrn3836
- Gough, P. B. , & Tunmer, W. E. (1986). Decoding, Reading, and Reading Disability. *Remedial and Special Education*, 7 (1), 6-10. doi:10. 1177/074193258600700104
- Griffiths, S. , & Frith, U. (2002). Evidence for an articulatory awareness deficit in adult dyslexics. *Dyslexia*, 8 (1), 14-21. doi:10. 1002/dys. 201
- Hayiou-Thomas, M. E. , Carroll, J. M. , Leavett, R. , Hulme, C. , & Snowling, M. J (2017). When does speech sound disorder matter for literacy? The role of disordered speech errors,

co-occurring language impairment and family risk of dyslexia. *Journal of Child Psychology and Psychiatry*, 58 (2), 197-205. doi:10. 1111/jcpp. 12648

Hayward, D. V. , Annable, C. D. , Fung, J. E. , Williamson, R. D. , Lovell-Johnston, M. A. , & Phillips, L. M (2017). Beyond the Total Score: A Preliminary Investigation into the Types of Phonological Awareness Errors Made by First Graders. *Early Childhood Education Journal*, 45 (6), 809-820. doi:10. 1007/s10643-016-0829-3

Hedges, L. V (2008). What Are Effect Sizes and Why Do We Need Them? *Child Development Perspectives*, 2 (3), 167-171. doi:10. 1111/j. 1750-8606. 2008. 00060. x

Hedges, L. V. , Pustejovsky, J. E. , & Shadish, W. R (2012). A standardized mean difference effect size for single case designs. *Research Synthesis Methods*, 3 (3), 224-239. doi:10. 1002/jrsm. 1052

Hitchcock, J. , Kratochwill, T. , & Chezan, L. (2015). What Works Clearinghouse Standards and Generalization of Single-Case Design Evidence. *Journal of Behavioral Education*, 24 (4), 459-469. doi:10. 1007/s10864-015-9224-1

Hjetland, H. N. , Brinchmann, E. I. , Scherer, R. , & Melby-Lervåg, M (2017). Preschool predictors of later reading comprehension ability: a systematic review

Hjetland, H. N. , Lervåg, A. , Lyster, S. -A. H. , Hagtvet, B. E. , Hulme, C. , & Melby-Lervåg, M (2018). Pathways to Reading Comprehension: A Longitudinal Study From 4 to 9 Years of Age. *Journal of Educational Psychology*. doi:10. 1037/edu0000321.

Hoonhorst, I. , Colin, C. , Markessis, E. , Radeau, M. , Deltenre, P. , & Serniclaes, W (2009). French native speakers in the making: From language-general to language-specific voicing boundaries. *Journal of Experimental Child Psychology*, 104 (4), 353-366. doi:10. 1016/j. jecp. 2009. 07. 005

Horlyck, S. , Reid, A. , & Burnham, D (2012). The Relationship between Learning to Read and Language-Specific Speech Perception: Maturation versus Experience. *Scientific Studies of Reading*, 16 (3), 218-239. doi:10. 1080/10888438. 2010. 546460

Horner, R. D. , & Baer, D. M (1978). MULTIPLE-PROBE TECHNIQUE: A VARIATION OF THE MULTIPLE BASELINE 1. *Journal of Applied Behavior Analysis*, 11 (1), 189-196. doi:10. 1901/jaba. 1978. 11-189

Hulme, C. , & Snowling, M. (2009). *Developmental disorders of language learning and cognition*. Chichester: Wiley-Blackwell.

Hulme, C. , Nash, H. M. , Gooch, D. , Lervåg, A. , & Snowling, M. J. (2015). The Foundations of Literacy Development in Children at Familial Risk of Dyslexia. *Psychological Science*, 26 (12), 1877-1886. doi:10. 1177/0956797615603702

Hulme, C. , Snowling, M. , Caravolas, M. , & Carroll, J. (2005). Phonological Skills Are (Probably) One Cause of Success in Learning to Read: A Comment on Castles and Coltheart. *Scientific Studies of Reading*, 9 (4), 351-365. doi:10. 1207/s1532799xssr0904\_2

Hulme, J. C. , & Snowling, J. M (2016). Reading disorders and dyslexia. *Current Opinion in Pediatrics*, 28 (6), 731-735. doi:10. 1097/MOP. 0000000000000411

Hämäläinen, J. A. , Salminen, H. K. , & Leppänen, P. H. T (2013). Basic Auditory Processing Deficits in Dyslexia: Systematic Review of the Behavioral and Event-Related Potential/ Field Evidence. *Journal of Learning Disabilities*, 46 (5), 413-427. doi:10. 1177/0022219411436213

Høien, T. , & Lundberg, I (1992). *Dysleksi* (2. utg. ed.). Oslo: Ad notam Gyldendal.

Joly-Pottuz, B., Mercier, M., Leynaud, A., & Habib, M. (2008). Combined auditory and articulatory training improves phonological deficit in children with dyslexia.

*Neuropsychological Rehabilitation*, 18(4), 402-429. doi:10.1080/09602010701529341

Juul, H. , Poulsen, M. , & Elbro, C (2014). Separating Speed from Accuracy in Beginning Reading Development. *Journal of Educational Psychology*, 106 (4), 1096-1106. doi:10. 1037/a0037100

Kausrud, T. (2003). PAS - steget inn i alfabetisk lesning : bruk av piktografisk-artikulatoriske symboler for utvikling av fonologisk bevissthet. In. Oslo: T. Kausrud.

Kim, Y. -S. , Wagner, R. K. , & Lopez, D (2012). Developmental Relations between Reading Fluency and Reading Comprehension: A Longitudinal Study from Grade 1 to Grade 2.



Journal of Experimental Child Psychology, 113 (1), 93-111. doi:10. 1016/j. jecp. 2012. 03. 002

Klinkenberg, J. E. & E. Skaar (2001). STAS. Standardisert Test for Avkoding og Staving.

Krasny-Pacini, A. , & Evans, J. (2018). Single-case experimental designs to assess intervention effectiveness in rehabilitation: A practical guide. *Annals of Physical and Rehabilitation Medicine*, 61 (3), 164-179. Retrieved from <http://www. sciencedirect. com/science/article/pii/S1877065717304542>. doi:<https://doi. org/10. 1016/j. rehab. 2017. 12. 002>

Kratochwill, T. R. , & Levin, J. R. (2014). Meta- and statistical analysis of single-case intervention research data: Quantitative gifts and a wish list. *Journal of School Psychology*, 52 (2), 231-235. doi:10. 1016/j. jsp. 2014. 01. 003

Kratochwill, T. R. , Hitchcock, J. H. , Horner, R. H. , Levin, J. R. , Odom, S. L. , Rindskopf, D. M. , & Shadish, W. R. (2013). Single-Case Intervention Research Design Standards. *Remedial and Special Education*, 34 (1), 26-38. doi:10. 1177/0741932512452794

Kvernbekk, T. (2018). Evidensbasert praksis: Utvalgte kontroverser. *Nordisk tidsskrift for pedagogikk og kritikk*, 4 (0). doi:10. 23865/ntpk. v4. 1153

Landerl, K. , Ramus, F. , Moll, K. , Lyytinen, H. , Leppanen, P. H. T. , Lohvansuu, K. , . . . Schulte-Korne, G (2013). Predictors of Developmental Dyslexia in European Orthographies with Varying Complexity. *Journal of Child Psychology and Psychiatry*, 54 (6), 686-694. doi:10. 1111/jcpp. 12029

Lane, J. D. , & Gast, D. L. (2014). Visual analysis in single case experimental design studies: Brief review and guidelines. *Neuropsychological Rehabilitation*, 24 (3-4), 445-463. doi:10. 1080/09602011. 2013. 815636

Ledford, J. R. (2018). No Randomization? No Problem: Experimental Control and Random Assignment in Single Case Research. *American Journal of Evaluation*, 39 (1), 71-90. doi:10. 1177/1098214017723110

Ledford, J.R, Wolery, M. & Gast, D.L (2014). Comtroversial and Critical Issues in Single Case Research. In Gast, D. L. , & Ledford, J. R. *Single case research methodology :*

applications in special education and behavioral sciences (2nd ed. ed.). New York: Routledge. 377-397.

Lervåg, A. , Hulme, C. , & Melby-Lervåg, M (2018). Unpicking the Developmental Relationship between Oral Language Skills and Reading Comprehension: It's Simple, but Complex. *Child Development*, 89 (5), 1821-1838. doi:10. 1111/cdev. 12861

Levine, A. J. , Miller, E. N. , Becker, J. T. , Selnes, O. A. , & Cohen, B. A. (2004). Normative Data for Determining Significance of Test-Retest Differences on Eight Common Neuropsychological Instruments. *The Clinical Neuropsychologist*, 18 (3), 373-384. doi:10. 1080/1385404049052420

Lewis, A. B. , Avrich, A. A. , Freebairn, A. L. , Taylor, G. H. , Iyengar, K. S. , & Stein, M. C. (2011). Subtyping Children With Speech Sound Disorders by Endophenotypes. *Topics in Language Disorders*, 31 (2), 112-127. doi:10. 1097/TLD. 0b013e318217b5dd

Liberman, A. M. , & Mattingly, I. G (1985). The motor theory of speech perception revised. *Cognition*, 21 (1), 1-36. doi:10. 1016/0010-0277 (85) 90021-6

Liberman, A. M. , & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, 4 (5), 187-196. doi:10. 1016/S1364-6613 (00)01471-6

Liberman, I. Y. , Shankweiler, D. , & Liberman, A. M (1989). The Alphabetic Principle and Learning To Read. In.

Lobo, A. M. , Moeyaert, A. M. , Baraldi Cunha, A. A. , & Babik, A. I (2017). Single-Case Design, Analysis, and Quality Assessment for Intervention Research. *Journal of Neurologic Physical Therapy*, 41 (3), 187-197. doi:10. 1097/NPT. 0000000000000187

Lohr, S. , Schochet, P. Z. , Sanders, E. , Westat, I. , Mathematica Policy Research, I. , & National Center for Education, R. (2014). Partially Nested Randomized Controlled Trials in Education Research: A Guide to Design and Analysis. NCER 2014-2000. In: National Center for Education Research.

Lohvansuu, K. , Hämäläinen, J. A. , Ervast, L. , Lyytinen, H. , & Leppänen, P. H. T (2018). Longitudinal interactions between brain and cognitive measures on reading development from

6 months to 14 years. *Neuropsychologia*, 108, 6-12. doi:10. 1016/j. neuropsychologia. 2017. 11. 018

Lundberg, I. , Frost, J. , & Petersen, O. -P (1988). Effects of an Extensive Program for Stimulating Phonological Awareness in Preschool Children. *Reading Research Quarterly*, 23 (3), 263-284. doi:10. 1598/RRQ. 23. 3. 1

Lyon, G. R. , Shaywitz, S. E. , & Shaywitz, B. A (2003). A Definition of Dyslexia. *Annals of Dyslexia*, 53, 1-14.

Maggin, D. M. , Cook, B. G. , & Cook, L. (2018). Using Single-Case Research Designs to Examine the Effects of Interventions in Special Education. *Learning Disabilities Research & Practice*, 33 (4), 182-191. doi:10. 1111/ldrp. 12184

Maggin, D. M. , Lane, K. L. , Pustejovsky, J. E. (2017). Introduction to the Special Issue on Single-Case Systematic Reviews and Meta-Analyses. *Remedial and Special Education*, 38 (6), 323-330. doi:10. 1177/0741932517717043

Mann, V. & Foy, J (2007). Speech development patterns and phonological awareness in preschool children. *An Interdisciplinary Journal of The International Dyslexia Association*, 57 (1), 51-74. doi:10. 1007/s11881-007-0002-1

McArthur, G. , Eve, P. , Jones, K. , Banales, E. , Kohonen, S. , Anandakumar, T. , . . . Castles, A (2012). Phonics training for English-speaking poor readers. In *Cochrane Database Syst Rev* (Vol. 12).

McBride-Chang, C (1995). Phonological processing, speech perception, and reading disability: An integrative review. *Educational Psychologist*, 30 (3), 109-121. doi:10. 1207/s15326985ep3003\_2

McBride-Chang, C (1996). Models of Speech Perception and Phonological Processing in Reading. *Child Development*, 67 (4), 1836-1856. doi:10. 2307/1131735

McBride-Chang, C. , Tong, X. , Shu, H. , Wong, A. M. Y. , Leung, K. -W. , & Tardif, T (2008). Syllable, Phoneme, and Tone: Psycholinguistic Units in Early Chinese and English Word Recognition. *Scientific Studies of Reading*, 12 (2), 171-194. doi:10. 1080/10888430801917290

McCaffrey, R. J. , Ortega, A. , & Haase, R. F. (1993). Effects of repeated neuropsychological assessments. *Archives of Clinical Neuropsychology*, 8 (6), 519-524. doi:10. 1016/0887-6177 (93)90052-3

Melby-Lervåg, M. , Lyster, S. -A. H. , Hulme, C. (2012). Phonological Skills and Their Role in Learning to Read: A Meta-Analytic Review. *Psychological Bulletin*, 138 (2), 322-352. doi:10. 1037/a0026744

Moher, D. , Hopewell, S. , Schulz, K. F. , Montori, V. , Gøtzsche, P. C. , Devereaux, P. J. , . . . Altman, D. G. (2010). CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. *Journal of Clinical Epidemiology*, 63 (8), e1-37. doi:10. 1016/j. jclinepi. 2010. 03. 004

Moll, K. , Loff, A. , & Snowling, M. J. (2013). Cognitive Endophenotypes of Dyslexia. *Scientific Studies of Reading*, 17 (6), 385-397. doi:10. 1080/10888438. 2012. 736439

Montgomery, D. (1981). Do dyslexics have difficulty accessing articulatory information? *Psychological Research*, 43 (2), 235-243. doi:10. 1007/BF00309832

Montgomery, P. , Grant, S. , Mayo-Wilson, E. , Macdonald, G. , Michie, S. , Hopewell, S. , . . . Yaffe, J. (2018). Reporting randomised trials of social and psychological interventions: the CONSORT-SPI 2018 Extension. 19 (1), 407-407. doi:10. 1186/s13063-018-2733-1

Navarra, J. , & Soto-Faraco, S (2007). Hearing lips in a second language: visual articulatory information enables the perception of second language sounds. *An International Journal of Perception, Attention, Memory, and Action*, 71 (1), 4-12. doi:10. 1007/s00426-005-0031-5

Noordenbos, M. W. , & Serniclaes, W (2015). The Categorical Perception Deficit in Dyslexia: A Meta-Analysis. *Scientific Studies of Reading*, 19 (5), 340-359. doi:10. 1080/10888438. 2015. 1052455

Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. H., Thompson, B., & Harris, K. R. (2005). *Research in Special Education: Scientific Methods and Evidence-Based Practices*. *Exceptional Children*, 71(2), 137-148. Retrieved from <https://journals.sagepub.com/doi/abs/10.1177/001440290507100201>. doi:10.1177/001440290507100201

Olson, R. K. , Keenan, J. M. , Byrne, B. , Samuelsson, S. , Coventry, W. L. , Corley, R. , . . . Hulslander, J (2011). Genetic and Environmental Influences on Vocabulary and Reading Development. *Scientific Studies of Reading*, 15 (1), 26-46. doi:10. 1080/10888438. 2011. 536128

Onghena, P. , Michiels, B. , Jamshidi, L. , Moeyaert, M. , & Van Den Noortgate, W. (2018). One by One: Accumulating Evidence by using Meta-Analytical Procedures for Single-Case Experiments. *Brain Impair.* , 19 (1), 33-58. doi:10. 1017/BrImp. 2017. 25

Ottem, E. & Kausrud , T. (2001). Use of pictographic-articulatory symbols to promote alphabetic reading in a language-impaired boy: Case study. *Augmentative and Alternative Communication*, 17 (1), 52. doi:10. 1080/aac. 17. 1. 52. 60

Parker, R. I. , Vannest, K. J. , & Davis, J. L. (2011). Effect Size in Single-Case Research: A Review of Nine Nonoverlap Techniques. In (Vol. 35, pp. 303-322). Los Angeles, CA.

Parker, R. I. , Vannest, K. J. , Davis, J. L. , & Sauber, S. B. (2011). Combining Nonoverlap and Trend for Single-Case Research: Tau-U. *Behavior Therapy*, 42 (2), 284-299. doi:10. 1016/j. beth. 2010. 08. 006

Pennala, R. , Eklund, K. , Hämäläinen, J. , Richardson, U. , Martin, M. , Leiwo, M. , . . . Lyytinen, H (2010). Perception of Phonemic Length and Its Relation to Reading and Spelling Skills in Children With Family Risk for Dyslexia in the First Three Grades of School. *Journal of Speech, Language and Hearing Research (Online)*, 53 (3), 710-724. doi:10. 1044/1092-4388 (2009/08-0133)

Pennington, B. F (2006). From single to multiple deficit models of developmental disorders. *Cognition*, 101 (2), 385-413. doi:10. 1016/j. cognition. 2006. 04. 008

Pennington, B. F. , & Bishop, D. V. M. (2009). Relations among speech, language, and reading disorders. *Annual Review of Psychology*, 60, 283.

Perfetti, C. , Beck, I. , Bell, L. , Hughes, C. , & Perfetti, C (1987). Phonemic Knowledge and Learning to Read Are Reciprocal: A Longitudinal Study of First Grade Children. *Merrill-Palmer Quarterly*, 33 (3), 283-319.

- Poulsen, M. , & Elbro, C (2018). The short- and long-term predictions of reading accuracy and speed from paired-associate learning. *Journal of Experimental Child Psychology*, 174, 77-89. doi:10. 1016/j. jecp. 2018. 05. 007
- Preston, J. L. , Hull, M. , & Edwards, M. L (2013). Preschool Speech Error Patterns Predict Articulation and Phonological Awareness Outcomes in Children with Histories of Speech Sound Disorders. *American Journal of Speech-Language Pathology*, 22 (2), 173-Language Pathology, 2013, Vol. 2022 (2012), p. 2173-2184. doi:10. 1044/1058-0360 (2012/12-0022)
- Protopapas, A., Simos, P. G., Sideridis, G. D., & Mouzaki, A. (2012). The Components of the Simple View of Reading: A Confirmatory Factor Analysis. *Reading Psychology*, 33(3), 217-240. doi:10.1080/02702711.2010.507626
- Pustejovsky, J. E. (2019). "Procedural Sensitivities of Effect Sizes for Single-Case Designs With Directly Observed Behavioral Outcome Measures." **Psychological Methods** 24 (2019): 217–235.
- Pustejovsky, J. E. (2018). Using response ratios for meta-analyzing single-case designs with behavioral outcomes. *J. Sch. Psychol.* , 68, 99-112. doi:10. 1016/j. jsp. 2018. 02. 003
- Pustejovsky, J. E. , Hedges, L. V. , & Shadish, W. R. (2014). Design-Comparable effect sizes in multiple baseline designs: A general modeling framework. *Journal of Educational and Behavioral Statistics*, 39 (5), 368-393. doi:10. 3102/1076998614547577
- Ramus, F. , Marshall, C. R. , Rosen, S. , & van der Lely, H. K. J (2013). Phonological Deficits in Specific Language Impairment and Developmental Dyslexia: Towards a Multidimensional Model. *Brain*, 136 (2), 630-645. doi:10. 1093/brain/aws356
- Robertson, E. K. , Joanisse, M. F. , Desroches, A. S. , & Ng, S (2009). Categorical speech perception deficits distinguish language and reading impairments in children. *Developmental Science*, 12 (5), 753-767. doi:10. 1111/j. 1467-7687. 2009. 00806. X
- Rose, J. Identifying and teaching children and young people with dyslexia and literacy difficulties. 2009. ••: DCSF, DCSF- 00659-2009
- Rosenthal, J. , & Ehri, L. C. (2011). Pronouncing New Words Aloud during the Silent Reading of Text Enhances Fifth Graders' Memory for Vocabulary Words and Their Spellings.

Reading and Writing: An Interdisciplinary Journal, 24 (8), 921-950. doi:10. 1007/s11145-010-9239-x

Sato, M. , Troille, E. , Ménard, L. , Cathiard, M. -A. , & Gracco, V (2013). Silent articulation modulates auditory and audiovisual speech perception. *Experimental Brain Research*, 227 (2), 275-288. doi:10. 1007/s00221-013-3510-8

SBU (2014). Dyslexi hos barn och ungdomar- Tester och insatser. En systematisk litteraturöversikt. Stockholm: Statens beredning för medicinsk utvärdering (SBU), rapport nr 225. ISBN 978-91-85413-66-9

Scarborough, H. S. , & Dobrich, W. (1990). Development of children with early language delay. *Journal of Speech & Hearing Research*, 33 (1), 70–83. <https://doi.org/10.1044/jshr.3301.70>

Schaadt, G. , Männel, C. , Meer, E. , Pannekamp, A. , & Friederici, A. D (2016). Facial speech gestures: the relation between visual speech processing, phonological awareness, and developmental dyslexia in 10-year-olds. *Developmental Science*, 19 (6), 1020-1034. doi:10. 1111/desc. 12346

Seidenberg, M. S. (1993). A Connectionist Modeling Approach to Word Recognition and Dyslexia. *Psychological Science*, 4 (5), 299-304. doi:10. 1111/j. 1467-9280. 1993. tb00568. x

Sénéchal, M. , Ouellette, G. , & Young, L (2004). Testing the Concurrent and Predictive Relations among Articulation Accuracy, Speech Perception, and Phoneme Awareness. *Journal of Experimental Child Psychology*, 89 (3), 242-269. doi:10. 1016/j. jecp. 2004. 07. 005

Serniclaes, W. , Sprenger-Charolles, L. , Carre, R. , & Demonet, J. -F (2001). Perceptual Discrimination of Speech Sounds in Developmental Dyslexia. *Journal of Speech, Language, and Hearing Research*, 44 (2), 384-399. doi:10. 1044/1092-4388 (2001/032)

Shadish, W. R. (2014). Statistical Analyses of Single-Case Designs: The Shape of Things to Come. *Current Directions in Psychological Science*, 23 (2), 139-146. doi:10. 1177/0963721414524773

Shadish, W. R. , Cook, T. D. , & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Shadish, W. R. , Hedges, L. V. , Horner, R. H. , Odom, S. L. , National Center for Education, R. , Westat, I. , & National Center for Special Education, R. (2015). *The Role of Between-Case Effect Size in Conducting, Interpreting, and Summarizing Single-Case Research*. NCER 2015-002. In: National Center for Education Research.

Shadish, W. , & Sullivan, K. (2011). Characteristics of single-case designs used to assess intervention effects in 2008. *Behavior Research Methods*, 43 (4), 971-980. doi:10.3758/s13428-011-0111-y

Skjelfjord, V. J. (1987a). Phonemic Segmentation. An Important Subskill in Learning to Read: I. *Scandinavian Journal of Educational Research*, 31 (1), 41-57. doi:10.1080/0031383870310104

Skjelfjord, V. J. (1987b). Phonemic Segmentation. An Important Subskill in Learning to Read: II. *Scandinavian Journal of Educational Research*, 31 (2), 82-98. doi:10.1080/0031383870310203

Snowling, M. J (2008). Specific disorders and broader phenotypes: The case of dyslexia. *The Quarterly Journal of Experimental Psychology*, 61 (1), 142-156. doi:10.1080/17470210701508830

Snowling, M. J., & Hulme, C. (2012). Annual research review: the nature and classification of reading disorders-a commentary on proposals for DSM-5. *J Child Psychol Psychiatry*, 53(5), 593-607. doi:10.1111/j.1469-7610.2011.02495.x

Snowling, M. J. , & Melby-Lervåg, M. (2016). Oral Language Deficits in Familial Dyslexia: A Meta-Analysis and Review. *Psychological Bulletin*. doi:10.1037/bul0000037

Snowling, M. J. , Gallagher, A. , & Frith, U (2003). Family Risk of Dyslexia Is Continuous: Individual Differences in the Precursors of Reading Skill. *Child Development*, 74 (2), 358-373. doi:10.1111/1467-8624.7402003



Snowling, M. J. , Lervåg, A. , Nash, H. M. , & Hulme, C (2019). Longitudinal Relationships between Speech Perception, Phonological Skills and Reading in Children at High-Risk of Dyslexia. *Developmental Science*, 22 (1). doi:10. 1111/desc. 12723

Snowling, M. , Hulme, C. , Wells, B. , & Goulandris, N (1992). Continuities between speech and spelling in a case of developmental dyslexia. *An Interdisciplinary Journal*, 4 (1), 19-31. doi:10. 1007/BF01027070

Spear-Swerling, L. , & Sternberg, R. J (1996). *Off track : when poor readers become “learning disabled”*. Boulder, Colo: Westview Press.

Swan, D. , & Goswami, U (1997). Phonological Awareness Deficits in Developmental Dyslexia and the Phonological Representations Hypothesis. *Journal of Experimental Child Psychology*, 66 (1), 18-41. doi:10. 1016/S0022-0965 (02) 00169-8

Tallal, P (1980). Auditory temporal perception, phonics, and reading disabilities in children. *Brain and Language*, 9 (2), 182-198. doi:10. 1016/0093-934X (80) 90139-X

Tate, R. L. , Perdices, M. , Rosenkoetter, U. , Shadish, W. , Vohra, S. , Barlow, D. H. , . . . Wilson, B. (2016). The Single-Case Reporting guideline In BEhavioural interventions (SCRIBE) 2016 statement. *Journal of School Psychology*, 56, 133-142. doi:10. 1016/j. jsp. 2016. 04. 001

Tate, R. , Rosenkoetter, U. , Wakim, D. , Sigmundsdottir, L. , Doubleday, J. , Togher, L. , . . . Perdices, M. (2015). *The Risk-of-Bias in N-of-1 Trials (RoBiNT) Scale: An Expanded Manual for the Critical Appraisal of Single-Case Reports*.

Thomas, E. , & Sénéchal, M (2004). Long-term association between articulation quality and phoneme sensitivity: A study from age 3 to age 8. In.

Torgesen, J. K. , Alexander, A. W. , Wagner, R. K. , Rashotte, C. A. , Voeller, K. K. S. , & Conway, T. (2001). Intensive Remedial Instruction for Children with Severe Reading Disabilities: Immediate and Long-Term Outcomes from Two Instructional Approaches. *Journal of Learning Disabilities*, 34 (1), 33-58. doi:10. 1177/002221940103400104

Torgesen, J. K. , Wagner, R. K. & Rashotte, C. A. (1999). *Test Of Word Reading Efficiency (TOWRE)*, Pearson.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Herron, J., Lindamood, P., & Torgesen, J. K. (2010). Computer-assisted instruction to prevent early reading difficulties in students at risk for dyslexia: Outcomes from two instructional approaches. *Annals of Dyslexia*, 60(1), 40-56. doi:10.1007/s11881-009-0032-y

Torppa, M. , Tolvanen, A. , Poikkeus, A. -M. , Eklund, K. , Lerkkanen, M. -K. , Leskinen, E. , & Lyytinen, H (2007). Reading Development Subtypes and Their Early Characteristics. *Annals of Dyslexia*, 57 (1), 3. doi:10. 1007/s11881-007-0003-0

Trainin, G. , Wilson, K. M. , Murphy-Yagil, M. , & Rankin-Erickson, J. L. (2014). Taking a Different Route: Contribution of Articulation and Metacognition to Intervention with At-Risk Third-Grade Readers. *Journal of Education for Students Placed at Risk*, 19 (3-4), 183-184), p. 183-195. doi:10. 1080/10824669. 2014. 972103

Valentine, J. C. , Tanner-Smith, E. E. , Pustejovsky, J. E. , & Lau, T. S. (2016). Between-case standardized mean difference effect sizes for single-case designs: a primer and tutorial using the scdhlms web application. *Campbell Systematic Reviews*, 12 (1), 1-31. doi:10. 4073/cmdp. 2016.

Watson, P.J. , Workman, E.A (1981) The non-concurrent multiple baseline across-individuals design: An extension of the traditional multiple baseline design, *Journal of Behavior Therapy and Experimental Psychiatry*, Volume 12, Issue 3, Pages 257-259, ISSN 0005-7916, [https://doi.org/10.1016/0005-7916\(81\)90055-0](https://doi.org/10.1016/0005-7916(81)90055-0). (<http://www.sciencedirect.com/science/article/pii/0005791681900550>)

Wagner, R. K. , Edwards, A. A. , Malkowski, A. , Schatschneider, C. , Joyner, R. E. , Wood, S. , & Zirps, F. A (2019). Combining Old and New for Better Understanding and Predicting Dyslexia. *New Directions for Child and Adolescent Development*, 2019 (165), 11-23. doi:10. 1002/cad. 20289

What Works Clearinghouse (2017). Standards handbook , version 4. [https://www.google.com/search?q=What+Works+Clearinghouse+2017&rlz=1C1GCEV\\_en&oq=What+Works+Clearinghouse+2017&aqs=chrome..69i57j0.8671j0j7&sourceid=chrome&ie=UTF-8&safe=active&ssui=on](https://www.google.com/search?q=What+Works+Clearinghouse+2017&rlz=1C1GCEV_en&oq=What+Works+Clearinghouse+2017&aqs=chrome..69i57j0.8671j0j7&sourceid=chrome&ie=UTF-8&safe=active&ssui=on)

What Works Clearinghouse (2020). Procedures handbook , version 4.1.

- Wise, B. W. , Ring, J. , & Olson, R. K. (1999). Training phonological awareness with and without explicit attention to articulation. *Journal of Experimental Child Psychology*, 271.
- Winn, B., Skinner, C.H , Allin, J.D. & Hawkins, J. (2004) Practicing School Consultants Can Empirically Validate Interventions. *Journal of Applied School Psychology*, 20:2, 109-128, DOI: 10.1300/J370v20n02\_07
- Wolfe, K. , Dickenson, T. S. , Miller, B. , & McGrath, K. V. (2019). Comparing Visual and Statistical Analysis of Multiple Baseline Design Graphs. *Behavior Modification*, 43 (3), 361-388. doi:10. 1177/0145445518768723
- Wolff, U (2016). Effects of a Randomized Reading Intervention Study Aimed at 9-Year-Olds: A 5-Year Follow-up. *Dyslexia*, 22 (2), 85-100. doi:10. 1002/dys. 1529
- Zhang, J. , & McBride-Chang, C (2010). Auditory Sensitivity, Speech Perception, and Reading Development and Impairment. *Educational Psychology Review*, 22 (3), 323-338. doi:10. 1007/s10648-010-9137-4
- Zhang, J. , McBride-Chang, C. , & Eccles, J. S (2014). Auditory Sensitivity, Speech Perception, L1 Chinese, and L2 English Reading Abilities in Hong Kong Chinese Children. *Developmental Psychology*, 50 (4), 1001-1013. doi:10. 1037/a0035086
- Ziegler, J. C. , & Goswami, U (2005). Reading Acquisition, Developmental Dyslexia, and Skilled Reading Across Languages: A Psycholinguistic Grain Size Theory. *Psychological Bulletin*, 131 (1), 3-29. doi:10. 1037/0033-2909. 131. 1.
- Ziegler, J. , Perry, C. , Ma-Wyatt, A. , Ladner, D. , & Schulte-Korne, G (2003). Developmental dyslexia in different languages: Language-specific or universal? *Journal of Experimental Child Psychology*, 86 (3), 169-193. doi:10. 1016/S0022-0965 (03) 00139-5
- Zimmerman, K. N. , Ledford, J. R. , Severini, K. E. , Pustejovsky, J. E. , Barton, E. E. , & Lloyd, B. P. (2018). Single-case synthesis tools I: Comparing tools to evaluate SCD quality and rigor. *Research in Developmental Disabilities*, 79, 19-32. doi:http s://doi. org/10. 1016/j. ridd. 2018. 02. 003



## **Artikkel 1**

Thurmann-Moe, A.C., Melby-Lervåg, M & Lervåg, A. Effects of Articulatory Consciousness Training in First Graders with a Reading Delay: A Randomised Control Trial.



## **Effects of Articulatory Consciousness Training in First Graders with a Reading Delay: A Randomised Control Trial**

### **Abstract**

This paper describes a randomised control trial (RCT) of a small group intervention targeting articulatory consciousness in first graders with a reading delay (control condition:  $n = 57$ , intervention condition:  $n = 64$ ). The program in the intervention condition consisted of basic decoding and spelling tasks using articulatory cards to symbolise speech sounds. Training took place four times per week for five weeks. The control condition received “business as usual”. The results show that the children in the intervention condition learned to use the articulatory symbols, but there were no effects of the intervention on reading and phonological awareness measures. Also, there was no relationship shown between initial phonological awareness skills and gains, or lack thereof, from the intervention.

Key words: reading, articulation, first grade, intervention, RCT

## Effects of Articulatory Consciousness Training in First Graders with a Reading Delay: A Randomised Control Trial

The importance of the early identification of and early intervention in children with a reading delay is emphasised by both policy makers and schools to close the gap between at-risk children and typical readers. Because of this, in Norway, all first graders take a mandatory reading test after eight months of schooling. This national test is administrated in Norwegian, and the aim of the test is to identify children with a reading delay and to take early measures to prevent a negative trend in future learning trajectories. Here we evaluate the effects of an intervention programme aimed at children below the critical limit on this national test. The study was conducted in collaboration with the school administration in one municipality. The intervention was delivered in small groups and took place during the last weeks of the school year. The main goal of the intervention was to give the children beyond the critical limit a boost in reading and phonological awareness before the summer holiday.

### **Phoneme Awareness and Reading Instruction**

Numerous studies show that phonological awareness, or more specifically phoneme awareness, is an important causal factor in the development of decoding skills (Ehri et al., 2001; Melby-Lervåg, Lyster, & Hulme, 2012; Suggate, 2016). Therefore, phonological awareness training has become a central part of basic reading instruction over the last 30–40 years in most alphabetic orthographies, including in Norway.

Norwegian is a Germanic language and can be categorised with medium orthographic consistency compared to Finnish and Italian in the consistent end of the continuum and English and French in the inconsistent end of the continuum (Landerl et al., 2013). Studies show that the predictors for learning to read (i.e. phoneme awareness, letter knowledge and rapid naming) are similar in consistent and inconsistent languages but that English-speaking children spend more time automatising their reading process than peers from more transparent languages (Caravolas, Lervåg, Defior, Seidlová Málková, & Hulme, 2013). Notably, in basic reading instruction, most first-grade teachers in Norwegian aim to use words with a consistent spelling pattern. However, words with inconsistent spelling occur in textbooks, and the children are therefore exposed to this from the very beginning of their reading development.

### **Speech and Articulation Problems as a Cause of Reading Difficulties**

Phonological awareness training typically focuses on listening tasks, for instance, by ‘sounding out’ the phonemes. This presumably presupposes relatively rapid auditory processing, and it has been suggested that a focus on articulation may function as a supplementary and compensatory tool in phonological awareness training for children with phonological deficits. This is because articulation, either aloud or silent, is assumed to give sensory feedback and offer a concrete way to distinguish between sounds (Boyer & Ehri, 2011; Rosenthal & Ehri, 2011; Wise, Ring, & Olson, 1999).

Arguments for this are most clearly formulated in the motor theory of speech perception (Liberman & Mattingly, 1985; Liberman & Whalen, 2000). This theory creates the



backdrop of several intervention studies (see below). According to this theory, the fundamental elements in the perception of human speech are a combination of articulatory gestures (mouth movements) and speech sounds, which together constitute the identity of each phonetic element. Within this framework, articulatory consciousness training is therefore a natural part of basic reading instruction (Lieberman, Shankweiler & Lieberman 1989; Lindamood & Lindamood, 1998).

As for the relationship between orthography and articulation, the ‘phonological recoding hypothesis’ (Share, 1999) suggests that orthographic learning is the product of a natural, ‘self-teaching’ process. When exposed to novel words during text reading, children process the words from prior knowledge of the grapheme/phoneme correspondence and construct new orthographic connections. The aloud or silent articulation of the phonemes are thought to play a role in this recoding process, and studies have shown that suppressing the recoding process, for instance, in experiments where participants are to display concurrent articulation during silent reading, disrupts orthographic learning (de Jong, Bitter, van Setten, & Marinus, 2009; Share, 1999). Orally pronouncing novel words aloud during text reading has also been shown to enhance the student’s memory in regard to both word meaning and the spelling structure of the words (Rosenthal & Ehri, 2011; Ehri, 2014). Although studies have supported the hypothesis of phonological recoding in orthographies with varying levels of transparency (such as Hebrew, English and Dutch), this may be limited to phonetic orthographies (Sahu 2000).

Another relevant perspective for the relationship between articulation and reading are the assumed connections between speech perception, speech production and the underlying phonological representations (Elbro, Petersen, & Borstrom, 1998; McBride-Chang, 1996; Sénéchal, Ouellette, & Young, 2004). The quality and distinctness of the phonological representations are vital in the development of phonological awareness (Elbro et al., 1998; Elbro & Jensen, 2005; Swan & Goswami, 1997) and have also been found to differ between different subgroups of language impaired children (Ramus, Marshall, Rosen & van der Lely, 2013). How the phonological representations are constituted and how different features contribute to the storing of these representations in memory is not clear. However, observational studies have found high correlations between speech perception and phoneme awareness (McBride-Chang, 1996; McBride-Chang, Wagner, Chang & Pressley, 1997), between articulation and speech perception (Sato, Troille, Ménard, Cathiard, & Gracco, 2013) and also between speech perception, articulation accuracy and phoneme awareness (Sénéchal et al., 2004).

Evidence for relations between reading and articulation has been investigated more directly in samples from different populations. Several longitudinal studies in samples of children with speech-sound disorders have shown that especially atypical speech-sound disorders in preschool children predict poor phonological awareness and reading delay in the early school years (Hayiou-Thomas, Carroll, Leavett, Hulme, & Snowling, 2013; Preston, Hull, & Edwards, 2013; Rvacheww, Chiang, & Evans, 2007). Correlations between poor articulation accuracy in preschool children and later reading problems are also reported in

other longitudinal studies of typically developed samples (Sénéchal et al., 2004; Thomas & Sénéchal, 2004).

However, that there is a correlation between articulation and reading does not imply a causal relationship. There are also several studies that do not support the notion that this reflects a causal relationship. Hulme, Nash, Gooch, Lervåg and Snowling, (2015) investigated the foundation of reading comprehension at eight years from a broad spectrum of language and speech measurements conducted when the children were three and five years old ( $N = 145$ ). In this study, only language measures predicted later reading comprehension, and articulation accuracy measured at three years was not significantly related to later reading development. Rvachew and Grawburg (2006) compared two structural equation models of the correlations between speech perception, articulation, receptive vocabulary, phonological awareness and emergent literacy skills in a sample of 90 children (4–5 years old) with speech sound disorders. They found that speech perception and articulation were significantly correlated, but the analysis did not reveal any direct influence from articulation to phoneme awareness.

### **Intervention Studies on Articulation Consciousness Training**

Although the articulation of speech sounds is indirectly embedded in most basic reading instruction programmes through reading-aloud exercises, studies that focus on instructional programmes with articulatory awareness as part of the basic reading instruction are rare. Most frequently cited in the literature is a programme called Auditory Discrimination in Depth (ADD) and a later version of the same programme called Lindamood Phonemic Sequencing (LiPS) (Lindamood & Lindamood, 1998; What Works Clearinghouse [WWC], 2008). This programme is inspired by the motor theory of speech perception and applies a multisensory approach to reading (Wise et al., 1999). The programme is constructed with learning activities that engage students in discovering the lip, tongue and mouth actions needed to produce specific sounds (WWC, 2008). Instructional means include mirrors and pictures of mouth movements in different positions. Adaptions of the original programme, some also including computer-assisted learning, have been tested in both the pre-reading phase and in reading-delayed samples.

As for studies of effect, Wise et al. (1999) compared the efficiency of three phonic-based computer assisted programmes in 122 second to fifth-grade children. Participants were all among the 10% lowest readers in their respective classes and were recruited from five schools in the same area. Children from three other schools in the area ( $n = 31$ ) received regular reading instruction and served as controls. One experimental group received a programme with a focus on sound manipulation, another received a programme with a focus on articulatory awareness (ADD) and a third group received a combination of these two programmes. All participants received 40 hours of training in small groups over six months. The results showed that all intervention groups outperformed the controls on all reading and spelling measurements (significant effect size difference  $d = .73$ – $1.73$ ). The three groups had almost the same gains from interventions across grades and initial levels of phonological awareness and reading.

Castiglioni-Spalten and Ehri (2003) compared the efficacy of two intervention programmes, one based on the ADD (the mouth group) and one based mainly on listening tasks (the ear group) on 45 preschool children (mean age 5.9). Both programmes aimed to improve phonological awareness and reading readiness in children at the ‘partial alphabetic’ level of reading development according to Ehri’s phase theory (1995). Participants were divided into equal triplets, and each triplet was randomly assigned to the ear group, the mouth group or the control group. Both experimental groups were individually taught by the same teacher and received same proportion of training. Controls remained in ordinary classrooms. The results indicated that both training types were efficient in promoting phonemic segmentation, enhancing the children’s ability to spell and to segment words and pseudo-words. However, the articulatory group had reliably better effects on a task measuring the children’s ability to use partial phonetic cues in reading (significant effect size difference  $d = .87$ ).

Boyer and Ehri (2011) replicated some of the features from the 2003 study on a slightly larger and younger sample ( $n = 60$ , mean age 4.9). The main research topic was to compare the efficiency of phonological awareness training with and without a focus on articulation consciousness in respect to moving the children to the next level of development according to Ehri’s theory, the partial alphabetic phase. The content of the two programmes was mainly as described for the 2003 study. The first group was taught to connect phonemes and pictures of articulatory gestures and then segment words first with pictures of articulatory gestures and then with letters. The second group was taught with letters and phonemes only. These groups were compared with a control group who received ‘business as usual’ instruction. The results showed that the two experimental groups both significantly outperformed the controls on all measures. Also significant was a better effect of training in favour of the articulatory awareness group for measures on phoneme segmentation, spelling, word reading and non-word repetition on the immediate post-test (significant effect size differences  $d = .79-1.12$ ). However, when controlling for group differences concerning the total time spent on phonological awareness training in the two conditions, only the word reading measure remained significant at the seven day delayed post-test.

Torgersen, Wagner, Rashotte, and Herron (2003) compared two intervention programmes, both partly including computer-assisted learning, on a sample of first graders with a reading delay. The first programme was an adapted version of the ADD programme and the second programme, Read, Write and Type (RWT), provided explicit instruction in phonetic spelling and writing through engaging the children in learning activities. The participants were 150 first-grade students with low achievement in five elementary schools. At two schools, students were randomly assigned to either the ADD or the RWT programme. At three additional schools, students were randomly assigned to either the ADD group, the RWT group or a regular instruction control group. Instruction was delivered for seven months in groups of three children. During this time, the children received four 50-minute sessions per week. The results document similar effects from the training in the two intervention groups, with a small but not significant advantage in favour of the ADD

group on two measures on phonemic decoding. Both intervention groups significantly outperformed the control group.

In another study, Torgersen et al. (2010) replicated these findings in a design where first graders from three cohorts were recruited after a screening procedure each autumn and then randomly assigned to either ADD, RWT or a control condition. The results documented equal efficiency of the two intervention programmes, both significantly outperforming the controls on measures in phonemic awareness, phonemic decoding, reading accuracy, rapid naming and reading comprehension (significant effect size difference  $d = .37-.77$ ). A follow-up test at the end of second grade showed similar patterns although some of the results at this time were non-significant.

Trainin, Wilson, Murphy-Yagil and Rankin-Erickson (2014) used another articulatory form of training material and examined the effect of two intervention programmes, both aimed to improve early reading skills. In total, 53 third-grade children with reading scores at least one year delayed from five schools in a low socio-economic status area were stratified by school and randomly placed in one of the two intervention conditions at their home schools. One programme, Word Work, aimed to teach the children how to decode and spell through a 'metaphonic' approach using articulatory awareness as a means to provide a conceptual understanding of the basic principles of alphabetic decoding. The other programme, Phonological Awareness Training for Reading, included traditional phonological awareness activities with a focus on segmenting, rhyming and blending and functioned as a control condition. Both interventions also included a component of reading fluency training. This was added from week three of the intervention and provided repeated practice in applying the skills learned in the groups. All training was delivered by graduate students from a local university and included three 45-minute sessions per week for 11 weeks. The results showed significant differences between the two programmes in favour of the Word Work condition in spelling and decoding (effect size difference  $d = .94$ ), oral reading fluency (effect size difference  $d = .79$ ) and metacognitive learning (effect size difference  $d = 1.1$ ). For the phonological awareness measure, there was no significant difference between the two conditions.

A recent study from Sweden investigated the effect of a phonological training programme using an articulation card system called Fonomix (Fälth, Gustafson, & Svensson, 2017). The Fonomix material contains a number of different letter-sound combinations with corresponding pictures representing the specific articulation movement for each. The intervention sample was 69 preschool children that were divided, but not randomised, into a treatment and control condition. During the intervention year, the children assigned to intervention were trained in Fonomix, and the children in the control condition received the same amount of training in a whole language-based programme for preschoolers. For the analysis, the sample was divided into four subgroups based on pre-test scores. The results show a significant effect in favour of the intervention condition on all outcome measures (effect size difference  $d = 1.06-4.27$ ). The results also documented that the 'at-risk' children showed the largest improvement from pre- to post-test on all measures compared to the other children.

Finally, a study evaluated the long-term effects of an intervention on third-grade children, also in Sweden (Wolff, 2016). The intervention programme entailed a multicomponent approach including phonemic decoding, reading comprehension strategies and fluency training. About 60% of the training sessions was spent on phonemic decoding and phonemic awareness tasks using both letter cards and cards with photographs of mouths pronouncing different speech sounds. This study found a sustained effect from the training after five years on word decoding ( $d = .37, p = .01$ ) but no significant effects on the other outcomes. This can give at least some support to the use of articulation cards in programmes that aim to increase phonemic decoding skills.

### **The Current Study**

As outlined above, most studies have compared the effects of articulatory consciousness training with other, more traditional approaches to phonological awareness training. However, in most studies the articulation consciousness training is embedded in programmes also emphasising listening tasks, letter-sound mapping and reading and writing exercises. Some of the programmes also include computer-assisted learning and metacognitive learning. Thus, the studies vary both in focus of instruction, sample composition, duration of training and quality of research designs. This makes the interpretation of the results somewhat difficult in respect to concluding on the unique contributions from the articulation consciousness training described. Few previous studies have used randomised samples, and in a review on the effectiveness of the ADD /LiPS programmes (WWC, 2008), only one out of 25 studies (the Torgersen et al. 2003 study) met the WWC evidence standards. In a more recently published report, WWC evaluated the effects of the LiPS programme on beginning reading (WWC, 2015). Here, two out of 48 studies were included due to the eligibility criteria. The WWC report concluded on a possible positive effect of the LiPS programme on reading comprehension and mixed effect on the alphabetic domain. Both conclusions are based on small amounts of evidence (WWC, 2015). Thus, both the mixed results from previous studies and the fact that the majority of the previous studies have used non-randomised designs underscore that it is important to do a randomised study.

In the present study, we aim to examine the effect of articulatory awareness training on first-grade children with a reading delay and randomise the children to produce a better design to draw conclusions about the effects. The current study also differs from the studies presented above in regard to the instructional material. Most studies described have used pictures of mouths in different positions. The material in the present study, Pictographic Articulatory System (PAS) (Kausrud, 2003), uses pictographic symbols of the speech sounds including a broader spectrum of phonological features, such as place, manner of articulation and acoustic cues. This material has been used in clinical practice on children with language disorders in Norway. The material has also been used by first-grade teachers in basic reading instruction both in full classes and for small-group delivery with positive results according to teachers' and parents' reports. However, there are no studies on the effect of the material, only a case study with one language-impaired child (Ottem &Kausrud, 2001). According to

this study the PAS symbols may have played a mediating and compensatory role by promoting the child's ability to read traditional orthography.

In line with several of the previous studies on novice readers, the intervention programme in the current study integrates articulatory consciousness training in the training of phonological awareness. Compared with most other studies in this area, ours has a larger sample size and also randomisation. Moreover, in contrast to most of the studies cited above, the present study also measures the extent to which the children actually learn to use the articulatory symbols and examines how this is correlated to their progression in phonological awareness, letter knowledge and alphabetic reading.

The research questions in the current study are as follows:

- To what extent did the intervention programme succeed in teaching the children assigned to the intervention condition to use articulatory symbols?
- Will the children that receive an intervention programme with focus on articulatory consciousness improve phonological awareness skills compared with the children assigned to the control condition?
- Will the children that receive an intervention programme with a focus on articulatory consciousness improve alphabetic knowledge compared with the children assigned to the control condition?
- Will the children that receive an intervention programme with a focus on articulatory consciousness improve word reading compared with the children assigned to the control condition?

## **Method**

### **Design**

The design of the study is a randomised controlled trial with an intervention and a control group at each school, 28 clusters in total, with four to six children in each. Pre-tests were administered the week before intervention onset, and the post-tests took place after 20 hours of training. The measurements were administered by the first author and experienced research assistants. All testing sessions were audio recorded for reasons of fidelity.

### **Sample**

A total of 129 first graders from a municipality outside Oslo were recruited based on their scores on the mandatory national reading test (see Figure 1 for an overview of the flow of participants in the study). The sample consists of 89 boys and 40 girls. Eighty-one (63%) have Norwegian as their first language, and 47 (36%) have another first language. Prior to the statistical analysis, we conducted an independent sampled T-test to consider whether it would be relevant to separate the sample from first language background in the analysis. The results from this analysis showed that there were no significant differences between children with Norwegian as their second language and children with Norwegian as their first language, neither on pre- nor post-tests. In further analysis, the sample was therefore considered as a whole.

The national test contains six subtests measuring phonological awareness, word reading, word spelling, letter knowledge and reading comprehension. The reliability for this screening test is not available at the subtest level, but the reliability for the test overall is coefficient alpha  $> .80$  (UDIR homepage: <https://www.udir.no/eksamen-og-prover/prover/rammeverk-for-kartleggingsprover-pa-1.-4.-trinn/krav/#reliabilitetskrav>).

Since articulatory training in previous studies has been most promising for children with phonological problems (Trainin et al., 2014; Fälth et al., 2017), we set the inclusion criteria from the children's scores on the four subtests related to phonological skills and identified children who had scored below the critical limit on at least two of four subtests. The percentile for the critical limit is not available to the public, but it is usually around the 20% weakest scoring part of the sample. Schools with more than six children in this category were by the school administration in the municipality committed to participate in the study and conduct the intervention programme as part of the instruction (see Figure 1 for an overview of the flow of participants in the study).

For each school, the children were distributed in two conditions based on two procedures, as follows: in schools with more than 10 children meeting the inclusion criteria, we first selected 10 children by using a randomiser programme (randomiser.org). In the next step, these children were randomly allocated to an intervention group and a control group. For schools with fewer than 10 children, all children were randomised. The reason for only selecting 10 participants from each school was that a group size of five was considered the maximum possible for the teacher to follow up on through the programme. We also aimed to have an intervention and a control group of nearly the same size at each school. The final sample consisted of 14 intervention clusters and 14 control clusters, each with between three and five participants. All parents gave written consent for their children to participate in the study.

### **Intervention Materials and Procedure**

The content of the programme can be described as an attempt to repeat the central principles in the basic reading instruction. However, instead of using alphabetic letters, the children were introduced to pictographic elements and taught how each speech sound was symbolised by these elements on articulation cards. The intervention programme aimed to increase the students' knowledge about the acoustic and articulatory features of each speech sound and thereby provide an alternative to auditory discrimination between phonemes. The Pictographic Articulatory System (PAS) symbols, which constitute the material used in this intervention, are based on all singular vowels and consonants in the Norwegian alphabet (Kausrud, 2003; Ottem, 2001). We aimed to avoid words with inconsistent spelling in the training sessions. This is because the PAS symbols represent single phonemes and are most convenient for the reading and spelling of words with a transparent spelling structure. Both the intervention material and the phonological awareness and reading tests conducted at pre- and post-tests therefore include transparently spelled words only. Figure 2 shows how the word ROSE, spelled in Norwegian, is depicted in PAS. For an overview of the PAS "Alphabet", see the supplementary material

The programme was constructed for the purpose of this study and consists of six learning activities, five activities for group work and one activity for individual work. The group activities were organised as a stepwise introduction to how to use the PAS symbols in representing speech sounds. First, the children were engaged in shared activities in exploring their own articulatory system, also using mirrors. Secondly, they learned the logic of the PAS system concerning how the articulatory and acoustic features of the sounds are represented in the symbols. Further, matching, spelling and reading tasks with the PAS symbols were also introduced. Additional materials for the shared activities included cards with words spelled in the PAS font and picture cards. The programme emphasises a collaborative atmosphere where the teacher and the children together investigate the PAS system and how to use the system in decoding and spelling. The teacher acts as a model for how to identify the articulatory and acoustic features of each speech sound, and the children also use mirrors to make the articulation more concrete.

The individual activities were different forms with fill-in transposition tasks. In this activity, the children read words written using the PAS system and were to transpose these words into alphabetic script. The programme has an increasing difficulty level that reflects the development from fluent and accurate knowledge of the PAS symbols to mastery of the symbols in reading and spelling (for details concerning the content of the programme, see the supplementary material).

The intervention was carried out at the children's local schools, and the teachers received three hours of group instruction in how to conduct the programme prior to the start of the intervention. The programme includes 20 training sessions of 40 minutes each. All teachers that delivered the programme were educated and experienced, some with additional training as special needs educators.

### **Fidelity**

All participating teachers had to fill in a record form for each lesson and briefly describe how the intervention programme had worked for that particular session. All lessons were also audio recorded to document that the teacher had followed the predefined structure of the programme. The correlation between the teachers' reports and the audio-recorded lessons was nearly 100%. However, some schools reported that they had failed to deliver the programme as planned and had spent less time per week on the intervention, mainly due to teacher absence. The intervention groups in these schools received the same number of hours – 20 in total – but spent more weeks completing the whole programme. For some groups, the delivery of the programme was also shared between two teachers.

For each intervention group, we rated 10% of the recordings to examine how the teachers followed the programme and how the interaction in the groups worked out. The results from this rating indicate that all groups delivered the details of the programme as intended. However, the recordings revealed some problems with the group interaction in two groups, and this may have reduced the effects from the training in these groups.



## **Control Condition**

The control condition in this study apply a “business as usual” approach. This means that the children got what the school usually provided children with a reading delay. All schools had to answer a questionnaire concerning the instruction of the children assigned to the control condition. The response rate was 85.7%. Six schools (50% of the answers) reported that the controls got some sort of special instruction in reading, spelling and phonological awareness during the weeks of the intervention. The content, organisation and dosage of this instruction varied between schools. One school reported daily individual lessons for all control children, and three other schools also offered individual instruction in addition to small-group instruction for the controls. Many schools had two teachers available for the first graders, and several schools also had school assistants present in the classrooms. This made it possible to use group instruction and tailor the reading instruction to individual needs.

Notably, in some schools the controls and the children in the intervention condition received instruction from the same teacher but at different times of the day; in other schools the delivery to the controls was given parallel to the intervention by another teacher. Importantly, the controls did not receive any of the instruction concerning the intervention materials. Because the assignment of the children to the two conditions was blinded to the professionals doing the post-intervention assessment, all children were shown two PAS symbols at the end of the measurement procedure and asked if they knew the symbols. The correspondence between the result from this check and the assignment key was 100%. This means that none of the children in the control condition had been exposed to the intervention material. The transposition test was added as a supplementary measure for those knowing the symbols.

The questionnaire for the teachers did not contain questions regarding instruction for those over the critical limit that remained in the classrooms. However, experience from a pilot study in several classrooms in the same municipality prior to the current study indicated that the general curricula across the classrooms included a broad spectrum of phonological awareness training and phonics reading exercises. It is also worth noting that in this municipality all first graders are offered a personal iPad, which is frequently used in the reading instruction and for practicing.

## **Measures**

### **Screening Test**

The national reading test is constructed as a group screening test and contains six subtests. For this purpose, only the four following subtests related to phonological skills are relevant:

- To identify sounds in words: phoneme position tasks (14 items)
- To segment sounds into words: picture-based tasks where the child must select a picture that corresponds to sounds read aloud (12 items)

- To spell words: the child must write and spell a word from a sentence that is read aloud (14 items)
- To read words: picture-based tasks where the children are asked to read a word silently and select the correct picture out of four (14 items)

### **Outcome Measures**

Pre- and post-test measures were mainly constructed for the purpose of this study and were tests of letter knowledge, phonological awareness, reading accuracy, word-reading tasks and one test concerning the intervention material measuring the children's understanding of the pictographic symbols used. In the construction of the phonological awareness tests, we split the tasks into three sets. The reason for this was to be able to compare the potential effect of measures with measures that had different degrees of proximity to the material used in the intervention. The first set (trained words) consisted of words used in exercises in the intervention programme. These words are all nouns that are easy to depict, and all had a regular, transparent spelling structure – that is, no diphthongs, double consonants or compounded phonemes. The words were selected from a database of the 500 most frequent words in the Norwegian language (Norwegian list of word frequencies) and from the database 'Ordforrådet' (<http://tekstlab.uio.no/ordforradet>). This database also provides information concerning usage frequency and the ability for a word to be depicted.

The second set (untrained words) consisted of words that were matched to the trained words in phonological complexity, ability to be depicted and usage frequency. The third set consisted of pseudo-words, which were also matched to the other sets in phonological complexity. For the one-minute word-reading tasks, we used a similar procedure; here we constructed two lists of regularly spelled words. The first list (trained words) consisted of words used in the intervention exercises, and the second list (untrained words) was constructed to match the trained words but was not used in the intervention.

The third set of tasks (unrelated to the intervention tasks) were two pseudo-word reading subtests from the Test of Word Reading Efficiency (TOWRE) battery (Rashotte, Torgersen, & Wagner, 1999). Reliability for all measures is reported in Table 1. For measures without time limits, we used coefficient alpha. For the other tests, the reliability measure is the correlation between equal tests (two word-reading tests and two pseudo-word-reading tests).

#### ***Letter knowledge***

This test contains 48 items. All the letters in the Norwegian alphabet (24 in total) written both in uppercase and lowercase letters were presented to the children, who were asked to identify both the name of the letter and its corresponding sound. Reliability for both subtests (letter sound knowledge and letter name knowledge) is reported in Table 1.

#### ***Phonological awareness***

Three different sets (trained words, untrained words and pseudo-words, respectively) were constructed, each set containing the following three subtests:

- Phoneme identification: e.g. ‘What is the first sound in ball?’
- Phoneme reduction: e.g. ‘What remains of the word ‘ball’ if you delete the ‘b’ sound?’
- Spoonerism: e.g. ‘What is ball with a ‘t’ first instead of a ‘b’?’

All subtests contain 12 items (36 per set) with increased levels of difficulty. The stop criteria were four consecutive errors in the first subtest, ‘phoneme identification’. The scoring was either right or wrong.

## **Decoding**

The decoding measures contained three subtests (reading accuracy, one-minute reading of regularly spelled words and one-minute reading of pseudo-words).

### *Accuracy*

Twenty-two pseudo-words with increasing difficulty were constructed for this task. The first words in the list were simple consonant-vowel (CV) words. The items had an increasing level of difficulty, ending in complex words with 14 letters and three consonant clusters. The children were asked to read the words as correctly as possible without any time limits. The stop criteria were four consecutive errors. When graphemes were either added, transposed or deleted, or if the reading was not fluent, this was defined as an error.

### *Regularly spelled words – trained*

Sixty regularly spelled nouns were selected from a database of the 500 most frequent Norwegian words (Norwegian list of word frequencies). The words were then listed based on their level of phonological complexity. The first words were simple CV or VC words with two graphemes, followed by CVC words and then successively VCC, CVCV, CVCC and finally CCVCCV. Scoring was based on the number of correct words read in one minute. Words read in a phonologically acceptable way (i.e. if all graphemes were articulated) were scored as correct.

### *Regularly spelled words – untrained*

The list of matched words was constructed using the same procedure and the same scoring criteria as described for the trained words.

### *Pseudo-words*

The pseudo-word measures consisted of two translated versions of the TOWRE test – phonemic reading, set A and B. A child’s score was calculated based on the number of correct words read in one minute. Words read in a phonologically acceptable way (i.e. all graphemes were articulated) were scored as correct.

### *Transposition of pictographic symbols*

The transposition test evaluated the progress in mastery of the articulatory elements in decoding and consisted of 24 listed pseudo-words with increased complexity, scripted in the font of the PAS system. The words all had a VC, CV, VCC or CVCV structure and consisted

of two to four symbols (letters). A child's score was based on the number of pictographic words transposed into correct alphabetic script in four minutes. Supplementary scoring also included the number of pictographic graphemes transposed correctly into alphabetic graphemes regardless of whether the word was spelled correctly.

### Analysis

To evaluate the effects of the intervention, multi-level regression models were conducted for all the outcome measures. Here, the outcome scores for each measure at post-test were regressed on the scores at pre-test and the group dummy variable (experiment or control group). In these analyses, we used the restricted maximum likelihood (REML) estimator in Stata 16. Before deciding which level to include in the regression models for the various outcomes, we first calculated intra-class correlations for all the outcome variables using three-level homoscedastic partially nested mixed-effects models. In these models, between-school variance was estimated as the third level, and the between-intervention group variance was estimated in a partially nested fashion (as suggested by Flight et al., 2016) as the second level.

As can be seen in Table 1, the Intra Class Correlations (ICC) varied from .253 to redundant. In the following regression analyses, we included the third and/or the second level as long as it was not redundant. In all the analyses in which the partially nested second level was included, we also used the Satterthwaite degrees of freedom correction, as suggested by Candlish and colleagues (2018). Further, as the residual variances might differ across the two conditions in the partially nested models, we also estimated partially nested heteroscedastic models where the variances were measured separately for the two conditions. However, as the results did not differ from the homoscedastic partially nested mixed-effects models, we only report the results from the latter ones.

### Results

Table 1 shows the descriptive together with Cohen's  $d$  between pre- and post-tests and the ICCs for the outcome variables. As shown in Table 1, all variables had acceptable normal distributions except for letter sound knowledge, where there were some ceiling effects.

As can be seen from Table 1, the children who received the intervention managed on average to transpose slightly less than half of the words (10.20 out of 24) in the transposition test. Supplementary scoring gave credit for each symbol correctly transposed even if the words were not fully spelled or wrongly spelled. The average scoring on this measure was 29.65 out of 66. Both these values were significantly different from zero; PAS words:  $t = 12.76$  (62),  $p < .001$  and PAS symbols:  $t = 15.38$  (62),  $p < .001$ . However, as there was variation around these means (see Table 1), we used partial correlation analyses to determine whether this variation on the PAS test was associated with the post-tests of the outcome measures after controlling for the same test at pre-test. As can be seen from Table 2, there

were significant partial correlations between the transposition of PAS words and phonological awareness of untrained words and pseudo-word reading (TOWRE phonemic B). None of the other correlations was significant.

### **Effects from the Intervention**

The fixed effects results from the multilevel mixed-effects models are shown in Table 2. As can be seen, no significant main effects were found from the intervention. Secondly, we analysed whether there were interaction effects between the pre-test scores on phonological awareness and the group variable on the post-test outcome. The results showed no significant interaction effects on the post-tests – that is, the gains, or lack thereof, from the intervention were unrelated to the initial phonological awareness skills.

### **Discussion**

This study evaluated the effects of a five-week intervention targeting first graders below the critical limit on the mandatory national test in reading. The focus of the intervention was to examine whether introducing articulatory consciousness training in basic reading instruction can function as a bridge from sounds to alphabetic script and increase the children's phonological awareness and word-reading skills.

To examine whether the children had learnt the system that the articulatory consciousness training is based on, they were assessed with a transposition test. On this test, the children allocated to the intervention condition read pseudo-words written with the PAS symbols and were asked to translate the PAS 'pictographic words' by writing them in ordinary alphabetic script. Thus, in some ways the effects on this measure may be a prerequisite for the effects on the secondary outcomes, such as reading and phoneme awareness. The results show that, on average, the children's mastery of the symbol system was not fully automated. However, it should be noted that the extent to which the children learnt the symbol system varied (Table 2). Some of the participants scored near the ceiling level, while others showed less progress concerning automation of the system. These diverging results might reflect both individual differences among the children and factors concerning the implementation of the programme in the different schools.

When we examined whether the effects from the interventions was related to the extent to which the children had learnt to master the PAS symbols, the results were not clear. As shown in Table 2, we found significant relations between the transposition of words and two of the post-test measures. On these two measures, there was a small to moderate relationship between the PAS test and the phonological awareness tasks for untrained words and for the pseudo-word reading measure (TOWRE B). Also, these correlations were not present for the other post-test measures of the phonological awareness of trained words and pseudo-words nor for the other pseudo-word reading measures (TOWRE – A and Accuracy). Since these measures are highly similar to those for which there was a significant relationship, it can clearly be questioned whether this is a true and robust relationship.

Still, in spite of some effects on learning the symbol system in the intervention, there were no effects from the intervention overall: the control condition and intervention condition had similar gains on the secondary outcome measures. Also, there were no significant relationships showing that the poorer the phonological awareness skills, the larger the effects from the intervention.

Several factors may explain the lack of effect from the intervention on the outcome measures. The most apparent explanation is, of course, that the intervention does not work since we had some effects on the children's learning on the PAS symbols but no effects on the outcome measures.

Still, there also might be some alternative explanations for why this intervention was not effective in this study and that, perhaps, can contribute to nuance the conclusion that this intervention is not at all effective. One is that this intervention might not be effective for the group with reading difficulties that we have examined here (20% with the weakest phonological skills based on the national test). Many of these children did not have weak phonological problems, and they can score below the critical limit for several reasons. That said, what speaks against the notion that this lack of effect is sample-specific is that there was no relationship between the initial phoneme awareness skills and gains, or lack thereof, from the intervention. But it should be noted that the power of these kinds of analyses is rather weak since it only concerns those in the intervention condition.

Another possible alternative explanation for why the intervention did not work is the rather limited time spent on the intervention due to the complexity of the symbol system. Five weeks is short a time, and the intervention took place in the last weeks of the school year. In this period there are several national holidays, and this might have caused discontinuity in the delivery of the programme. As mentioned, there were also some implementation issues as some of the groups also met problems in delivering the programme in time. Problems with the social interaction in some of the groups may also have reduced the benefits from training.

Concerning the time limitation, it is also worth mentioning that the children in the intervention condition spent time on learning the new symbols of PAS, while the controls got 'more of the same' by continuing to take part in training inspired by the more traditional approach to phonological awareness training, either in the general classrooms or in training groups. Previous studies have noted that it takes more time to teach the children how to use articulatory cues as a supplementary tool in segmenting tasks compared to a 'letter only approach' (Boyer & Ehri, 2011). The results from the transposition test also show that not all the symbols were automatized at post-test, indicating that the training period may have been too short. From this perspective, it is possible that the training effect from the intervention could have increased with more time spent on the intervention. Other studies on the effects from articulatory training in similar populations, for instance that of Fälth et al. (2017), provided one year of training.

Finally, since all the children both in the intervention and control conditions improved from pre- to post-test, an alternative explanation for the lack of effect in favour of the intervention programme could be the quality of the instruction for the controls. Both the

results from the teachers' questionnaire and the experience from the pilot in the general classrooms prior to the intervention indicate that the controls received well-tailored phonological awareness training and reading instruction during the intervention period. Still, as mentioned, only 50% of the schools reported that the controls got some sort of special instruction in reading, spelling and phonological awareness during the weeks of the intervention.

Given that the control group also received high-quality, tailored instruction, the results from the present study are in line with previous intervention studies. Thus, as mentioned, previous studies have shown a similar size of effects from training in phonological awareness and basic decoding with and without a focus on articulatory consciousness (Wise et al., 1999; Torgersen et al., 2003, 2010). Since the post-test measures used listening tasks and word-reading tasks with alphabetic script, the content of the measures may also have favoured the children in the control condition who possibly spent more time on listening tasks and alphabetic decoding exercises during the intervention period than the children in the intervention condition.

## **Limitations**

The study has several limitations. First, the study was conducted in collaboration with the school administration in the municipality. This means that the schools and teachers were committed to participate in the study. A post-study questionnaire for the teachers revealed mixed attitudes in regard to motivation to participate in the study. Although a large majority of the teachers gave positive responses on the evaluation of the intervention programme, negative attitudes prior to the intervention may have inflected the outcome in some of the groups. In the meetings with the school administration prior to the study onset, we could have specified that we wanted to avoid the recruitment of teachers with negative attitudes towards conducting external intervention programmes. A more motivated staff would possibly have improved the efficacy of the intervention in some of the groups.

Since the learning of the new symbols in the intervention condition in itself appeared to be time consuming, this may have prevented the children from getting enough time to apply the articulatory consciousness in reading and spelling exercises during the intervention period. We could therefore have added additional post-tests looking for delayed effects in the intervention condition. Since the intervention took place in the last weeks of the school year and since the school office for ethical reasons wanted to break the assignment into the two conditions after the intervention period, this was not possible.

Another issue concerns the curricula-based measures applied for the pre- and post-test assessments. The study intended to measure the effects on near and distal measures and evaluate any transfer effects to pseudo-words both on phonological awareness tasks and on reading tasks. The near transfer tasks were directly based on the words used in the intervention material, while the distal measures were matched to the words used in the

intervention material but were different words. Since the intervention material partly used pictures, the words selected both for the phonological awareness tasks and the word-reading tasks were depictable, high frequency nouns. This may have caused a high degree of overlap with words often used in the basic reading instruction in the classrooms and might partially explain the lack of effect on both the taught and distal words in the experiment condition. We could perhaps have controlled for this by including word knowledge in the pre-test measures. However, this would have extended the time spent on the measurement procedure, which was undesirable for both practical and ethical reasons.

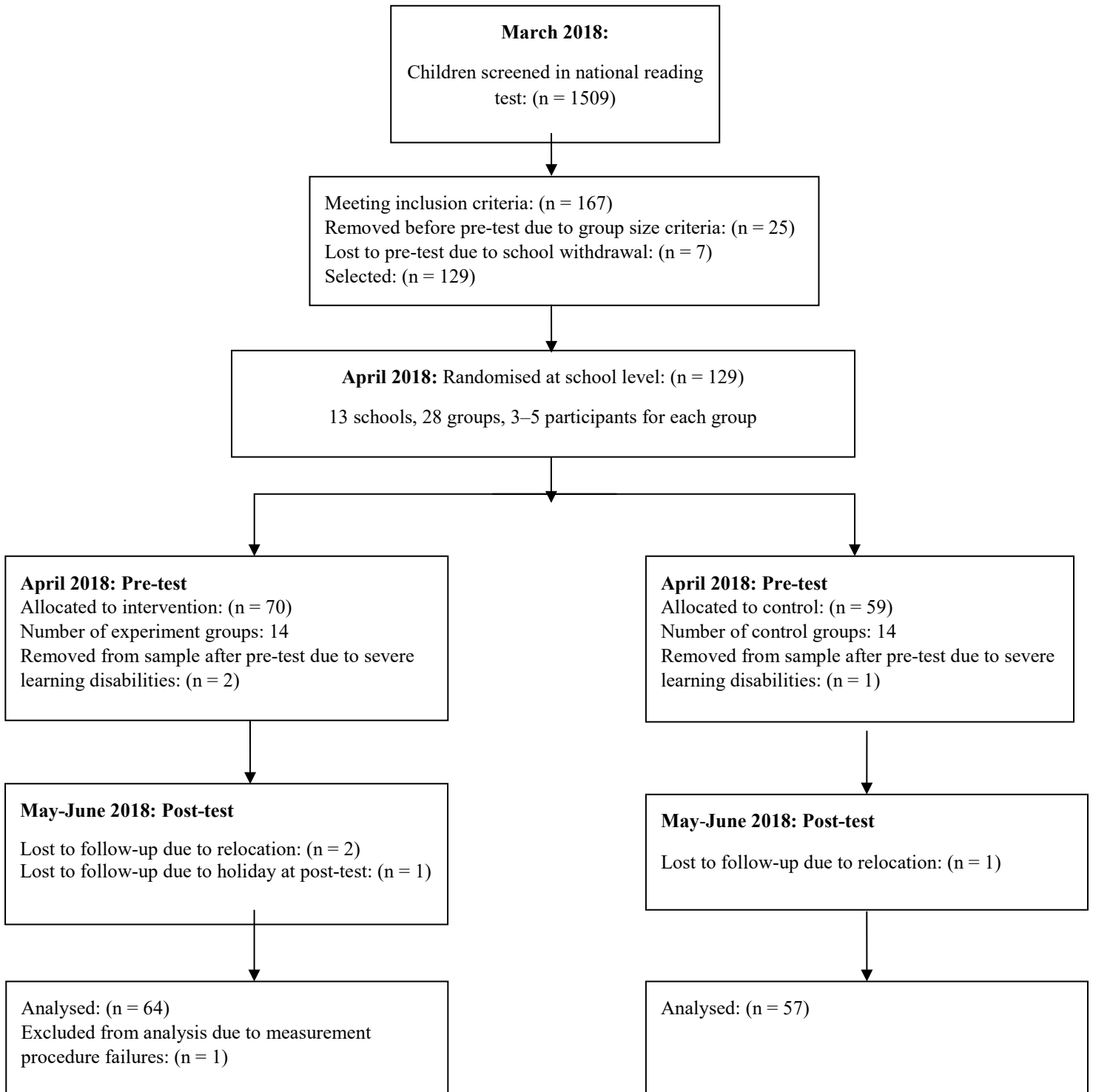
The sample was based on the weakest first-grade readers in one municipality. Inclusion criteria were established to select children with a presumed phonological deficit. However, the results from the individual pre-tests show that the sample is quite heterogeneous and includes children with minor problems concerning phonological awareness skills. The national test is administered in full classrooms, and the instructional procedure is standardised, not allowing additional assistance. This may have included some false positives in the sample, considering that these children are young and not familiar with formal testing procedures. We could have changed the inclusion criteria after the individual pre-tests and randomised the sample from more restrictive criteria concerning phonological awareness, but this would have reduced the sample size. A smaller sample would have reduced the power of the statistical analyses. For ethical reasons, the school administration in the municipality also desired less restrictive criteria for inclusion in the intervention programme since many resources the weeks after the mandatory test were spent on the intervention programme.

### **Conclusions and Future Directions**

The results from the current study indicated that training with a focus on articulatory consciousness was as effective as the ‘business as usual’ delivery to first graders in this municipality. The most apparent conclusion from this is that the intervention was not effective. However, it should be noted that the control condition in this study might have had an intervention dosage similar to the intervention group but with different content. An alternative interpretation of the results is therefore that phonological awareness training both with and without a focus on articulation consciousness are equally effective to facilitate basic decoding skills in novice readers. This conclusion is in line with the main conclusions from most previous research on similar populations (Boyer & Ehri, 2011; Torgersen et al., 2003, 2010). To investigate the effect of articulatory consciousness training from the perspective of compensation, further studies should consider a more clinical sample with severe problems in regard to phonological skills.



Figure 1. Flow of Participants



## Figure 2

### *Pictographic Symbol System (PAS)*



*Figure 2.* PAS 'spelling' of the word ROSE [² ru:sə ] (spelled in Norwegian): Blue script for consonants and red script for vowels. Consonant cards contain indicators for voice, placement of tongue and acoustic cues. Vowel cards symbolise the shape and the opening of the mouth when pronouncing a vowel.

Table 1.

*Means, standard deviations, range, reliability and effect size for the sample (N = 121).*

Measure	Intervention Condition (SD)	Min-max	Control Condition (SD)	Min-max	$\alpha$	Cohens <i>d</i>	ICC	
							School level	Partial intervention level
Letter sound								
Pre-test	19.89(5.20)	3-24	20.75(4.84)	3-24	.92 <sup>a</sup>			
Post-test	22.57(3.05)	8-24	22.22(2.87)	10-24	.87 <sup>a</sup>	.22	.252	<.001
Letter name								
Pre-test	15.14(6.37)	0-24	16.47(6.85)	2-24	.91 <sup>a</sup>			
Post-test	18.01(5.70)	0-24	19.98(4.57)	7-24	.91 <sup>a</sup>	-.24	.052	.092
Phonological awareness – trained words								
Pre-test	9.72(6.79)	0-24	11.22(7.23)	0-28	.93 <sup>a</sup>			
Post-test	15.28(6.49)	2-28	15.31(7.04)	0-34	.90 <sup>a</sup>	.15	.190	<.001
Phonological awareness – untrained words								
Pre-test	8.98(6.87)	0-24	10.65(6.70)	0-23	.93 <sup>a</sup>			
Post-test	12.51(6.44)	0-27	13.71(6.70)	2-28	.90 <sup>a</sup>	-.06	.075	<.001
Phonological awareness – pseudo-words								
Pre-test	7.63(6.38)	0-22	9.31(6.83)	0-28	.923 <sup>a</sup>			
Post-test	11.34(5.72)	2-24	12.71(6.2)	0-29	.894 <sup>a</sup>	-.06	.201	<.001
Word reading – trained words								
Pre-test	9.71(5.99)	0-24	10.44(5.74)	1-27	.83 <sup>b</sup>			
Post-test	13.59(6.91)	2-29	15.31(7.62)	2-35	.91 <sup>b</sup>	-.07	.033	<.001
Word reading – untrained words								
Pre-test	7.93(6.04)	0-23	9.85(6.38)	0-25	.83 <sup>b</sup>			
Post-test	10.71(6.66)	0-27	13.71(7.65)	0-31	.91 <sup>b</sup>	-.13	.041	<.001
TOWRE phonemic A								
Pre-test	7.80(5.78)	0-20	9.85(6.38)	0-25	.79 <sup>b</sup>			
Post-test	11.34(5.67)	2-24	13.10(6.04)	0-28	.90 <sup>b</sup>	.003	<.001	.050
TOWRE phonemic B								
Pre-test	5.63(4.97)	0-17	7.37(5.67)	0-24	.79 <sup>b</sup>			
Post-test	9.84(5.36)	0-26	11.52(5.74)	0-25	.90 <sup>b</sup>	-.006	.052	.006
Accuracy								
Pre-test	5.38(4.22)	0-17	5.47(4.07)	0-16	.84 <sup>a</sup>			
Post-test	9.03(5.39)	0-21	8.17(4.20)	0-21	.88 <sup>a</sup>	.19	.130	<.001
Transposition of PAS symbols								
Pre-test	29.65(15.30)	5-66						
Transposition of PAS words								
Pre-test	10.20(6.34)	0-24						

Note: <sup>a</sup> = Coefficient alpha; <sup>b</sup> = Pearson's correlation; Cohens *d* = difference between the intervention condition and the control condition in standard deviation units controlled for pre-

test. PAS symbols unknown to the children before intervention and transposition tests therefore only present at post-test.

Table 2.

*Partial correlations between the transposition test and the outcome measures at post-test (n = 62).*

Measure	Transposition of PAS Words		Transposition of PAS Symbols	
	Correlation	<i>p</i> - value	Correlation	<i>p</i> - value
Letter sound	.245	.057 n.s.	.162	.212 n.s.
Letter name	.018	.889 n.s.	.084	.521 n.s.
Phonological awareness – trained words	.116	.367 n.s.	.098	.450 n.s.
Phonological awareness – untrained words	.267	.036 *	.153	.234 n.s.
Phonological awareness – pseudo-words	.161	.210 n.s.	.227	.077 n.s.
Word reading – trained words	.179	.167 n.s.	.090	.492 n.s.
Word reading – untrained words	.119	.361 n.s.	.131	.316 n.s.
TOWRE phonemic-A	.082	.529 n.s.	.098	.451 n.s.
TOWRE phonemic-B	.275	.032*	.243	.059 n.s.
Accuracy	-.243	.057 n.s.	-.190	.138 n.s.

Note: \* =  $p > .05$ ; n.s. = non significant values  $p = > .05$

Table 3.

*Effects of the treatment: Unstandardized regression coefficients, standard errors, p values and 95% confidence interval. Fixed effect estimates controlled for pre-test.*

	<i>Reg. coef.</i>	<i>S.E.</i>	<i>p value</i>	<i>Confidence interval (95%)</i>
Letter Sound <sup>a</sup>	.698	.426	.101 n.s.	-.137, 1.533
Letter Name <sup>b</sup>	-1.193	.770	.134 n.s.	-2.778, .392
Phonological Awareness – trained words <sup>a</sup>	1.158	.842	.169 n.s.	-1.576, 3.705
Phonological Awareness – untrained words <sup>a</sup>	-.318	.981	.746 n.s.	-2.240, 1.605
Phonological Awareness – pseudo-words <sup>a</sup>	-.342	.807	.672 n.s.	-1.923, 1.239
Word Reading – trained words <sup>a</sup>	-.343	.824	.677 n.s.	-1.957, 1.271
Word Reading – untrained words <sup>a</sup>	-.935	.874	.291 n.s.	-2.697, .827
TOWRE Phonemic A <sup>c</sup>	-.016	.735	.983 n.s.	-1.460, 1.491
TOWRE Phonemic B <sup>a</sup>	-.370	.769	.631 n.s.	-1.88, 1.238
Accuracy <sup>a</sup>	.915	.672	.173 n.s.	-.402, 2.232

*Note.* <sup>a</sup> Fixed effects results from a two-level mixed-effects model with school as the second level. <sup>b</sup> Fixed effects results from a three-level regression homoscedastic partially nested mixed-effects model with the intervention groups in the treatment arm as the second level and

school as the third level. <sup>c</sup>Fixed effects results from a two-level regression homoscedastic partially nested mixed-effects model with the intervention groups in the treatment arm as the second level.

## References

- Boyer, N., & Ehri, L. C. (2011). Contribution of phonemic segmentation instruction with letters and articulation pictures to word reading and spelling in beginners. *Scientific Studies of Reading, 15*(5), 440–470. <https://doi.org/10.1080/10888438.2010.520778>
- Candlish, J., Teare, M. D., Dimario, M., Flight, L., Mandelfield, L., & Walters, S. J. (2018). Appropriate statistical methods for analysing partially nested randomised controlled trials with continuous outcomes: a simulation study. *BMC Medical Research Methodology, 18*, 105. <https://doi.org/10.1186/s12874-018-0559-x>
- Caravolas, M., Lervåg, A., Defior, S., Seidlová Málková, G., & Hulme, C. (2013). Different patterns, but equivalent predictors, of growth in reading in consistent and inconsistent orthographies. *Psychological Science, 24*(8), 1398–1407. <https://doi.org/10.1177/0956797612473122>
- Castiglioni-Spalten, M. L., & Ehri, L. C. (2003). Phonemic awareness instruction: Contribution of articulatory segmentation to novice beginners' reading and spelling. *Scientific Studies of Reading, 7*(1), 25–52. [https://doi.org/10.1207/S1532799XSSR0701\\_03](https://doi.org/10.1207/S1532799XSSR0701_03)
- de Jong, P. F., Bitter, D. J. L., van Setten, M., & Marinus, E. (2009). Does phonological recoding occur during silent reading, and is it necessary for orthographic learning? *Journal of Experimental Child Psychology, 104*(3), 267–282. <http://www.sciencedirect.com/science/article/pii/S0022096509001337>. <https://doi.org/10.1016/j.jecp.2009.06.002>
- Ehri, L. C. (1995). Phases of development in learning to read words by sight. *Journal of Research in Reading, 18*(2), 116–125.
- Ehri, L. C. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading, 18*(1), 5–21. <https://doi.org/10.1080/10888438.2013.819356>.
- Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36*(3), 250–287. <https://doi.org/10.1598/RRQ.36.3.2>
- Elbro, C., & Jensen, M. N. (2005). Quality of phonological representations, verbal learning, and phoneme awareness in dyslexic and normal readers. *Scandinavian Journal of Psychology, 46*(4), 375–384. <https://doi.org/10.1111/j.1467-9450.2005.00468.x>
- Elbro, C., Petersen, D. K., & Borstrom, I. (1998). Predicting dyslexia from kindergarten: The importance of distinctness of phonological representations of lexical items. *Reading Research Quarterly, 33*(1), 36–60. <https://doi.org/10.1598/RRQ.33.1.3>



- Fälth, L., Gustafson, S., & Svensson, I. (2017). Phonological awareness training with articulation promotes early reading development. *Education, 137*(3), 261–276.
- Flight, L., Allison, A., Dimario, M., Lee, E., Mandefield, L., & Walters, S J. (2016). Recommendations for the analysis of individually randomised controlled trials with clustering in one arm – A case of continuous outcomes. *BMC Medical Research Methodology, 16*, 165. <https://doi.org/10.1186/s12874-016-0249-5>
- Hayiou-Thomas, M. E., Carroll, J. M., Leavett, R., Hulme, C., & Snowling, M. J. (2017). When does speech sound disorder matter for literacy? The role of disordered speech errors, co-occurring language impairment and family risk of dyslexia. *Journal of Child Psychology and Psychiatry, 58*(2), 197–205. <https://doi.org/10.1111/jcpp.12648>
- Hulme, C., Nash, H. M., Gooch, D., Lervåg, A., & Snowling, M. J. (2015). The foundations of literacy development in children at familial risk of dyslexia. *Psychological Science, 26*(12), 1877–1886. <https://doi.org/10.1177/0956797615603702>
- Kausrud, T. (2003). PAS - steget inn i alfabetisk lesning: bruk av piktografisk-artikulatoriske symboler for utvikling av fonologisk bevissthet. Master thesis, Universitetet i Oslo. URN:NBN:no-6248
- Landerl, K., Ramus, F., Moll, K., Lyytinen, H., Leppanen, P. H. T., Lohvansuu, K., . . . Schulte-Korne, G. (2013). Predictors of developmental dyslexia in european orthographies with varying complexity. *Journal of Child Psychology and Psychiatry, 54*(6), 686–694. <https://doi.org/10.1111/jcpp.12029>
- Lieberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*(1), 1–36. [https://doi.org/10.1016/0010-0277\(85\)90021-6](https://doi.org/10.1016/0010-0277(85)90021-6)
- Lieberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences, 4*(5), 187–196. [https://doi.org/10.1016/S1364-6613\(00\)01471-6](https://doi.org/10.1016/S1364-6613(00)01471-6)
- Lieberman, I. Y., Shankweiler, D., & Liberman, A. M.(1989). *The alphabetic principle and learning to read*. National Inst. of Child Health and Human Development (NIH), Bethesda, MD. Reprinted from *Phonology and reading disability: Solving the reading puzzle*. International Academy for Research in Learning Disabilities, Monograph Series. <https://eric.ed.gov/?id=ED427291>
- Lindamood, P., & Lindamood, P. (1998). Lindamood Phoneme Sequencing Program for Reading, Spelling, and Speech (LiPS). <https://lindamoodbell.com/program/lindamood-phoneme-sequencing-program>
- McBride-Chang, C. (1996). Models of speech perception and phonological processing in reading. *Child Development, 67*(4), 1836–1856. <https://doi.org/10.2307/1131735>
- McBride-Chang, C., Wagner, R. K., Chang, L., & Pressley, G. M. (1997). Growth modeling of phonological awareness. *Journal of Educational Psychology, 89*(4), 621–630. <https://doi.org/10.1037/0022-0663.89.4.621>

- Melby-Lervåg, M., Lyster, S.-A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin*, *138*(2), 322–352. <https://doi.org/10.1037/a0026744>
- Ottem, E. & Kausrud, T. (2001). Use of pictographic-articulatory symbols to promote alphabetic reading in a language-impaired boy: Case study. *Augmentative and Alternative Communication*, *17*(1), 52. <https://doi.org/10.1080/aac.17.1.52.60>
- Preston, J. L., Hull, M., & Edwards, M. L. (2013). Preschool speech error patterns predict articulation and phonological awareness outcomes in children with histories of speech sound disorders. *American Journal of Speech-Language Pathology*, *22*(2), 173ff. [https://doi.org/10.1044/1058-0360\(2012/12-0022\)](https://doi.org/10.1044/1058-0360(2012/12-0022))
- Ramus, F., Marshall, C. R., Rosen, S., & van der Lely, H. K. J. (2013). Phonological deficits in specific language impairment and developmental dyslexia: Towards a multidimensional model. *Brain*, *136*(2), 630–645. <https://doi.org/10.1093/brain/aws356>
- Rashotte, C., Torgersen, J. K., & Wagner, R. (1999). Test Of Word Reading Efficiency (TOWRE).
- Rosenthal, J., & Ehri, L. C. (2011). Pronouncing new words aloud during the silent reading of text enhances fifth graders' memory for vocabulary words and their spellings. *Reading and Writing: An Interdisciplinary Journal*, *24*(8), 921–950. <https://doi.org/10.1007/s11145-010-9239-x>
- Rvachew, S., Chiang, P.-Y., & Evans, N. (2007). Characteristics of speech errors produced by children with and without delayed phonological awareness skills. *Language, Speech, & Hearing Services in Schools*, *38*, 60.
- Rvachew, S., & Grawburg, M. (2006). Correlates of phonological awareness in preschoolers with speech sound disorders. *Journal of Speech, Language, and Hearing Research*, *49*(1), 74–87.
- Sahu, S. (2000). Lexical access in Oriya orthography: The Impact of Suppression condition. In S. Sahu (ed.), *Psychology of reading: role of orthographic features* (pp. 41–51). Concept Publishing Company. 110059, ISBN 91- 7022-832-8.
- Sato, M., Troille, E., Ménard, L., Cathiard, M.-A., & Gracco, V. (2013). Silent articulation modulates auditory and audiovisual speech perception. *Experimental Brain Research*, *227*(2), 275–288. <https://doi.org/10.1007/s00221-013-3510-8>
- Sénéchal, M., Ouellette, G., & Young, L. (2004). Testing the concurrent and predictive relations among articulation accuracy, speech perception, and phoneme awareness. *Journal of Experimental Child Psychology*, *89*(3), 242–269. <https://doi.org/10.1016/j.jecp.2004.07.005>

- Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology*, 72(2), 95–129. <https://doi.org/10.1006/jecp.1998.2481>
- Suggate, S. P. (2016). A meta-analysis of the long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions. *Journal of Learning Disabilities*, 49(1), 77–96. <https://doi.org/10.1177/0022219414528540>
- Swan, D., & Goswami, U. (1997). Phonological awareness deficits in developmental dyslexia and the phonological representations hypothesis. *Journal of Experimental Child Psychology*, 66(1), 18–41. [https://doi.org/10.1016/S0022-0965\(02\)00169-8](https://doi.org/10.1016/S0022-0965(02)00169-8)
- Thomas, E., & Sénéchal, M. (2004). Long-term association between articulation quality and phoneme sensitivity: A study from age 3 to age 8. *Applied Psycholinguistics*, 25, 513–541.
- Torgesen, J., Wagner, R. K., Rashotte, C. A., Herron, J., Lindamood, P. (2010). Computer-assisted instruction to prevent early reading difficulties in students at risk for dyslexia: Outcomes from two instructional approaches. *An Interdisciplinary Journal of The International Dyslexia Association*, 60(1), 40–56.
- Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Herron, J. (2003/2018). Summary of outcomes from first grade study with ‘Read, Write, and Type’ and ‘Auditory Discrimination in Depth’ instruction and software with at-risk children. FCRR Technical Report #2. Florida Center for Reading Research. <http://files.eric.ed.gov/fulltext/ED587689.pdf>
- Trainin, G., Wilson, K. M., Murphy-Yagil, M., & Rankin-Erickson, J. L. (2014). Taking a different route: Contribution of articulation and metacognition to intervention with at-risk third-grade readers. *Journal of Education for Students Placed at Risk*, 19(3-4), 183–195. <https://doi.org/10.1080/10824669.2014.972103>
- What Works Clearinghouse, Institute of Education Science, U.S. Department of Education. (2008). WWC Intervention Report <https://files.eric.ed.gov/fulltext/ED503694.pdf>
- What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education. (2015). *Beginning reading intervention report: Lindamood Phoneme Sequencing® (LiPS®)*. <http://whatworks.ed.gov> <https://ies.ed.gov/ncee/wwc/Intervention/665>
- Wise, B. W., Ring, J., & Olson, R. K. (1999). Training phonological awareness with and without explicit attention to articulation. *Journal of Experimental Child Psychology*, 72(4), 271–304.
- Wolff, U. (2016). Effects of a randomised reading intervention study aimed at 9 year olds: A 5 year follow-up. *Dyslexia*, 22(2), 85–100. <https://doi.org/10.1002/dys.1529>



## **Artikkel 2**

Thurmann-Moe, A.C., Melby-Lervåg, M & Lervåg, A. Single Case Design i evaluering av spesialpedagogiske tiltak: En gjennomgang av aktuelle analysemetoder for vurdering av effekt.



## **Single Case Design i evaluering av spesialpedagogiske tiltak: En gjennomgang av aktuelle analysemetoder for vurdering av effekt**

### **Sammendrag**

Artikkelen drøfter hvordan Singel Case Design (SCD) kan brukes i vurdering av spesialpedagogiske tiltak og diskuterer spørsmål knyttet til validitet og alternative innfallsvinkler for vurdering av effekt. Det presenteres også et konkret eksempel fra en gjennomført studie. Det argumenteres for at SCD er spesielt egnet til vurdering av effekt av spesialpedagogiske tiltak for elever med lavfrekvente og/eller sammensatte vansker. I skolen kan SCD bidra til å øke kvaliteten av spesialundervisningen gjennom mer nøyaktige vurderinger av hva som virker for den enkelte elev, og også bidra til utvikling av forskningsbasert kunnskap for praksis..

### **Summary/ Abstract**

#### **Single Case Design in Evaluation of Special Needs Education: A Methodological Review and an Example.**

This paper presents how Single - Case Designs (SCD) can be used to evaluate interventions in the field of special needs education and discuss validity and alternative approaches for assessing efficacy. An example from a completed study is also presented. We argue that SCD is particularly suitable for assessing the effect of interventions for students with low-frequency and / or compound difficulties. In schools, SCD can help to improve the quality of special education delivery through more accurate assessments of what works.

**Nøkkelord: Spesialundervisning, Singel Case Design, effektstudier, evaluering**

## Innledning

Andelen av elever i Norge som mottar spesialundervisning etter enkeltvedtak er i skoleåret 2019/2020 ca. 8% (UDIR). Samtidig konkluderer forskning med uklart og i noen tilfeller også negativt utbytte av spesialundervisning (Kvande, Bjørklund, Lydersen, Belsky & Wichstrøm, 2018; Opsvik & Haug, 2017; Lekhal 2017; Skorpen, 2017).

Offentlige utredninger, sist utført av Barneombudet (2017) og det regjeringsoppnevnte Nordahl-utvalget (2018) samt stortingsmelding 6 (Regjeringen, 2019-2020) har også satt spørsmålsteget ved om bruken av ressurser til spesialundervisning fører fram. Rapportene konkluderer med svikt i mange ledd i tiltakskjeden knyttet til planlegging, gjennomføring, medvirkning fra elev og foresatte, samt evaluering og rapportering av igangsatte spesialpedagogiske tiltak.

Både av hensyn til den enkelte elev og med tanke på bruk av offentlig ressurser synes det i lys av dette å være et stort behov for nærmere vurdering av effekten av konkrete spesialpedagogiske tiltak. Vi vil i denne artikkelen drøfte hvordan Single Case Design(SCD) kan bidra til en sikrere evaluering av hvorvidt igangsatte tiltak for enkeltelever etter vedtak om spesialundervisning har effekt. Vi vil også diskutere hvordan SCD kan bidra til utvikling av mer forskningsbaserte tiltak i praksisfeltet, noe som i økende grad etterspørres av utdanningsmyndighetene (se f.eks. Stortingsmelding 21(Regjeringen 2016-2017)).

De fleste undersøkelser av effekt av spesialundervisning er gjort på gruppenivå der grupper av elever som har vedtak om spesialundervisning sammenliknes med grupper av elever som ikke har vedtak om spesialundervisning. Denne typen design har flere svakheter. For det første blir gruppene systematisk ulike fordi det er en grunn til at noen fikk enkeltvedtak mens andre ikke fikk det. For det andre kan gruppen som mottar spesialundervisning være svært heterogen (Skorpen, 2017).

Randomiserte kontrollerte studier (RCT) regnes i utgangspunktet for å være den sikreste metoden for å evaluere effekten av et tiltak, fordi flest mulig alternative forklaringer kan utelukkes (Maggin , Lane & Pustejovsky,2017; Odom et. al, 2005; WWC, 2017). Av etiske årsaker er det imidlertid vanskelig å designe studier der likeverdige elever med vedtak om spesialundervisning randomiseres i grupper der bare halvparten får tiltaket (Lekhal, 2017). I spesialpedagogikk arbeider man også ofte med lavfrekvente grupper hvor det er stor innen-gruppe heterogenitet. Dette kan gjøre det krevende å bruke dette designet til å vurdere effekt av tiltak rettet mot lavfrekvente grupper (Lobo et al., 2017; Onghena, Michiels, Jamshidi, Moeyaert & Van den Noortgate, 2018).

I et mer klinisk perspektiv har gruppedesign også begrensninger når det gjelder å vurdere effekt av tiltak for enkeltelever fordi en gjennomsnittseffekt basert på en gruppe bare gir en tentativ indikasjon for hvordan tiltaket faktisk virket for hver enkelt deltaker (Onghena, Michiels, Jamshidi, Moeyaert, & Van Den Noortgate, 2018).

I SCD prøves tiltaket direkte på en deltaker eller eventuelt en liten gruppe deltakere og gir direkte mål på effekt for dem som har deltatt i den situasjonen tiltaket er prøvd ut.



Grunnlaget for konklusjoner fra gruppestudier og SCD bygger dermed på noe ulik logikk (Onghena et al., 2018). I gruppedesign går man bredt ut og prøver tiltaket på mange deltakere som basis for å konkludere om antatt effekt for enkeltpersoner fra samme populasjon. I SCD blir sammenhengen mellom et gitt tiltak og eventuell endring av atferd hos den enkelte deltaker målt direkte.

SCD er særlig relevant når det gjelder å vurdere effekt av tiltak for enkeltelever eller små grupper innen vanskeområder som er for lavfrekvente eller for heterogene til at det er hensiktsmessig eller praktisk mulig å rekruttere store og randomiserte utvalg. Dette gjør metoden aktuell for store deler av det spesialpedagogiske fagfeltet (De Bruin, 2017; Horner et al., 2005; Maggin et al., 2018). SCD er også egnet som utgangspunkt for pilotering før gruppestudier (Krasny-Pacini & Evans, 2018). Ikke minst er SCD velegnet i studier der målet ikke først og fremst er å predikere effekt for fremtiden (generalisert effekt), men der man vil kvalitetssikre utprøving av tiltak for en enkelt person eller en definert klinisk gruppe i en praktisk kontekst (Onghena et al., 2018).

### Ulike Single Case Design

Single Case Design (SCD) er en av de eldste kjente metodene for å vurdere effekt av tiltak, og er særlig brukt i klinisk virksomhet på områder innenfor medisin, psykologi og spesialpedagogikk (Shadish, 2014; Shamseer et al., 2015; Tate et al., 2016). Single Case metodologien ble først etablert for å kunne igangsette kontrollerte eksperimenter med kun en deltaker, såkalte  $N=1$  studier (Kazdin, 2016; Kratochwill & Levin, 2010). Senere har designet blitt utviklet til også å omfatte flere  $N=1$  deltakere i en kjede (Multiple baseline design). (Gast & Ledford, 2014; Kazdin, 2016). Grunnlaget for alle SCD er først å definere observerbar atferd som kan måles over tid. Utover dette er metoden fleksibel med hensyn til valg av målemetode og både direkte observasjon, selvrapporteringsskjemaer og repeterte kartlegginger kan brukes. Av disse er direkte observasjon den tradisjonelt mest brukte (Klingbeil, Van Norman & Nelson, 2017).

Den overordnede logikken i SCD er at hver deltaker tjener som sin egen kontroll gjennom at det foretas repeterte målinger i en baselinefase (A) før man introduserer fasen med tiltak (B). Resultatet fra målingene i baselinefasen antas å være representativt for atferden eller prestasjonene før intervensjonen starter og validiteten styrkes ved at det gjennomføres flere målinger, ideelt sett minst fem (Kratochwill et al. 2013; Tate et al., 2016). Målingene fortsetter gjennom intervensjonsfasen og deretter sammenliknes resultatene fra de to fasene. Slike enkle tofase AB design er utgangspunktet for alle andre SCD (Ledford, 2018) og kan brukes til å vurdere kvalitative endringer hos enkeltelever over tid (Shadish, 2014) men fyller ellers ikke forskningsmessige krav til eksperimentell kontroll (Ledford, 2018; Lobo, Moeyaerth, Cunha & Babik, 2017; Tate et al., 2016).

I konsensusrapporten fra The Single Case Reporting Guideline in Behavioural Interventions (SCRIBE) (Tate et al., 2016) skilles det derfor mellom to fase design og *eksperimentelle* design. Skillet angår først og fremst i hvilken grad designet inkorporerer replikasjoner enten ved å prøve tiltaket på flere elever, situasjoner eller fag i en kjede der tiltaket introduseres til ulik tid (multiple baseline design), eller ved å systematisk «skru» tiltaket av og på for samme deltaker gjennom flere enn to faser. For å kunne konkludere med funksjonell sammenheng mellom tiltaket og resultatet har det vært foreslått at effekten bør repliseres tre ganger, det vil si på tre deltakere, i tre situasjoner eller gjennom tre faser av studien (Ledford, 2018; Lobo et

al., 2017; Maggin, Cook & Cook, 2018). En underliggende forutsetning er også at data fyller anbefalte kvalitetskrav knyttet til både design og egenskaper ved innsamlede data (se nedenfor).

### Reversible design – ABAB-design

For undersøkelser der man ønsker å vurdere en umiddelbar atferdsendring hos elever, for eksempel om intensivert muntlig ros fra lærer fører til økt elevaktivitet i timen, er det mulig å bruke design som skruer tiltaket av og på fra fase til fase, såkalte ABAB-design. I ABAB-design er vurdering av effekt knyttet til hvorvidt avhengig variabel (for eksempel økt elevaktivitet) systematisk øker hver gang tiltaket i form av positiv forsterkning settes i gang og så avtar når tiltaket fjernes. ABAB-design kan også utvides med flere faser (f.eks. ABABAABAAB) eller man kan introdusere et tiltak til (C; så designet blir: ABABACAC) (Gast & Backey, 2014).

### Irreversible design - Multiple baseline Design (MBD)

For tiltak som tar sikte på en varig endring av elevens atferd, for eksempel ved å bedre ferdigheter i et akademisk fag, kan ikke tiltaket reverseres. I *Multiple Baseline Design* (MBD) (Baer, Wolf & Risley, 1968; Gast, Lloyd & Ledford, 2014) introduserer man samme tiltak til et avgrenset antall elever (som oftest tre eller flere) på ulike tidspunkter langs en tidslinje. Målingene i baselinefasen foregår kontinuerlig for alle deltakerne, men tiltaket settes i verk til ulik tid. Forutsatt at øvrige kvalitetskrav er oppfylt konkluderes det med positiv effekt av tiltaket dersom det demonstreres positiv endring av avhengig variabel som for minst tre deltakere kan assosieres med tidspunktet tiltaket ble satt i verk. Forholdstallet synes å ta utgangspunkt i at MBD design vanligvis inkluderer 3-4 deltakere (Shadish & Sullivan, 2011) og at noen deltakere potensielt må ekskluderes fra vurderingen på grunn av ustabile data, for få datapunkter e.a.

En variant av multiple baselinedesign er multiple probe design (Horner & Baer, 1978) der målingene langs tidslinjen ikke foregår kontinuerlig, men organiseres i planlagte bolker. Denne typen design er spesielt egnet hvis det er usannsynlig med store variasjoner i baselinefasen. Fordelen med denne varianten er også at den minsker risikoen for at elevene blir enten lei av eller for godt kjent med kartleggingsprosedyrene (Horner & Baer, 1978). MBD/probe design kan også brukes til å undersøke om et tiltak for samme elev virker i flere fag eller situasjoner, og tiltaket introduseres da på ulikt tidspunkt i de ulike settingene (minst tre).

## Kausalitet og evidens

I SCD er den indre validiteten knyttet til i hvilken grad man har eksperimentell kontroll med at eventuell endring fra baseline til intervensjonsfase skyldes det igangsatte tiltaket og ikke andre forhold. Nedenfor vil vi gå gjennom de viktigste truslene mot den indre validiteten i SCD og konkretisere hvordan disse truslene kan møtes (Gast, 2014; Tate et al., 2016).

### Trusler mot indre validitet og begrepsvaliditet

**Historie:** I den perioden tiltaket foregår kan det også inntre andre hendelser som påvirker den variabelen man ønsker å undersøke. Det kan være at konkurrerende tiltak som man ikke kontrollerer settes i verk relativt samtidig, det kan være større endringer i elevens liv, eller mindre forbigående episoder (Gast, 2014). Slike hendelser er ikke lett å kontrollere for, men

nøyaktige nedtegnelser, gjerne som loggføring, både fra undervisning og kartlegging kan bidra til å forklare endringer som skyldes eksterne faktorer.

**Modning:** Hvis tiltaket foregår over lang tid vil man måtte ta i betraktning at det kan skje en modning av ferdigheter ettersom tiden går, uavhengig av tiltaket. Ifølge Gast (2014) er dette spesielt aktuelt dersom tiltaket gjelder unge barn og strekker seg over fire til seks måneder.

**Test- /retest- effekt:** Repeterte målinger av samme fenomen er kjernen i ethvert single-case eksperiment, men innebærer samtidig en trussel mot validiteten. Repeterte målinger kan føre til at eleven går lei og presterer svakt som følge av dette. En annen mulighet er at kjennskap til testprosedyren fører til bedre prestasjoner (Gast, 2014). For å unngå at eleven blir for godt kjent med oppgavene anbefales det å randomisere rekkefølgen av oppgaver og testledd fra gang til gang (ibid). Ved visuell inspeksjon (se nedenfor) regnes endring av data som kan assosieres med tidspunkt for oppstart av intervensjonen, og som ikke forekommer eller er mindre synlige i baselinefasen som indikasjon på at resultatet i intervensjonsfasen ikke er påvirket av testeffekt (Gast, Lloyd & Ledford, 2014).

**Hawthorne-effekt.** Repeterte målinger innebærer økt oppmerksomhet rundt eleven. Eleven blir også kjent med at han er med på et eksperiment og det er relevant å diskutere hvorvidt det å være gjenstand for oppmerksomhet i seg selv virker inn på prestasjonene (Shadish, Cook, & Campbell, 2002). Det at prosedyrene i forbindelse med kartleggingen er nye for eleven kan innvirke både i positiv og negativ retning. For å styrke validiteten anbefales det å la eleven bli kjent med prosedyrene og eventuelle fremmede testledere før man starter datainnsamlingen (Gast, 2014). Såkalt *Performance effekt* kan også oppstå både hos elev og andre involverte, for eksempel testledere og lærere. Dette innebærer at atferden økes i positiv retning fordi man blir observert. Dette kan bety at den målte effekten ikke er kun knyttet til tiltaket, men også til det å delta i et eksperiment.

**Stabilitet:** Ustabile data, enten som følge av faste sykliske variasjoner, for eksempel hvis data ved daglige målinger alltid er bedre på en fredag enn en mandag, eller data med stor variasjon fra måling til måling internt i fasen, er en trussel mot validiteten og kan forstyrre tolkingen av resultatene (Gast, 2014). Stabilitet i data er også en sentral forutsetning for at en studie skal oppnå de kvalitetskravene som kreves for å kunne vurdere potensiell effekt av et tiltak (Kratowill et al., 2010).

**Implementeringsvaliditet:** Implementeringsvaliditet er avhengig av at prosedyrer i forbindelse med datainnsamling og selve tiltaket gjennomføres som planlagt. For å sikre dette bør det benyttes opptak enten med lyd/bilde eller bare lyd som gjør det mulig for en annen fagperson å vurdere og tolke data (Gast, 2014).

**Sosial og økologisk validitet:** I SCD legges det vekt på sosial validitet, altså hvorvidt tiltaket oppfattes positivt og relevant i det aktuelle miljøet, og økologisk validitet, som dreier seg om hvorvidt tiltaket er prøvd ut på en måte og i en situasjon som er representativ for praksisfeltet (Gast, 2014). Det er derfor viktig at tiltak prøves ut i en naturlig kontekst. Vurdering av sosial validitet gjennom intervju/spørreskjema til viktige nærstående personer, f.eks. lærere og foreldre, tas i betraktning når det gjelder vurdering av videre bruk av tiltaket, og anbefales også brukt for å validere resultatene (Tate et al., 2016)

## Ytre validitet og effektbegrepet i SCD

Historisk har SCD vært et verktøy for evaluering av tiltak i en klinisk kontekst. Tradisjonelt er en enkeltstående SCD studie også først og fremst et eksempel på hvordan et aktuelt tiltak virker i den situasjonen det er prøvd ut og resultatene fra en enkelt studie kan derfor i utgangspunktet ikke generaliseres. Ved å sikre at antall målepunkter i baselinefasen er tilstrekkelige og ved å møte aktuelle trusler mot den indre validiteten kan imidlertid enkeltstående SCD oppnå god eksperimentell kontroll (Tate et al., 2016). Når studien også fyller krav til presise casebeskrivelser og nøyaktige beskrivelser av det utprøvde tiltaket kan resultatene gi et forskningsbasert grunnlag for utprøving på nye elever med liknende status (Maggin et al., 2018).

Dersom SCD skal bidra til å dokumentere *generalisert effekt* av et gitt spesialpedagogisk tiltak kreves det at effekt repliseres gjennom gjentatte SCD studier og metaanalyser. For at et tiltak vurdert med SCD skal kunne regnes som forskningsbasert har det vært foreslått at positiv effekt av tiltaket må være dokumentert i minst fem uavhengige studier med til sammen minst 20 N=1 design, utført av minst tre ulike forskningsmiljøer («3-5-20 kriteriet»)(Hitchcock, Kratochwill & Chezan, 2015; Maggin et al., 2018; WWC, 2017). Inkludert i disse kriteriene er det også kvalitetskrav knyttet både til overordnet design og til egenskaper ved innsamlede data. For nærmere diskusjon om kvalitetskrav og kriterier for vurdering av effekt i SCD se ellers Zimmerman et al. (2018 a) for beskrivelse og sammenlikning av ulike evidensstandarder.

Som en konsekvens av etterspørselen etter såkalt forskningsbaserte undervisningstiltak har det i USA, og etter hvert også i Europa, blitt etablert åpne databaser over undervisningstiltak som fyller kriterier for å regnes som effektive (se f.eks. *What Works Clearing House* (WWC)) Siden SCD ofte foregår i praksisnære kontekster har det vært ønskelig å kunne inkludere resultater fra SCD i vurdering av konkrete undervisningsprogrammer (Horner et al., 2005; Kratochwill et al., 2013; Shadish, 2012, 2014). Flere har imidlertid vært kritiske til at databasenes standarder primært har tatt utgangspunkt i kvalitetskrav innen gruppedesign metodologi og beklaget konsekvensen dette har fått for metodeutviklingen i SCD i retning av økte krav til generaliserbarhet (de Bruin, 2017; Horner et al., 2005; Ledford, Wolery & Gast, 2014; Maggin et al., 2017).

Hovedinnvendingene har vært at lavfrekvente vanskeområder risikerer å bli underrepresentert i forskningen og at fordelene ved småskalastudier blir oversett (Onghena et al., 2018).

Parallelt har det, primært med tanke på å kunne gjennomføre metaanalyser, blitt lansert statistiske analyseverktøy for vurdering av samlet deltakereffekt i SCD (Shadish, Hedges, & Pustejovsky, 2013). Dette har hatt som formål å kunne presentere synteser fra flere enkeltstående SCD på en slik måte at de kan sammenliknes med resultater fra gruppestudier (se f.eks. Shadish et al., 2015). I oppdaterte retningslinjer fra WWC (2020) er bruk av design - uavhengige effektmål et krav og det antydes at resultater fra denne typen mål delvis kan erstatte det tidligere 3-5-20 kriteriet.

## Metoder for dataanalyse

Analyse av data i SCD har tradisjonelt basert seg på metoder som ikke forutsetter tilgang på programvare eller kompetanse som ikke er tilgjengelig i en klinisk hverdag. Dataanalysen baseres på visuelle inspeksjoner av grafiske framstillinger av resultatene og/eller ulike kvantitative effektmål. De siste årene har det også blitt utviklet nettbaserte kalkulatorer for utregning av effekt. Nedenfor vil vi presentere noen aktuelle visuelle analysemetoder og et utvalg kvantitative effektmål. Listen er absolutt ikke fullstendig og i utvelgelsen er det blitt lagt vekt på metoder som er ofte referert i publiserte studier og som har god tilgjengelighet ved at de utføres enten for hånd, ved hjelp av ordinær programvare på PC eller via gratis nettressurser.

### Visuell inspeksjon av data

Visuell inspeksjon har historisk sett vært den mest brukte metoden for å vurdere effekt i SCD (Gast & Ledford, 2014; Lane & Gast 2014; Shadish, 2014). Den visuelle inspeksjonen av data har primært to formål. Det første er å undersøke hvorvidt egenskaper ved data er tilfredsstillende som grunnlag for å vurdere eventuell effekt. Her vurderes særlig stabilitet og trender i data. I klinisk sammenheng kan dette også inngå som et arbeidsredskap i gjennomføring av studien for eksempel ved at man avventer oppstart av intervensjonsfasen til data i intervensjonsfasen er stabile (Gast & Spriggs, 2014). Det andre er å vurdere hvorvidt det kan påvises en funksjonell sammenheng mellom oppstart av tiltak og endring i avhengig variabel. Denne vurderingen angår grad av overlapp i datapunkter mellom fasene, eventuell umiddelbar endring i grafens helningsgrad ved oppstart av intervensjon og nivåforskjeller mellom fasene. Visuelle analyser angir ikke grad eller størrelse på sammenheng, men er dikotom (Gast & Ledford, 2014). Det er utviklet retningslinjer for vurdering av trender, stabilitet og overlapp av data som bidrar til å øke validiteten av de visuelle analysene (Gast & Spriggs, 2014; Lane & Gast, 2014). Hovedpunkter når det gjelder disse retningslinjene presenteres nedenfor (Kratochwill et al., 2010; Lane & Gast, 2014).

#### Stabilitet

Stabilitet i data regnes ut for hver fase ved at man først finner medianverdien i baselinefasen og så etablerer et intervall rundt medianen som utgjør +/- 25 prosent av verdien. Dette intervallet brukes også i intervensjonsfasen (Gast & Spriggs, 2014; Lane & Gast, 2014). For at data i fasen skal vurderes som stabile kreves det at 80 % av datapunktene skal være innenfor intervallet. Stabilitet internt i fasene er viktig for å kunne foreta en gyldig sammenlikning av det gjennomsnittlige nivået i hver fase. Ideelt sett er det anbefalt å fortsette baselinemålingene og vente med å introdusere tiltaket inntil man har oppnådd et stabilt resultat (ibid).

#### Trender

Trender i data vurderes gjennom at man tegner opp trendlinjer for hver fase. Det er relevant å avdekke hvorvidt det skjer en endring i trendens helningsgrad ved oppstart av intervensjonen, og det er også relevant å vurdere hvorvidt det er tydelige og stabile trender i baselinefasen fordi dette kan være en begrensning når det gjelder vurdering av effekt.

Ved «split-middle»-metoden danner medianverdien fra hver halvpart av hver fase utgangspunkt for å trekke opp trendlinjer (Gast & Spriggs, 2014). Stabiliteten vurderes ved å

trekke opp et intervall rundt trendlinjen og regne ut hvor stor prosentandel av punktene som faller innenfor intervallet (Gast & Spriggs, 2014; Lane & Gast, 2014; Manolov et al., 2018). Størrelsen på intervallet og utregningsmåten er den samme som beskrevet i avsnittet om stabilitet over. Intervallet angir +/-25 % av medianverdien i baselinefasen og avvik fra dette intervallet for mer enn 20 % av datapunktene tolkes som at observert trend er upålitelig. En pålitelig trend skal være visuelt observerbar, men kan også regnes ut ved en enkel regresjonsanalyse. Her vil en lav korrelasjon/forklart variasjon ( $R^2$ ) mellom tidsvariabelen og avhengig variabel angi en upålitelig trend og vice versa (Manolov, et al., 2018). Bruk av regresjonsanalyse forutsetter imidlertid at trenden er lineær og i tilfeller man er usikker anbefales den manuelle metoden (Manolov et al., 2018).

En metode for vurdering av trender er også å predikere trendlinjen basert på data i baselinefasen og så se i hvilken grad den observerte trenden i intervensjonsfasen avviker fra den predikerte (Kazdin, 2011). Vurdering av effekt av et tiltak basert på ekstrapolering av trendlinjen fra baselinefasen forutsetter imidlertid at observert trend i baselinefasen automatisk videreføres over tid. Undersøkelser av publiserte SCD studier har vist at dette i mange tilfeller ikke er tilfelle (Parker, Vannest, Davis & Sauber, 2011).

En begrensning ved trendanalysene er at det ikke finnes gyldige standarder for hva som regnes som en akseptabel endring av grafenes helningsgrad for vurdering av effekt i ulike design og med ulike målemetoder (Busse, Mc Gill & Kennedy 2015; Manolov & Vannest, 2019). Det har også blitt påpekt at kriteriet om umiddelbar endring av trend i terapeutisk retning ved oppstart av intervensjonen ikke er relevant for studier der målet er bedrede akademiske ferdigheter, siden dette krever tid å etablere (Klingbeil, Norman & Nelson, 2017).

### Nivåendringer

Nivåendringer mellom fasene kan uttrykkes gjennom sammenlikning av gjennomsnittet eller medianverdien av resultatene for hver fase.

### Overlapp

Graden av overlapp mellom datapunkter i hver fase regnes ut ved at man teller opp antall punkter i intervensjonsfasen som er høyere (eventuelt lavere) enn det høyeste/laveste datapunktet i baselinefasen (se også PND nedenfor).

### Kvantitative effektmål

Effektstørrelser har betydning for å kunne kvantifisere størrelsen eller graden av effekt og skal ideelt fungere slik at det er mulig å sammenlikne ulike studier (Hedges 2008). Effektmålene i SCD er ikke direkte sammenliknbare på tvers av de ulike måleverktøyene, men bruker egne skalaer. Det skilles mellom tre «familier» av kvantitative mål i SCD: Overlappingsteknikker, Parametriske effektmål og Mål for samlet deltakereffekt (Pustejovsky & Ferron, 2017).

### Overlappingsteknikker

Det grunnleggende felles prinsippet for overlappingsteknikkene er å vurdere graden av overlapp mellom målinger gjort i baseline fasen sammenliknet med intervensjonsfasen. Dette prinsippet er basis for en rekke utregningsmetoder og teknikker for vurdering av effekt (se Parker, Vannest & Davis, 2011 for en gjennomgang).

*Percent of Nonoverlapping Data (PND)* (Scruggs, Mastropieri, & Casto, 1987).

Prosentandelen av datapunkter fra intervensjonsfasen som ikke overlapper med datapunkter i

baselinefasen (PND) er den enkleste av overlappingsteknikkene, og regnes ut ved at man først finner det høyeste datapunktet i baselinefasen (eller laveste hvis målet for studien er reduksjon av verdiene). Deretter teller man opp hvor mange datapunkt i intervensjonsfasen som ligger over (eller under) dette. Prosenttallet regnes ut ved at dette antallet divideres på antall målepunkter i intervensjonsfasen. For PND er den foreslåtte standarden at 50 og mindre viser dårlig effekt, skårer mellom 50 og 70 usikker effekt, skårer fra 70 til 90 god effekt og skårer over 90 svært god effekt (Scruggs & Mastropieri, 1998). Kritikken mot PND er knyttet til at et isolert (og mulig lite representativt) målepunkt i baselinefasen tillegges for stor vekt ved utregning (Maggin et al., 2011). Som et alternativ er det derfor foreslått å bruke medianverdien i baselinefasen, PEM (*The Percentage of datapoints in phase B exceeding the Median of the Baseline Phase*), som sammenlikningsgrunnlag (Ma, 2006; Parker et al., 2011). Andre aktuelle varianter av PND er PAND (*Percentage of All Nonoverlapping Data*) og NAP (*Nonoverlap of All Pairs*) (Parker et al., 2011).

*Tau-U* (Parker, Vannest, Davis & Sauber, 2011) er en effektstørrelse som kombinerer graden av overlapp mellom fasene med kontroll for trend i baselinefasen. Det anbefales å kun kontrollere for trend i baselinefasen dersom trenden er statistisk signifikant (Parker et al., 2011). *Tau-U* er ikke-parametrisk og forutsetter dermed ikke normalfordelte data. Utregningen baseres på kombinasjon av to metoder, Mann-Whitney U og Kendalls Tau (Parker et al., 2011). Begge disse utregningsmåtene baseres på sammenlikning av alle teoretisk mulige kombinasjoner av dataverdier fra de to fasene, og *Tau U*-estimatet er uttrykket for prosentverdien av alle datapunkter som har økt verdi i intervensjonsfasen sammenliknet med baselinefasen. Det er mulig å regne konfidensintervall og signifikansnivå for effektstørrelsen.

Effektstørrelsen angis som et estimat mellom 0 og 1. For *Tau-U* regnes estimat under .20 som liten effekt, mellom .20 og .60 som moderat effekt, fra .60 - .80 som stor effekt og over .80 som svært stor effekt (Vannest & Ninci, 2015). Diskusjoner om *Tau-U* som effektmål har blant annet omhandlet vurderingen av baselinetrender (Manolov et al., 2018; Tarlow, 2017) og andre utregningsmetoder enn den som er gjort tilgjengelige av opphavspersonene er foreslått (se f.eks. Tarlow 2016; 2017 og Pustejovsky & Swan, 2018) for alternativer). Et viktig moment i denne diskusjonen er imidlertid at de alternative metodene for trendkontroll forutsetter en lineær baselinetrend, mens den originale *Tau-U* kontrollerer for trender med både lineær og ikke lineær variasjon

Online kalkulatorer for utregning av PND og *Tau-U* er tilgjengelig både via Vannest, Parker, Gonen & Adiguzel, (2016) og fra Pustejovsky & Swan (2018).

### Parametriske effektmål

*Standardized Mean Difference* (SMD) og varianter av dette (Busk & Serlin, 1992; Cohen, 1988; Hegdes, Pustejovsky & Shadish, 2012) angir gjennomsnittlig endring fra baselinefasen til intervensjonsfasen og regnes ut ved at gjennomsnitt for baselinemålingene for hver elev trekkes fra gjennomsnittsverdien fra intervensjonsfasen og deles på standardavviket i baselinefasen eller på standardavviket for både baseline og intervensjonsfasen. Tolkning av effektstørrelsen for SMD basert på gjennomgang av 52 publiserte SCD fra 2010-årgangen av *Journal of Applied Behavior Analysis* (Harrington & Velicer, 2015), har foreslått følgende tommelfingerregler: 0–1 er liten effekt, 1–2,5 er middels effekt og større enn 2,5 er stor effekt. Kalkulator for utregning av SMD er tilgjengelig online (Pustejovsky & Swan, 2018).

*Log Respons Ratio* (LRR) (Pustejovsky, 2018 a) er et effektmål som også sammenlikner gjennomsnittet fra intervensjonsfasen med gjennomsnittet fra baselinefasen. Det er ikke utviklet standard for LRR, men effekttestimatet kan konverteres til et mål på prosentvis økning eller reduksjon av målt atferd i intervensjonsfasen sammenliknet med baseline. LRR er kun aktuell dersom måleskalaen som brukes har et naturlig nullpunkt og er primært prøvd ut i studier basert på direkte atferdsobservasjon (ibid).. I situasjoner med gulv- og takeffekter, for eksempel hvis man måler utvikling av en ny ferdighet og baselinemålene er på null, vil enhver økning i intervensjonsfasen innebære en stor prosentvis endring og vil derfor være misvisende (ibid). Også for *Log Respons Ratio* (LRR) er det kalkulator tilgjengelig (Pustejovsky & Swan, 2018).

#### Mål for samlet deltakereffekt

*Between Case – Standardized Mean Difference* (BC-SMD) er et nylig lansert effektmål som bruker samme standard som i gruppestudier (Valentine, Tanner Smith, Pustejovsky & Lau, 2016). Denne utregningsmåten angir ikke individuell effekt, men en samlet effektstørrelse for alle deltakerne. BC-SMD er foreløpig kun tilgjengelig for ABAB design og for MBD/probe design mellom deltakere. Effektstørrelsen regnes ut ved flernivåanalyse med mulighet for å la de individuelle deltakerne variere både med hensyn til utgangspunkt, oppnådd effekt og eventuelle trender i data. Utregningen av effektstørrelse baseres på Restricted Maximum Likelihood (REML) (Valentine et al., 2016). Effektmålet oppgis med konfidensintervall og signifikansnivå og bruker samme skala som for Cohens  $d$ . Det vil si at  $d = .20$  regnes som liten effekt,  $d = .50$  som moderat effekt og  $d = .80$  som stor effekt (Cohen, 1988). Denne utregningsmåten gjør at resultater med BC-SMD teoretisk kan sammenliknes med resultater fra gruppestudier på samme skala (Shadish, Hedges, Horner & Odom, 2015). Det finnes en online-kalkulator for BC-SMD der man laster inn filer med egne data (se Pustejovsky, 2016)

Også ved bruk av online kalkulatoren for TAU-U (se over) er det mulig å regne ut en samlet deltakereffekt. Denne regnes ut som en vektet gjennomsnittsverdi av de individuelle effektstørrelsene for et gitt utvalg av deltakere i samme studie. Det kombinerte effekttestimatet bruker samme standard som Tau-U for individuell effekt og oppgis også med konfidensintervall og signifikansnivå.

#### Vurdering av de ulike effektmålene

Et stort utvalg av ulike analysemetoder og teknikker for vurdering av effekt i SCD er altså tilgjengelig og flere studier har sammenliknet de ulike metodenes egnethet både med bruk av simulerte data og ved reanalyse av allerede publiserte studier (Manolov et al., 2010; Wolfe, Dickenson, Miller, and McGrath, 2019; Zimmerman et al., 2018 a).

Manolov og kolleger (2010) undersøkte i hvilken grad effektmålene PND og variantene PEM og PAND ble påvirket av grad av autokorrelasjon og antall målepunkter (serielengde) i hver fase. De brukte simulerte data der de manipulerte både autokorrelasjon og serielengde og vurderte så hvordan dette påvirket effektstørrelsen fra de forskjellige målemetodene. Resultatene viste at PND og PAND begge ble påvirket av autokorrelasjon og at dette førte til både overestimering og underestimering av resultatene. PEM var mer robust både for autokorrelasjon og serielengde.

Pustejovsky (2018 b) undersøkte i hvilken grad ulike effektmål påvirkes av datainnsamlingsprosedyrer. Han sammenliknet fem forskjellige effektmål basert på overlapp



(PND, PEM, NAP, RIRD, Tau) med to parametriske effektmål (SMD og LRR). Han gjennomførte analyser med simulerte data og vurderte så de ulike effektmålenes sensitivitet for variasjon i antall målepunkter i hver fase, observasjonslengde og system for datainnsamling. Han fant at de parametriske metodene generelt var mer robuste når det gjaldt påvirkning fra forhold knyttet til prosedyrene enn effektmålene basert på overlapp. Dette gjaldt særlig for LRR som i denne undersøkelsen viste seg å gi mer presise estimater enn SMD. En annen undersøkelse som sammenliknet de samme effektmålene med utgangspunkt i publiserte SCD studier som vurderte effekten av sensorisk stimulering i autistiske utvalg (Zimmerman et al., 2018 b), kom til samme konklusjon. Også her ble LRR vurdert som det effektmålet som samlet sett var det mest valide, mens Tau-U ble vurdert som det mest valide av teknikkene basert på overlapp.

I en annen undersøkelse som brukte simulerte data til å vurdere ulike effektmåls sensitivitet i forhold til autokorrelasjon kom imidlertid SMD godt ut (Manolov, Solanas & Sierra, 2010). Denne undersøkelsen indikerte også at SMD er egnet til å fange opp effekt også i datasett med relativt få målepunkter.

Andre undersøkelser har konkludert med at det til tross for ulikheter synes å være relativt stort samsvar både mellom ulike kvantitative effektmål og også mellom visuelle analyser og kvantitative effektmål. Wolfe et al. (2019) undersøkte i hvilken grad konklusjoner fra visuelle analyser samsvarte med resultater fra fire utvalgte statistiske metoder (IRD, Tau-U, Hedges g og BC-SMD). De rekrutterte et utvalg på 52 eksperter og disse ble bedt om å gjøre selvstendige visuelle analyser av 25 grafer fra publiserte Multiple Baseline Design studier. Forskerne gjorde selv de statistiske analysene ved å gå inn i det publiserte materialet. De sammenliknet så råskårene fra de fire statistiske analysene for å vurdere i hvilken grad disse kom til samme resultat. De vurderte også i hvilken grad ekspertene var samstemte i konklusjonene og i hvilken grad resultatene fra de visuelle analysene og de statistiske analysene samsvarte. Resultatene viste at for 67 % av de publiserte studiene kom de visuelle analysene og samtlige statistiske analyser til samme konklusjon når det gjaldt hvorvidt studien hadde hatt effekt. Det var også relativt høy grad av samsvar mellom de ulike måleverktøyene (Spearman's  $r = .65 - .89$ ). For de resterende studiene, der ekspertene var uenige i konklusjonene, viste de statistiske målene både stor grad av samstemmighet og en gjennomgående mindre streng vurdering.

Som beskrevet over kan visuelle analyser fange opp detaljer i studien som ikke kommer fram ved statistisk analyse (Lane & Gast 2014). Samtidig har undersøkelser vist at både hvem som tolker dataene og hvordan grafene teknisk framstilles virker inn på vurderingene, særlig i tilfeller med variable data (Busse, Mc Gill & Kennedy, 2015; Maggin & Odom, 2014; Pustejovsky, 2018 b). Når det gjelder kvantitative effektmål er den viktigste kritikken mot de parametriske effektmålene at de fleste effektmålene i utgangspunktet forutsetter egenskaper ved data som er vanskelig å realisere i SCD. Dette gjelder særlig utfordringer knyttet til autokorrelasjon, trender og data som ikke er normalfordelte (Lobo et al., 2017). De siste årene har det blitt lansert flere alternative metoder for å inkludere kontroll for autokorrelasjon og trend i de parametriske statistiske modellene (f.eks. Pustejovsky, Hegdes & Shadish, 2014; Swaminathan et al., 2014). Et problem er imidlertid at mange av modellene forutsetter lineære trender og gir upresise estimater dersom de brukes på data som ikke er lineære (Manolov, Solanas & Sierra, 2018). Disse metodene er forutsetter også tilgang til spesiell programvare og statistisk kompetanse, noe som kan gjøre dem lite anvendelige for praksisfeltet. Et unntak er den omtalte BS-SMD der det finnes en effektkalkulator. Denne er imidlertid kun for utregning av samlet effekt og kan ikke brukes for å vurdere effekt hos enkeltdeltakere.

De ikke-parametriske overlappingsteknikkene, som tradisjonelt har vært antatt å være mindre påvirket av autokorrelasjon (Kratochwill et al., 2013), har på sin side blitt kritisert for å være upresise. Overlappingsteknikker påvirkes også i større grad av forhold knyttet til prosedyrer for datainnsamling, som for eksempel antall målepunkter, observasjonsmetode og observasjonsintervall (Manolov, Solanas & Leiva, 2010; Pustejovsky, 2018 b).

Som vist har både visuelle analyser og samtlige kvantitative effektmål begrensinger. I henhold til standarder for SCD (Kratochwill et al., 2013; Tate et al., 2016) er det per i dag heller ikke enighet om hvilke effektmål som er best egnet. Det synes imidlertid å være enighet om at visuelle analyser og kvantitative mål utfyller hverandre og at det derfor er nyttig å bruke begge metoder i vurdering av resultatene (Kratochwill et al., 2010; Maggin & Odom, 2014; Manolov et al. Sierra, 2018; Shadish et al., 2015). Det har også blitt foreslått at visuelle analyser brukes først og at man dersom man finner at studien har hatt effekt går videre med statistiske analyser for å kunne kvantifisere effektstørrelsen (Lobo et. al, 2017). Når det gjelder statistiske løsninger har det blitt foreslått å bruke Tau-U for vurdering av individuell effekt dersom det er tydelige observerte trender i baselinedata (Rakap, 2015). I nylig oppdaterte retningslinjer for evidensstandarder i SCD (WWC, 2020) er det også inkludert krav om vurdering av samlet deltakereffekt.

## SCD – et eksempel

I det følgende vil vi presentere data fra en multiplebaseline/probe studie der vi vurderte effekten av et konkret spesialpedagogisk tiltak. Elevene som presenteres her kom fra fire forskjellige skoler og ble henvist Statped på grunn av vedvarende lese og skrivevansker (dysleksi) og lite utvikling til tross for langvarig spesialpedagogisk bistand. Elevenes leseproblemer var særlig knyttet til upresis fonologisk avkoding og tiltaket ble rettet mot å øke elevenes lesenøyaktighet. Tiltaket ble gjennomført som individuell trening fire timer pr uke over åtte uker og tok sikte på å lære elevene å bruke artikulasjon som støtte i ordavkoding. Intervensjonsprogrammet besto av konkrete læringsaktiviteter knyttet til å opparbeide artikulatorkompetanse. Studien ble realisert gjennom et samarbeid mellom lokal skole, Pedagogisk Psykologisk Tjeneste (PPT) og Statped.

Vi vil her kort redegjøre for den praktiske gjennomføringen av denne studien. Presentasjonen inneholder bare et utvalg data fra den originale undersøkelsen og de aktuelle elevene er valgt ut for å illustrere problemstillinger ved visuelle og statistiske analyser knyttet til ulikheter i individuelle data. Lærervurderinger, nærmere casebeskrivelser og presentasjon av tiltaket er utelatt.

## Beskrivelse av design

Mål for studien var å vurdere effekten av et konkret tiltak på et avgrenset område knyttet til elevenes lesestrategier. Vi ønsket også å vurdere hvorvidt den eventuelle effekten holdt seg eller ble borte når tiltaket ble avsluttet. Tiltaket ble gjennomført med utgangspunkt i elevenes ordinære timer til spesialundervisning, men for noen av elevene som deltok måtte skolene omdisponere ressurser i åtte uker for å realisere individuell trening. For å oppnå et sikkert grunnlag for å vurdere effekt ble det gjennomført fem kartlegginger i baselinefasen (Tate et al., 2016). Målingene i denne fasen er uttrykk for elevenes utgangspunkt før tiltaket startet. Kartleggingene ble gjennomført av PPT og tidspunkt ble delvis tilpasset deres rutiner for besøk på skolene. Siden intervensjonen tok sikte på en varig endring av ferdigheter ble det

valgt et multiple baseline design, men med elementer fra multiple probe design ved at målingene i baseline og postfasen ikke ble foretatt kontinuerlig, men organisert som blokker

### Operasjonalisering av avhengig variabel

Målet for tiltaket var å bedre elevenes fonologiske lesestrategi. Non-ord lesing regnes som et valid mål på fonologisk lesing (Pennington & Bishop, 2009) og det ble valgt å operasjonalisere begrepet «fonologisk lesestrategi» til «antall non ord lest korrekt på ett minutt». Dette er dermed avhengig variabel, mens uavhengig variabel er treningen som ble gitt gjennom de åtte ukene tiltaket varte. Som mål ble det brukt både antall korrekt leste ord pr minutt (råskåre) og prosentvis andel av ord lest riktig.

### Tiltak for å styrke validiteten

I planlegging av studien ble det gjort tiltak for å styrke validiteten. Det ble særlig lagt vekt på å minske risikoen for såkalt testeffekt av de gjentatte målingene. For å redusere risiko for at elevene skulle gå lei av testingen ble det som nevnt benyttet en «probe» variant av multiple baseline design, noe som innebar færre målinger i pre og post fasen av studien sammenliknet med et klassisk multiple baseline design (Horner & Baer, 1978). For å redusere risiko for at elevene skulle huske oppgavene fra gang til gang laget vi unike ordlister til hver kartlegging som besto av likeverdige, men ikke identiske testledd. Siden lesing av ordlister med tidsbegrensning kan være sårbart for forbigående uoppmerksomhet leste elevene to unike lister hver gang de ble kartlagt og et gjennomsnitt av disse to ble lagt til grunn. For å minske risiko for «performance» effekt ble kartleggingen foretatt av en annen fagperson enn den som hadde ansvaret for undervisningen.

### Vurdering av resultatene

I dette eksemplet vil vi legge vekt på å vise noen grunnleggende metoder som kan være egnet som utgangspunkt for evaluering av spesialpedagogiske tiltak i praksis.

Vi vil først vise hvordan visuelle analyser kan brukes for å vurdere kvaliteten av data som er samlet inn og gi tentative indikasjoner på virkning av tiltaket. Vi vil deretter vise hvordan statistisk analyse av data kan supplere visuelle analyser både for vurdering av individuell effekt og for samlet deltakereffekt (Lobo et al. 2017; Pustejovsky et al., 2014).

#### Visuell inspeksjon av data

Som nevnt har visuell inspeksjon av data to funksjoner. For det første å vurdere om innsamlede data har den nødvendige stabiliteten i baselinefasen slik at det kan foretas en nivåmessig sammenlikning mellom baseline og intervensjonsfase. For det andre kan visuell inspeksjon også indikere hvorvidt det har skjedd en endring som en følge av intervensjonen og sammen med loggføringer fra kartlegging og undervisning generelt bidra til tolkning av data.

#### **Stabilitet**

Det ble lagt vekt på å vurdere hvorvidt data for hver deltaker var stabile i baselinefasen. Utrekning av dette fulgte prosedyrene som beskrevet over (Lane & Gast, 2014). Resultatene viste at tre av elevene hadde stabile data i baselinefasen ved at 80 % av datapunktene var innenfor 25 prosent intervallet av medianverdien i fasen. For elev 7 var data i fasen ikke tilstrekkelig stabile, da to av datapunktene avvek med mer enn 25 % fra medianverdien.

### **Trender**

Siden intervensjonen i dette tilfellet handlet om å lære nye strategier i ordavkodning forventet vi ingen umiddelbar effekt ved oppstart av intervensjonen, men en eventuell gradvis progresjon. Kriteriet om umiddelbar endring i helningsgrad ble derfor vurdert irrelevant (Klingbeil et al., 2017). I trendanalysen ble det derfor primært undersøkt hvorvidt det var uønskede trender i baselinefasen siden dette kan svekke validiteten av vurderingene (Gast & Ledford, 2014). Som støtte i vurderingen ble trendlinjen lagt til ved hjelp av Microsoft Excel (figur 2). Stabilitet rundt trendlinjen ble så vurdert gjennom å regne ut korrelasjonen mellom tidsvariabelen og avhengig variabel ved hjelp av Excel. Denne metoden forutsetter som nevnt at data er lineære, noe som bare delvis er tilfelle i dette eksemplet. Inspeksjonen avdekket en tydelig, men ikke lineær trend i baselinefasen for elev 7, mens trendene for de øvrige var enten nøytrale (elev 2 & 3) eller svakt positiv (elev 10). Stabilitet rundt trendlinjen ble målt med  $R^2$  (figur 2). Her viser resultatene en moderat stabilitet for tre av elevene (2, 3 & 7) og en stabil lineær trend for elev 10.

For øvrig sees en viss endring i data assosiert med tidspunkt for oppstart av intervensjonen og for elev 3, 7 og 10 en svak nedgang i postfasen. For elev 7 sees også en markant nedgang etter seks ukers intervensjon, assosiert med avbrekk i treningen som følge av juleferie (figur 1). For elev 2 sees generelt en større variasjon i data i intervensjons og postfasen sammenliknet med baseline.

### **Nivåendringer**

Nivåendringer ble vurdert gjennom utregning av gjennomsnittsskåre for hver fase både når det gjaldt råskåre og når det gjaldt målet for i hvilken grad elevene utviklet seg positivt i retning av å tilegne seg en mer nøyaktig lesestrategi. Som vist i Figur 3 er det for tre av elevene en positiv endring i form av at prosentandelen av ord som blir lest riktig øker i intervensjonsfasen sammenliknet med baseline. Dette gjelder også for råskårene, om enn i mindre tydelig grad. For to av elevene (elev 7 & elev 10) opprettholdes endringen også i postfasen av studien, mens elev 2 viser en svak nedgang og elev 3 en tydelig reduksjon i postfasen. Resultatene for elev 3 må generelt sees i lys av at denne eleven opplevde store endringer i livssituasjon i løpet av intervensjonsperioden, noe som i henhold til informasjon fra lærer gjorde eleven mindre mottakelig for treningen.

### **Overlapp**

For hver elev ble det telt opp antall datapunkter som økte i henhold til den høyeste verdien i baselinefasen (figur 2). Prosentandelen av datapunkter i intervensjonsfasen som ikke overlapper (PND) er vist i figur 2.

### **Statistiske analyser**

For å vurdere hvorvidt den observerte endringen i intervensjonsfasen var signifikant og for å vurdere størrelsen av eventuell effekt for hver elev brukte vi to parametriske mål, SMD og LRR og et ikke parametrisk mål, Tau-U. For vurdering av samlet deltakereffekt brukte vi vektet Tau-U og BC-SMD. Alle utregninger av effekt ble basert på råskårer. Siden effekt av tiltaket innebar å lære og bruke et ukjent treningsmateriale beregnet vi, ut fra tidligere erfaring med treningsmaterialet, en treningsfase på to uker ved oppstart og inkluderte derfor kun data fra og med uke tre av intervensjonen. LRR og SMD ble regnet ut ved hjelp av online kalkulatoren (Pustejovsky & Swan 2018).

Vi valgte å konvertere effektmålet fra LRR til prosentvis økning og i utregningen av SMD ble det brukt samlet (pooled) standardavvik. Resultater er vist i figur 2.

Tau-U ble regnet ut med online kalkulatoren (Vannest, Parker, Gonen & Adiguzel, 2016). Vi vurderte først hvorvidt det var signifikante trender i baselinefasen. Til tross for at visuell inspeksjon viste at elev 7 hadde en tydelig oppadgående trend var denne trenden ikke signifikant på 95% nivå ( $p = 0,14$ ). Trenden for elev 10 var derimot signifikant og det ble derfor kontrollert for baseline trend i utregning av effektestimater (Parker et al., 2011) (figur 2).

Samlet effekt for deltakerne basert på vektet gjennomsnittlig effektstørrelse med Tau -U gir et effektestimater på 0,72 (Konfidensintervall ( 95%) = 0,40-1,0). Dette regnes i henhold til den foreslåtte standarden (se over) som stor effekt.

Samlet effekt ble også målt med BC-SMD. Siden RMLE estimatoren er regnet som mest fleksibel i vurdering av data i små og heterogene utvalg valgte vi denne (Valentine et al., 2016). Utregningsmetoden i kalkulatoren krever at det spesifiseres både gjennomsnittsnivå (fixed level) og individuell variasjon (random level) i baselinefasen samt gjennomsnittlig effekt ( fixed level ) i intervensjonsfasen. For videre spesifikasjoner fulgte vi kriteriene foreslått av Wolfe et al.(2019). I henhold til disse skal det inkluderes spesifisering av individuell variasjon i intervensjonsfasen ( random level) dersom den gjennomsnittlige intervensjonseffekten mellom deltakerne varierer med mer enn 10% av verdien på y akse. I dette tilfellet avviker elev 3 med mer enn 10% fra de andre. For å vurdere om det også skulle spesifiseres trender brukte vi visuell inspeksjon av data som utgangspunkt( Wolfe et al., ibid). Siden en av elevene ( elev 10) viste en tydelig trend i baselinefasen spesifiserte vi random trend i denne fasen. Trendene i intervensjonsfasen var uklare og delvis inkonsistente og det ble derfor valgt å ikke spesifisere trender i denne fasen.. Samlet effekt med BC-SMD viser  $d = 0,64$  med konfidensintervall -8.81- 10.11. Dette resultatet er ikke signifikant.

### Oppsummering av resultater

Kort oppsummert kan man si at de visuelle analysene angir hvorvidt individuelle data er egnet for å vurdere effekt og også gir et overblikk over fluktuasjoner i data som sammen med logg fra lærer kan avdekke om variasjonen er tilfeldig eller systematisk. De individuelle effektestimaterne kan brukes til å vurdere effektstørrelsen hos en gitt elev, mens det samlede effektmålet indikerer for eventuell videre utprøving på nye elever. Dersom effektvurdering av tiltaket skulle gjøres på bakgrunn av visuelle analyser alene ville trolig elev 7 bli tatt ut siden data i baselinefasen er ustabile. Til en viss grad kan imidlertid de statistiske analysene kompensere for ustabile data og gir generelt en mindre restriktiv vurdering enn visuelle analyser (Wolfe et al., 2019). De gjennomgående positive resultatene for elev 7 på tvers av de ulike effektmålene tydeliggjør dermed at statistiske analyser kan fange opp effekt som ikke framkommer visuelt (ibid.).

Effektmålene synes å være enige i henhold til vurdert effekt hos tre av elevene (elev 2, 7 og 10), mens for elev 3 er endringen marginal og ikke signifikant. Det samlede målet (BC SMD) viser også et ikke signifikant resultat med stort konfidensintervall. Det samlede resultatet med Tau - U viser imidlertid et signifikant resultat, noe som eksemplifiserer at vurderingen basert på overlapp i dette tilfellet gir en mindre konservativ vurdering enn effektmålet basert på flernivåanalyse.

Utregning av samlet effekt med BC-SMD kan, ifølge veilederen (Valentine et al. , 2016), i prinsippet gjøres med multiple baseline design med så få som 3-4 deltakere. Tidligere simuleringsstudier med flernivåanalyser i SCD har imidlertid indikert at det oppnås usikre

resultater med så få deltakere og at det er nødvendig med 6-8 deltakere for å minske feilmarginene (Ferron et al., 2009). Siden vi i dette eksemplet har valgt ut enkelt deltakere fra den originale studien primært for å vise kontraster i oppnådd effekt, gir utregning av samlet deltakereffekt med disse fire elevene et upresist mål med stort konfidensintervall og dermed ikke signifikante resultater. Vi har likevel valgt å eksemplifisere bruk av BC-SMD her fordi vi antar at dette effektmålet kan ha positive kvaliteter når det gjelder mulighet til å vurdere effekt av konkrete tiltak på tvers av deltakere (se nedenfor). I den originale studien (under fagfelle vurdering) eksemplifiseres bruk av BC-SMD med 11 deltakere basert på de samme kriteriene for utregning som vist her. Siden BC-SMD er et relativt nylig utviklet verktøy for SCD er både kriterier for bruk og retningslinjer når det gjelder spesifisering av modeller under utvikling. Vi har her valgt å bruke kriterier som har vist seg å ha stort grad av samsvar med visuelle analyser og individuelle effektmål ( Wolfe et.al., 2019), men også andre kriterier er tilgjengelige (se f.eks. WWC, (2020) for alternative innfallsvinkler).

## Diskusjon

Vi har i denne artikkelen drøftet og eksemplifisert hvordan SCD kan brukes som verktøy i evaluering av spesialpedagogiske tiltak. Vi har også diskutert utfordringer knyttet til validitet og presentert forskjellige innfallsvinkler for vurdering av effekt. Eksemplet viser hvordan det er mulig å designe en SCD studie innenfor rammen av spesialpedagogiske tiltak i skolen. I eksemplet brukes repeterte tester som utgangspunkt for vurdering av effekt. Også andre målemetoder er mulige for å vurdere utbytte av intervensjoner rettet mot å øke akademiske ferdigheter. I en nylig publisert studie som evaluerte effekten av et tiltak med sikte på å øke elevenes bruk av relevante planleggings strategier i tekstproduksjon («The STOP and LIST strategy»)(Grunke , Nobel & Bracht , 2019), ble det brukt en mer kvalitativ innfallsvinkel ved at utvalg av elevtekster ble analysert og kodet etter bestemte kriterier både i baselinefasen og intervensjonsfasen. Resultatene ble så sammenliknet og effekt vurdert utfra differansen mellom tekstene i de to fasene utfra de målte kriteriene.

Dersom tiltak rettes mot atferdsendring er design som bruker direkte observasjon en egnet innfallsvinkel. Da kan både reversible design, der tiltaket «skrus» av og på gjennom gjentatte faser, og multiple baseline design brukes. Ved multiple baseline design med bruk av atferds observasjon som målemetode kan for eksempel den uavhengige variabelen introduseres som tiltak først i et skolefag og deretter i de neste fagene suksessivt og effekt vurderes utfra hvorvidt det forekommer en systematisk økning eller reduksjon av observert atferd assosiert med tiltaket (Gast et al., 2014)).

Som vist i eksemplet kan SCD med relativt enkle grep tilpasses den situasjonen man arbeider i. For å undersøke om et tiltak virker for en elev i en konkret undervisningssituasjon vil det være tilstrekkelig å bruke et enkelt tofasedesign (AB) der man gjør et visst antall målinger (minst tre, men ideelt sett fem) før man starter tiltaket og så fortsetter målingene etter at tiltaket har startet. Resultatet vil da kunne si noe om tiltaket virker i akkurat denne situasjonen, men vil ikke ha gyldighet for andre fag, i andre klasserom, med andre lærere eller for andre elever. Likevel innebærer en slik metode en systematisk måling av atferd som gir et grunnlag for vurdering av om og hvordan tiltaket har virket på eleven og på om tiltaket bør fortsette eller erstattes med et annet tiltak (Gast & Ledford, 2014). Flere av online-kalkulatoren som er tilgjengelige, baserer også utregning av effekt på tofase design, slik at det selv med et enkelt AB-design er mulig å regne ut effektstørrelse for en elev, slik det er vist i eksemplet over. Selv om AB-designet ikke er et fullverdig forskningsdesign kan likevel kvantifisering av effekt og bruk av veiledende normer for vurdering av effektstørrelse være nyttig i en klinisk sammenheng og supplere læreres subjektive vurderinger av hva som virker

for eleven. Ikke minst vil bruk av kvantitative effektmål være egnet til å sammenlikne effekt av ulike tiltak på samme elev.

Effektmålinger i SCD har også den fordel at de relateres til individuell utvikling ut fra elevens eget utgangspunkt på et definert utviklingsfelt. Denne fleksibiliteten er særlig relevant i spesialpedagogisk virksomhet der man ofte arbeider med konkrete delferdigheter innenfor et fagområde. Fagbaserte måleverktøy som er standardisert på klassetrinnet, for eksempel kartleggingsprøver for lesing og matematikk, vil på sin side sjelden være detaljerte nok til å fange opp endringer hos en elev som følge av et avgrenset spesialpedagogisk tiltak.

Individuelle effektmål av denne typen er også egnet som grunnlag for årlige rapporteringer og også for samarbeid med elev og foresatte, samt eventuelle andre aktuelle faginstanser rundt eleven. For PPT vil det å motivere for bruk av SCD i det spesialpedagogiske arbeidet på skolene eller i barnehagene også danne utgangspunkt for å kunne forske i egen praksis. Ved å sikre at kravene til eksperimentell kontroll følges for eksempel ved å følge retningslinjene presentert i konsensusrapporten fra SCRIBE (Tate et al 2016), og ved å replisere resultater på flere elever vil PPT også kunne bidra til utvikling av forskningsbasert kunnskap. Dette vil både være relevant som grunnlag for videre rådgivning på feltet og samtidig kunne bidra til å knytte praksisfeltet tettere til forskningsmiljøene. Statistiske analyseverktøy som gjør det mulig å vurdere samlet effekt basert på replikasjoner vil her trolig kunne være nyttige verktøy (se f.eks. Pustejovsky, 2016).

Erfaring viser også at småskalaforskning av denne typen kan ha positive ringvirkninger i miljøet der de settes i gang og føre til økt faglig engasjement i virksomheten (Dexter & Seden, 2012). I en norsk kontekst vil bruk av SCD i evaluering av spesialpedagogiske tiltak trolig kunne bidra både til økt engasjement i planlegging og gjennomføring av enkeltvedtak og indirekte også fungere som et styringsverktøy for økt kvalitet i alle ledd i tiltakskjeden.

### Begrensninger

Vi har i denne artikkelen beskrevet grunnleggende trekk ved SCD-metodologien, og diskutert validitet knyttet til design, effektmål og analysemetoder. En viktig begrensning ved denne gjennomgangen er selvsagt at den kun omhandler et utvalg metoder og problemstillinger. Utvalget gjengir heller ikke den historiske utviklingen på området. Metodeutviklingen innen SCD er i stadig endring og gamle metoder nyanseres og videreutvikles parallelt med at tradisjonelle metoder relanseres og metodiske nyvinninger testes ut og diskuteres (Shadish, 2014). Når det gjelder analysemetodenes validitet er det noen verktøy som i litteraturen er hyppigere diskutert enn andre, og det preger også denne framstillingen. Det kan gi et skjevt bilde av at metoder som er oftest omtalt og dermed oftest kritisert, er de som også har svakest validitet og kvalitet. Dette er ikke nødvendigvis tilfelle.

## Referanser

Baer, D. M., Wolf, M. M. & Risley, T. R. (1968). SOME CURRENT DIMENSIONS OF APPLIED BEHAVIOR ANALYSIS 1. *Journal of Applied Behavior Analysis*, 1 (1) , 91-97. doi:10.1901/jaba.1968.1-91

Barneombudet (2017): Uten mål og mening. [http://barneombudet.no/wp-content/uploads/2017/03/Bo\\_rapport\\_enkeltsider.pdf](http://barneombudet.no/wp-content/uploads/2017/03/Bo_rapport_enkeltsider.pdf)

Busk, P. L. and R. C. Serlin (1992). Meta-analysis for single-case research. Single-case research design and analysis: New directions for psychology and education. Hillsdale, NJ, US, Lawrence Erlbaum Associates, Inc: 187-212.

Busse, R., McGill, R., & Kennedy, K. (2015). Methods for Assessing Single-Case School Based Intervention Outcomes. *The Official Journal of the California Association of School Psychologists*, 19(3), 136-144. doi:10.1007/s40688-014-0025-7

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed. ed.). Hillsdale, N. J: Laurence Erlbaum.

de Bruin, C. L. (2017). Conceptualizing effectiveness in disability research. *International Journal of Research & Method in Education*, 40 (2) , 113-136. doi:10.1080/1743727X.2015.1033391

Dexter, B., & Seden, R. (2012). 'It's really making a difference': how small-scale research projects can enhance teaching and learning. *Innovations in Education and Teaching International*, 49(1), 83-93. doi:10.1080/14703297.2012.647786

Ferron, J.M. Bell, B.A, Hess, M.R. Rendina-Gobioff, G. & Hibbard, S.T. Making. treatment effect inferences from multiple-baseline data: The utility of multilevel modeling approaches. *Behavior Research Methods* 41, 372–384 (2009). <https://doi.org/10.3758/BRM.41.2.372>

Gast, D. & Baekey, D.H., 2014. Withdrawal and Reversal Designs. I Gast, D. L., & Ledford, J. R. *Single case research methodology : applications in special education and behavioral sciences* (2nd ed. ed.). New York: Routledge, p. 211-251.

Gast, D. & Spriggs, A.M., 2014. Visual Analysis of Graphic Data. I Gast, D. L., & Ledford, J. R. *Single case research methodology : applications in special education and behavioral sciences* (2nd ed. ed.). New York: Routledge, p. 176-211.

Gast, D. L., & Ledford, J. R. (2014). *Single case research methodology : applications in special education and behavioral sciences* (2nd ed. ed.). New York: Routledge.

Gast, D., 2014. General Factors in Measurement and Evaluation. I Gast, D. L., & Ledford, J. R. *Single case research methodology : applications in special education and behavioral sciences* (2nd ed. ed.). New York: Routledge, p. 85-105.

Gast, D., Lloyd, B.P. & Ledford, J. (2014). Multiple Baseline and Multiple Probe Designs. I: Gast, D. L., & Ledford, J. R. *Single case research methodology : applications in special education and behavioral sciences* (2nd ed. ed.). New York: Routledge, p. 251-297.

Grünke, Matthias & Nobel, Kerstin & Bracht, Janine. (2019). Effects of the STOP and LIST



Strategy on the Writing Performance of Struggling Fourth Graders. 8. 11310.5539/jel.v8n2p1.

Harrington, M., & Velicer, W. F. (2015). Comparing Visual and Statistical Analysis in Single-Case Studies Using Published Studies. *Multivariate Behavioral Research, 50* (2), 162-183. doi:10.1080/00273171.2014.973989

Haug, P. (2019). *Spesialundervisning : innhald og funksjon* (1. utgåva. ed.) Samlaget, Oslo.

Hedges, L. V. (2008). What Are Effect Sizes and Why Do We Need Them? *Child Development Perspectives, 2* (3), 167-171. doi:10.1111/j.1750-8606.2008.00060.x

Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2012). A standardized mean difference effect size for single case designs. *Research Synthesis Methods, 3* (3), 224-239. doi:10.1002/jrsm.1052

Horner, R. D., & Baer, D. M. (1978). MULTIPLE-PROBE TECHNIQUE: A VARIATION OF THE MULTIPLE BASELINE 1. *Journal of Applied Behavior Analysis, 11* (1), 189-196. doi:10.1901/jaba.1978.11-189

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The Use of Single-Subject Research to Identify Evidence-Based Practice in Special Education. *Exceptional Children, 71* (2), 165-179. doi:10.1177/001440290507100203

Kazdin, A. E., & American Psychological, A. (2016). *Methodological issues & strategies in clinical research* (Fourth edition. ed.). Washington, DC: American Psychological Association.

Kazdin, A. E., & Kazdin, A. E. (2011). Evidence-based treatment research: Advances, limitations, and next steps. *The American psychologist, 66*(8), 685-698. doi:10.1037/a0024975

Klingbeil, D., Norman, E., & Nelson, P. (2017). Precision of Curriculum-Based Measurement Reading Data: Considerations for Multiple-Baseline Designs. *Journal of Behavioral Education, 26* (4), 433-451. doi:10.1007/s10864-017-9282-7

Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the Scientific Credibility of Single-Case Intervention Research: Randomization to the Rescue. *Psychological Methods, 15* (2), 124-144. doi:10.1037/a0017736

Krasny-Pacini, A., & Evans, J. (2018). Single-case experimental designs to assess intervention effectiveness in rehabilitation: A practical guide. *Annals of Physical and Rehabilitation Medicine, 61*(3), 164-179. Retrieved from <http://www.sciencedirect.com/science/article/pii/S1877065717304542>. doi:<https://doi.org/10.1016/j.rehab.2017.12.002>

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-Case Intervention Research Design Standards. *Remedial and Special Education, 34* (1), 26-38. doi:10.1177/0741932512452794

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M. & Shadish, W. R. (2010). Single-case designs technical documentation. Retrieved from What Works Clearinghouse website: [http://ies.ed.gov/ncee/wwc/pdf/wwc\\_scd.pdf](http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf).

- Kvande, M. N., Bjørklund, O., Lydersen, S., Belsky, J., & Wichstrøm, L. (2019). Effects of special education on academic achievement and task motivation: a propensity-score and fixed-effects approach. *European Journal of Special Needs Education, 34* (4) , 409-423. doi:10.1080/08856257.2018.1533095
- Lane, J. D., & Gast, D. L. (2014). Visual analysis in single case experimental design studies: Brief review and guidelines. *Neuropsychological Rehabilitation, 24* (3-4) , 445-463. doi:10.1080/09602011.2013.815636
- Ledford, J. R. (2018). No Randomization? No Problem: Experimental Control and Random Assignment in Single Case Research. *American Journal of Evaluation, 39* (1) , 71-90. doi:10.1177/1098214017723110
- Ledford, J.R., Wolery, M. & Gast, D.I. 2014. Controversial and Critical Issues in Single Case Research. I Gast, D. L., & Ledford, J. R. *Single case research methodology : applications in special education and behavioral sciences* (2nd ed. ed.). New York: Routledge, p. 377-397.
- Lekhal, R. (2017). Elever med vedtak om spesialundervisning : hva vet vi, hvordan har de det, og trives de på skolen? I Haug, P. (2019). *Spesialundervisning : innhold og funksjon* (1. utgåva. ed.) (pp. 368-385). Oslo: Samlaget.
- Lobo, A. M., Moeyaert, A. M., Baraldi Cunha, A. A., & Babik, A. I. (2017). Single-Case Design, Analysis, and Quality Assessment for Intervention Research. *Journal of Neurologic Physical Therapy, 41* (3) , 187-197. doi:10.1097/NPT.0000000000000187
- Ma, H.- H. (2006). An Alternative Method for Quantitative Synthesis of Single-Subject Researches: Percentage of Data Points Exceeding the Median. *Behavior Modification, 30* (5) , 598-617. doi:10.1177/0145445504272974
- Maggin, D. M., Cook, B. G., & Cook, L. (2018). Using Single-Case Research Designs to Examine the Effects of Interventions in Special Education. *Learning Disabilities Research & Practice, 33* (4) , 182-191. doi:10.1111/ldrp.12184
- Maggin, D. M., Lane, K. L. & Pustejovsky, J. E. (2017). Introduction to the Special Issue on Single-Case Systematic Reviews and Meta-Analyses. *Remedial and Special Education, 38*(6), s. 323-330. doi:10.1177/0741932517717043
- Maggin, D. M., Swaminathan, H., Rogers, H. J., Amp, Apos, Keeffe, B. V., . . . Horner, R. H. (2011). A generalized least squares regression approach for computing effect sizes in single-case research: Application examples. *Journal of School Psychology, 49*(3), 301-321. doi:10.1016/j.jsp.2011.03.004
- Maggin, D. M., Swaminathan, H., Rogers, H. J., O'Keeffe, B. V., Sugai, G., & Horner, R. H. (2011). A generalized least squares regression approach for computing effect sizes in single-case research: Application examples. *Journal of School Psychology, 49*(3), 301-321. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0022440511000203>. doi:https://doi.org/10.1016/j.jsp.2011.03.004
- Manolov, R., & Vannest, K. J. A.(2019). Visual Aid and Objective Rule Encompassing the

Data Features of Visual Analysis. *Behavior Modification*, 0(0), 0145445519854323. Retrieved from <https://journals.sagepub.com/doi/abs/10.1177/0145445519854323>. doi:10.1177/0145445519854323

Manolov, R., Solanas, A., & Leiva, D. (2010). Comparing “Visual” Effect Size Indices for Single-Case Designs. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 6(2), 49-58. doi:10.1027/1614-2241/a000006

Manolov, R., Solanas, A., & Sierra, V. (2018). Extrapolating baseline trend in single-case data: Problems and tentative solutions. *Behavior Research Methods*, doi:10.3758/s13428-018-1165-x

Nordahl, T. et al., (2018). Inkluderende felleskap for barn og unge. Fagbokforlaget. <https://nettsteder.regjeringen.no/inkluderende-barn-unge/files/2018/04/INKLUDERENDE-FELLESSKAP-FOR-BARN-OG-UNGE-til-publisering-04.04.18.pdf>

NOU (2019:3). Nye sjanser - bedre læring - Kjønnforskjeller i skoleprestasjoner og utdanningsløp. Regjeringen/Kunnskapsdepartementet. <https://www.regjeringen.no/no/dokumenter/nou-2019-3/id2627718/sec1>

Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. H., Thompson, B., & Harris, K. R. (2005). Research in Special Education: Scientific Methods and Evidence-Based Practices. *Exceptional Children*, 71 (2) , 137-148. doi:10.1177/001440290507100201

Olive, M. L., & Smith, B. W. (2005). Effect Size Calculations and Single Subject Designs. *Educational Psychology*, 25, 313-313) , p.313-324.

Onghena, P., Michiels, B., Jamshidi, L., Moeyaert, M., & Van Den Noortgate, W. (2018). One by One: Accumulating Evidence by using Meta-Analytical Procedures for Single-Case Experiments. *Brain Impair.*, 19 (1) , 33-58. doi:10.1017/BrImp.2017.25

Opsvik, F., & Haug, P. (2017). Læringsutbyttet i matematikk. I Haug, P. (2019). *Spesialundervisning : innhald og funksjon* (1. utgåva. ed.). p. 324-349). Oslo: Samlaget.

Parker, R. I., Vannest, K. J., & Brown, L. (2009). The Improvement Rate Difference for Single-Case Research. *Exceptional Children*, 75 (2) , 135-150. doi:10.1177/001440290907500201

Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect Size in Single-Case Research: A Review of Nine Nonoverlap Techniques. In (Vol. 35, pp. 303-322). Los Angeles, CA.

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining Nonoverlap and Trend for Single-Case Research: Tau-U. *Behavior Therapy*, 42 (2) , 284-299. doi:10.1016/j.beth.2010.08.006

Pennington, B. F., & Bishop, D. V. M. (2009). Relations among speech, language, and reading disorders. *Annual Review of Psychology*, 60, 283.

Pustejovsky, James E. (2016). scdhlms: A web-based calculator for between-case standardized mean differences (Version 0.3.1) [Web application]. Retrieved from: <https://jepusto.shinyapps.io/scdhlms>

Pustejovsky, J. (2018 a). Using response ratios for meta-analyzing single-case designs with behavioral outcomes. *J. Sch. Psychol.*, 68, 99-112. doi:10.1016/j.jsp.2018.02.003

Pustejovsky, J. E. (2018 b). Procedural Sensitivities of Effect Sizes for Single-Case Designs with Directly Observed Behavioral Outcome Measures. *Grantee Submission*.

Pustejovsky, J. E., & Ferron, J. M. (2017). Research synthesis and meta-analysis of single-case designs. In J. M. Kaufmann, D. P. Hallahan, & P. C. Pullen (Eds.), *Handbook of Special Education*, 2nd Edition. New York, NY: Routledge.

Pustejovsky, J. E., Hedges, L. V., & Shadish, W. R. (2014). Design-Comparable effect sizes in multiple baseline designs: A general modeling framework. *Journal of Educational and Behavioral Statistics*, 39 (5) , 368-393. doi:10.3102/1076998614547577

Pustejovsky, J. E. & Swan, D. M. (2018). Single-case effect size calculator (Version 0.5) Web application. Retrieved from <https://jepusto.shinyapps.io/SCD-effect-sizes/>

Pustejovsky, J. E., Swan, D. M., English, K. W., & Pustejovsky, J. E. (2019). An Examination of Measurement Procedures and Characteristics of Baseline Outcome Data in Single-Case Research. *Behavior Modification*, 145445519864264-145445519864264. doi:10.1177/0145445519864264

Rakap, S. (2015), Effect sizes as result interpretation aids in single-subject experimental research: description and application of four nonoverlap methods. *British Journal of Special Education*, 42: 11-33. doi:[10.1111/1467-8578.12091](https://doi.org/10.1111/1467-8578.12091)

Regjeringen 2016-2017: Lærelyst – tidlig innsats og kvalitet i skolen. <https://www.regjeringen.no/no/dokumenter/meld.-st.-2120162017/id2544344/>

Regjeringen 2019-20 : Tett på – tidlig innsats og inkluderende fellesskap i barnehage, skole og SFO <https://www.regjeringen.no/no/dokumenter/meld.-st.-6-20192020/id2677025/>

Scruggs, T. E., & Mastropieri, M. A. (1998). Summarizing Single-Subject Research: Issues and Applications. In (pp. 221-242).

Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The Quantitative Synthesis of Single-Subject Research: Methodology and Validation. *Remedial and Special Education*, 8(2), 24-33. doi:10.1177/074193258700800206

Shadish, W. R. (2014). Statistical Analyses of Single-Case Designs: The Shape of Things to Come. *Current Directions in Psychological Science*, 23 (2) , 139-146. doi:10.1177/0963721414524773

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Shadish, W. R., Hedges, L. V., & Pustejovsky, J. E. (2013). Analysis and meta-analysis of single-case designs with a standardized mean difference statistic: A primer and applications. *Journal of School Psychology*. doi:10.1016/j.jsp.2013.11.005

Shadish, W. R., Hedges, L. V., Horner, R. H., Odom, S. L., National Center for Education, R., Westat, I., & National Center for Special Education, R. (2015). The Role of Between-Case Effect Size in Conducting, Interpreting, and Summarizing Single-Case Research. NCER 2015-002. In: National Center for Education Research.

Shadish, W. & Sullivan, K. (2011). Characteristics of single-case designs used to assess intervention effects in 2008. *Behavior Research Methods*, 43 (4) , 971-980. doi:10.3758/s13428-011-0111-y

Shamseer, L., Sampson, M., Bukutu, C., Schmid, C. H., Nikles, J., Tate, R., . . . Vohra, S. (2015). CONSORT extension for reporting N-of-1 trials (CENT) 2015: Explanation and elaboration. In: British Medical Journal Publishing Group.

Skorpen, L.B. (2017). Elevar med matematikkvanskar og deira utvikling i løpet av eit år. I Haug, P. (2019). *Spesialundervisning : innhald og funksjon* (1. utgåva. ed.) (p. 296-324). Oslo, Samlaget.

Swaminathan, H., Rogers, H. J., Horner, R. H., Sugai, G., & Smolkowski, K. (2014). Regression models and effect size measures for single case designs. *Neuropsychological Rehabilitation*, 1-18. doi:10.1080/09602011.2014.887586

Tarlow, K. R. (2016). Baseline Corrected Tau-U Calculator. <http://www.ktarlow.com/stats/tau/>

Tarlow, K. R. (2017). An improved rank correlation effect size statistic for single-case designs: Baseline Corrected Tau. *Behavior Modification*, 41(4), 427-467. <http://dx.doi.org/10.1177/0145445516676750>

Tate, R. L., Perdices, M., Rosenkoetter, U., Shadish, W., Vohra, S., Barlow, D. H., . . . Wilson, B. (2016). The Single-Case Reporting guideline In BEhavioural interventions (SCRIBE) 2016 statement. *Journal of School Psychology*, 56, 133-142. doi:10.1016/j.jsp.2016.04.001

Valentine, J. C., Tanner-Smith, E. E., Pustejovsky, J. E., & Lau, T. S. (2016). Between-case standardized mean difference effect sizes for single-case designs: a primer and tutorial using the scdhlms web application. *Campbell Systematic Reviews*, 12 (1) , 1-31. doi:10.4073/cmdp.2016.1

Vannest, K. J., & Ninci, J. (2015). Evaluating Intervention Effects in Single-Case Research Designs. *Journal of Counseling & Development*, 93(4), 403-411. doi:10.1002/jcad.12038

Vannest, K.J., Parker, R.I., Gonen, O., & Adiguzel, T. (2016). Single Case Research: web based calculators for SCR analysis. (Version 2.0) [Web-based application]. College Station, TX: Texas A&M University.. Available from [singlecaseresearch.org](http://singlecaseresearch.org)

What Works Clearinghouse (2017) Procedures Handbook Version 4.

[https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc\\_procedures\\_handbook\\_v4.pdf](https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_handbook_v4.pdf)

What Works Clearinghouse (2020) Current standards version 4.1. – summary of changes.

<https://ies.ed.gov/ncee/wwc/Handbooks>

Zimmerman, K. N., Ledford, J. R., Severini, K. E., Pustejovsky, J. E., Barton, E. E., & Lloyd, B. P. (2018 a). Single-case synthesis tools I: Comparing tools to evaluate SCD quality and rigor. *Research in Developmental Disabilities, 79*, 19-32. Retrieved from

<http://www.sciencedirect.com/science/article/pii/S0891422218300301>.

doi:<https://doi.org/10.1016/j.ridd.2018.02.003>

Zimmerman, K. N., Pustejovsky, J. E., Ledford, J. R., Barton, E. E., Severini, K. E., & Lloyd, B. P. (2018 b). Single-case synthesis tools II: Comparing quantitative outcome measures. *Research in Developmental Disabilities, 79*, 65-76.

doi:<https://doi.org/10.1016/j.ridd.2018.02.001>



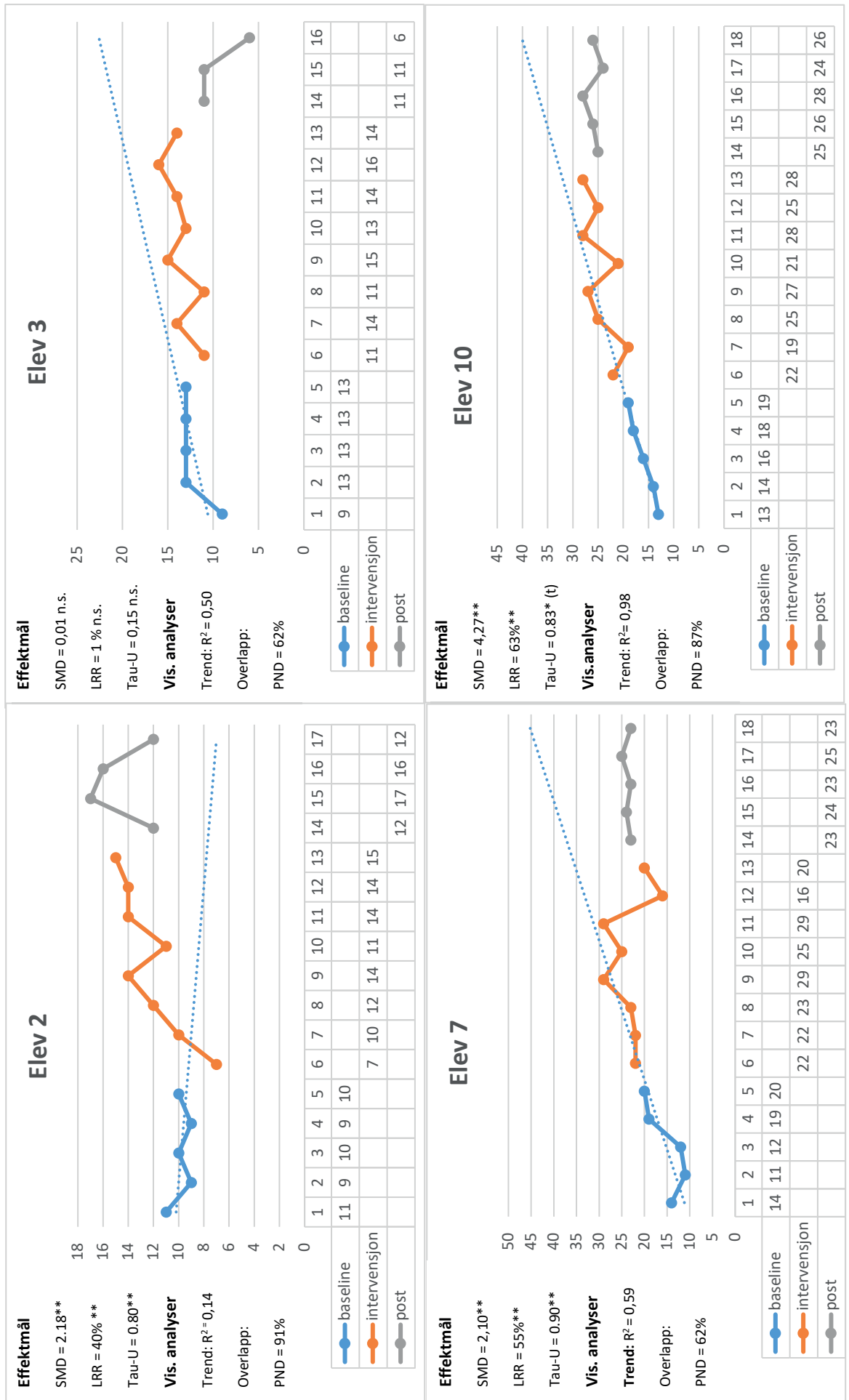
**Figur 1 Kartleggingsplan, Multiple baseline/ probe design for fire elever med totalt 18 målinger pr elev**

Elev/ Uke	39	40	41	42	43	44	45	46	47	48	49	50	51	52	1	2	3	4	5	6	7	8	9	1	11	12	13	14	15	16	17	18		
2	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
3	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
7	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
10	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Merk: Uke 40 (41), 51, 52 og 8 skolefri på grunn av ferie. Baselinefase markert i grått og post fasen markert i blått.



**Figur 2** Visuelt framstilling av resultater i tre faser (ett minutt non-ord lesing) for fire elever



**Merk:** Blå linje: Baselinefase, Rød linje: Intervensjonsfase, Grå linje: Postfase.

Grafikken gir et inntrykk av sammenhengen av datapunkter i de tre fasene uten at opphold mellom målingene er visualisert (Se figur 1 for korrekt oversikt over tidsplan for kartlegging for hver elev).

Stiplet linje: Lineær representasjon av trendlinjen basert på data i baselinefasen.

Alle effektmål: Utregningene er basert på samlet utvikling i intervensjonsfasen fra og med uke tre samt postfasen .

SMD = Standardized Mean Difference

LRR = Log Response Ratio – Increased

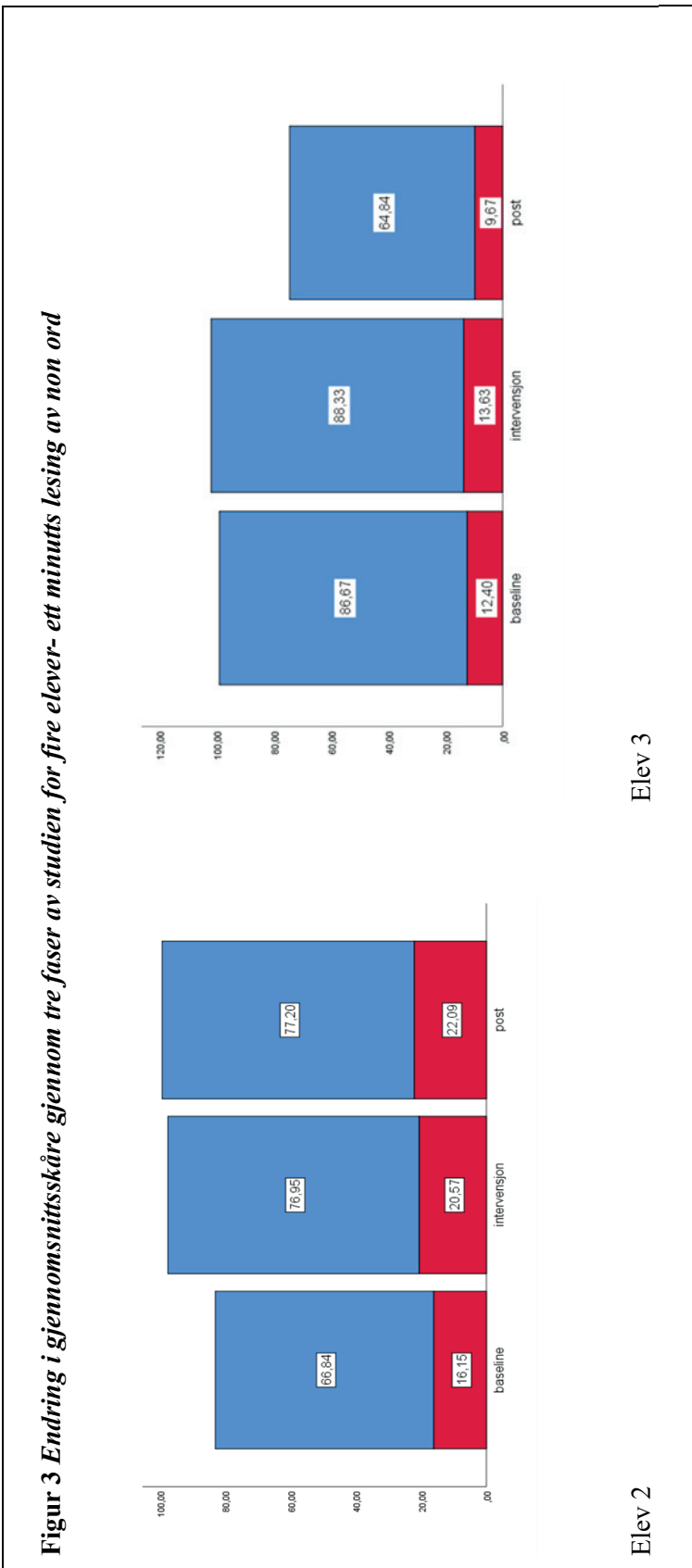
Utregninger av SMD og LRR: <https://jepusto.shinyapps.io/SCD-effect-sizes/>

Utregninger av Tau-U : <http://www.singlecaseresearch.org/calculators>.

Utregning av BC-SMD : <https://jepusto.shinyapps.io/scdhlm>

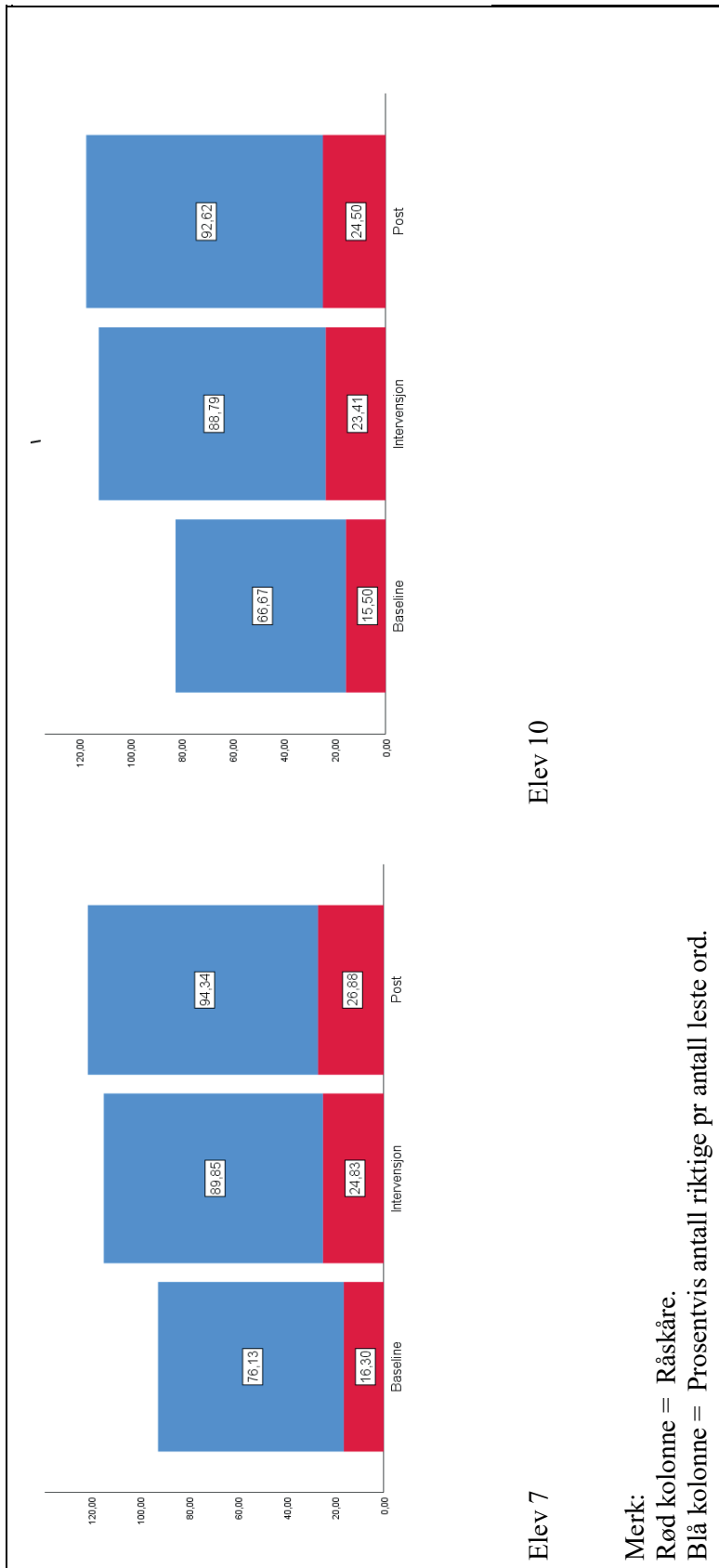
\*\* =  $p < 0,05$ , n.s. =  $p > 0,05$

(t) = Kontrollert for baseline trend



Elev 3

Elev 2



## **Artikkel 3**

Thurmann-Moe, A.C., Melby-Lervåg, M & Lervåg, A. The Impact of Articulatory Consciousness Training on Reading and Spelling Literacy in Students with Severe Dyslexia: An Experimental Single Case Study.



**The Impact of Articulatory Consciousness Training on Reading and Spelling Literacy in Students with Severe Dyslexia: An Experimental Single Case Study.**

### **Abstract**

This study evaluates the effect of an intervention whose aim is to make articulatory consciousness a tool in decoding and spelling. The sample comprises 11 students with severe dyslexia (2 *SD* below the mean pseudoword scores), and the intervention programme consists of 32 individual sessions over eight weeks. The study applies a multiple baseline/probe design with five baseline tests that correspond to a control condition, eight tests during the intervention and five post-intervention tests. On average, the results show significant improvement in all reading and spelling outcomes. However, there were also significant effects on an irrelevant control task (the pegboard test), perhaps indicating testing effects on the dependent variables, making it difficult to draw firm conclusions from the study. Consequently, testing the intervention in randomised trials of children with severe dyslexia is recommended so as to draw conclusions about its efficacy for this group.



## **The Impact of Articulatory Consciousness Training on Reading and Spelling Literacy in Students with Severe Dyslexia: An Experimental Single Case Study.**

Dyslexia is defined as a specific disorder of reading and spelling, which is primarily caused by a deficit in the phonological system (Lyon et al., 2003). The aetiology of this deficit is not fully known, but both genetic factors and environment play a role (Byrne et al., 2006; Christopher et al., 2013; Elwér et al., 2013; Olson et al., 2011). Thus, there is a large heterogeneity in children with dyslexia in terms of severity and the impact that the reading disorder has on reading, spelling and academic performance in general (Snowling & Melby-Lervåg, 2016). The heterogeneity within dyslexic samples also actualises the need for fine-grained measurement tools in the process of diagnosis and a wider spectrum of tailored remediation programmes. In this multiple probe study of students with severe dyslexia, we evaluate the effect of an instructional programme whose aim is to teach students how to use knowledge about articulatory and acoustic features of speech sounds as a tool in spelling and decoding.

### **The Phonological Deficit Hypothesis versus the Multifactorial Model**

According to the phonological deficit hypothesis, reading and spelling problems in dyslexia are caused by underspecified phonological representations (Swan & Goswami, 1997). Precursors of the weak phonological representations have often been explained with deficits in the auditory sensory system (Hämäläinen et al., 2013) or from deviant auditory perception (McBride-Chang, 1996; Noordenbos & Serniclaes, 2015).

However, more recently, the phonological deficit hypothesis has been criticised as overly simplistic, with the claim that dyslexia is not likely to be due to a single cause. This has been supported by studies showing that not all children with severe reading problems have a phonological deficit (Ramus et al., 2013).

As phonological deficit theory has not been able to accommodate important findings in children with dyslexia, researchers have suggested a multifactorial model for explaining reading disorders (Pennington et al., 2012). Recent longitudinal studies have also demonstrated that dyslexia does not stem exclusively from phonological deficits and have focused on a wider spectrum of sensory, cognitive and environmental factors (Eklund et al., 2015; Snowling et al., 2003).

### **Speech Production Deficits in Dyslexia**

In addition, it has been suggested that that speech perception is causally related to dyslexia (Hulme & Snowling, 2009), and that the quality of phonological representations can be understood in the context of speech production skills (Elbro et al., 1998; Snowling et al., 1992). Elbro et al. (1998) found that the distinctness of phonological representations, measured by a task where preschool children corrected a hand-held toy which incorrectly pronounced target words, correlated with later decoding skills. In a case study on a boy with speech sound disorders, Snowling et al. (1992) reported that as time went on, the boy's speech sound problems were resolved, but the same error typology remained as a spelling problem.

Correlations between speech production deficits and reading impairment are frequently reported in samples with speech sound disorders (Lewis et al., 2011; Preston et al., 2013) and are also present in dyslexic samples (Lalain et al., 2003; Sénéchal et al., 2004). Studies also show that individuals with dyslexia perform more poorly than controls both in articulatory awareness (Griffith & Frith, 2002; Montgomery, 1981) and articulatory speed (Duranovic & Sehic, 2013; Fawcett & Nicolson, 2002). However, studies employing a longitudinal perspective have shown that speech production skills do not seem to have unique explanatory value regarding the development of reading disorders (Hulme et al., 2015). Rather, it appears that speech production and speech perception are closely linked (Hulme & Snowling, 2009). This is perhaps most clearly formulated in the motor theory of speech

perception (Liberman & Mattingly, 1985; Liberman & Whalen, 2000). According to this theory, speech perception and speech production cannot be separated; they are parts of the same process, where the perception of sounds is synchronised with the observation of associated articulatory gestures.

### **Articulatory Consciousness Training to Ameliorate Dyslexia**

Based on the theory that dyslexia can, at least partly, be caused by speech production problems, the next step was to try to train students in features related to speech production to ascertain whether this can enhance decoding and spelling. An instructional focus on articulation is also embedded in most basic reading programmes through read-aloud exercises. Moreover, teachers working in first-grade classrooms often report that novice readers spontaneously use loud or silent articulation as a form of support when working with segmentation and spelling tasks. This indicates that articulation may function as a ‘natural’ tool in reading and spelling (Skjelfjord, 1987). According to the self-teaching hypothesis, silent articulation may also play a role in the spontaneous process of phonological recoding (print-to-sound translation) that occurs during text reading (Share, 1999).

Interventions using articulatory training have mainly been conducted on typically developing preschool children and novice typical readers (Boyer & Ehri, 2011; Castiglioni-Spalten & Ehri, 2003; Fälth et al., 2017; Torgesen et al., 2001). In reading-delayed samples, the most frequently studied articulatory intervention programme was the Auditory Discrimination in Depth (ADD) and a later version of the same programme called LIPs (Lindamood Phonemic Sequences) (Lindamood & Lindamood, 1998; What Works Clearinghouse (WWC) 2008; 2015). This programme applies a multisensory approach to reading and aims to teach children to identify the mouth movements involved in the production of speech sounds. Several studies have compared the effects of this programme with other phonic-based programmes. Some studies of the ADD/LIPs programmes have failed

to demonstrate clear advantages of articulatory training over more traditional phonic approaches in improving phonological awareness and basic decoding skills, concluding that the two approaches provide similar gains (Torgesen et al., 2010; Wise et al., 1999). However, other studies have reported significant advantages of articulatory training compared to traditional phonics instruction in samples with reading problems (Joly-Pottuz et al., 2008; Trainin et al., 2014), particularly for those with the most severe reading problems (Fälth et al., 2017; Trainin et al., 2014).

In clinical contexts has a system based on pictographic symbols of both articulatory and accoustic features of the speech sounds, Pictographic Articulatory System (PAS) (Kausrud, 2003) been used on children with language disorders. The results of a case study on an eight-year-old boy with developmental language disorder suggest that a combined intervention, using both Bliss and PAS symbols, played a compensatory and mediating role by ameliorating the child's ability to read (Ottem & Kausrud, 2001). In a recent randomised controlled trial, Authors (in press) examined the effects of the PAS material in a five-week intervention aimed at improving phonological awareness and basic decoding skills in a sample of reading-delayed (approximately below the 20 percentile) first graders ( N = 129). For this group of delayed children, no effect was found beyond a 'business as usual' control group using ordinary phonics and phonological awareness training. However, as most of these children were beginning readers with a delay, and not dyslexic readers, we could not rule out that this kind of intervention would not work on children and young people with more severe dyslexic problems.

### **The current study**

Based on prior research indicating that articulatory consciousness training would produce benefits particularly in those with severe reading disorders, the current study

examines the effects of an intervention in a sample of children with persistent dyslexia. The intervention material is predominantly the same as in our previous study. Since children with persistent and severe dyslexia are a hard-to-reach population, we investigate the intervention by using a single-case experimental design (SCED) (Tate et al., 2016). The hypothesis underlying the intervention is that increased articulatory consciousness makes the discrimination of speech sounds easier for dyslexic students and, therefore, makes the representations of phonemes in the memory more distinct, producing a sustained training effect.

The research questions for the study are as follows:

Will articulatory consciousness training improve the students' reading efficiency and reading accuracy concerning regular words, pseudowords and irregular words?

Will articulatory consciousness training improve the students' spelling efficiency?

## **Method**

### **Sample**

The sample was recruited from the Regional Department for Speech and Language Disorders at The National Service for Special Needs Education. The criteria for participating in the study were scores below 2 *SD* on two standardised pseudo-reading subtests (STAS, Klinkenberg & Skaar 2001), i.e. below the second percentile; the participants' first language should be Norwegian; and they should be aged 10 years or older. Students with more complex diagnoses were excluded. For the flow of participants through the study, see Figure 1.

Additional sample characteristics were collected on a standardised pseudoword spelling test, also from the STAS battery, and on selected subtests from the Norwegian

standardised version of the Wisc-IV (Wechsler, 2014). The students' parents also answered a questionnaire concerning their children's early development, family risk factors and socio-economic status. For further details concerning the sample descriptives, see Table 1.

For all the participants, the reason for the referral to the Regional Department for Speech and Language Disorders was the need for new directions in reading instruction. Concerning individual dyslexic profiles, Cases 2, 3 and 12 were described in the school reports as 'non-readers'. The remainder of the sample had slightly better reading skills, although not fluent readers. Rather than using decoding, most of the children recorded a high presence of different sorts of guessing strategies, typically displayed by decoding from random phonological cues instead of decoding each letter.

Most of the participants were students in public schools, but Cases 7 and 11 were enrolled in classes based on alternative pedagogy due to school refusal problems. Case 6 attended a private sports academy. Case 3 experienced a relocation of residence and was placed in a new family during the intervention period.

The study followed the ethical guidelines of the national ethics committee.

## **Intervention**

The intervention programme aims to teach students a supplementary path to develop decoding and spelling skills by using pictographic symbols from the acoustic and articulatory features of each speech sound. The Pictographic cards are based on singular vowels and consonants in the Norwegian alphabet. Consonant cards consist of indicators for voice, placement of tongue and acoustic cues. Vowel cards symbolise the shape and opening of the mouth when pronouncing a vowel. Figure 2 depicts how the word 'ROSE' is spelled in PAS.

The intervention programme was constructed for the purpose of this study and consists of five learning activities. The content of activities was phonological awareness

training and basic decoding and spelling exercises, also including matching games. In the activities the articulatory cards were used as means in the instruction (See supplemental material for details of the content of the activities). The programme was introduced to the students as ‘a new way of reading’, and they were told that they were going to learn ‘secret signs’. Local teachers received individual supervision (mean of 1.5 hours) in how to teach the programme. They were encouraged to apply a dialogic approach and facilitate a collaborative atmosphere. The intervention material consisted of PAS cards and a poster with the PAS symbols. Additional materials included mirrors, alphabet letter cards, pictures of objects and the right level of texts for the reading exercises.

### **Design and procedure**

The study is a multiple baseline/probe design across participants (Baer et al., 1968; Horner & Baer, 1978), with repeated testing over 18 weeks. More specifically, there were five pre-intervention measurement occasions, corresponding to a control condition, eight measurement occasions during the intervention and five post-intervention measurement occasions. For the baseline and post-phase, we used a probe design, with breaks of varying lengths between the measurement occasions. For the intervention phase, we used weekly measurements for all the occasions.

Cases 2–10 were separated in two groups, and the intervention onset were staggered across the participants in randomised order. Cases 11, 12 and 13 were recruited later in the process, introduced to the intervention at different time points, and followed the structure of the first set of participants in the original design. For details concerning the structure of the measurement procedure, see Figure 3.

At all the measurement points, the assessment followed the same procedure and took about 20 minutes. The testing was conducted by the first author or by professionals from the School Psychology Service.

The children received 32 hours of intervention distributed in eight weekly sessions. Further, the teachers answered a short questionnaire about the organisation, content and extent of the special needs education delivery in the periods before and after the intervention.

### **Fidelity**

All teachers filled in a log form for each lesson and briefly described how the intervention programme had worked out. All lessons were audio recorded. Ten percent of the recordings were rated, and the correspondence between the recordings and the log form were close to 100%. The measurement sessions were also audio recorded. A random selection of 15% of the sessions were picked out for each participant and rated to make sure that the procedure was followed. One student was removed from the sample due to procedural failures. There was 100% procedural fidelity for the remainder of the sample.

### **Outcome measures**

#### **Weekly measures**

Measurement tools for the baseline, intervention and post-phases were constructed for the purpose of this study. To minimise threats to internal validity from the testing effects, we constructed alternate forms for the reading and spelling tests for each testing point (described in detail below). The reading measures were made in 36 (2 x 18) versions and the spelling test in 18 versions.

#### ***Decoding***

#### ***Regular words***



We selected regular words from a database of the 500 most frequent Norwegian words (Norwegian word frequency list), which were separated in nine groups based on their level of phonological complexity. The words in Groups 1 and 2 were simple c/v or v/c words with two graphemes, followed by Group 3 consisting of cvc words and then, successively, vcc, cvcv, cvcc, ccvccv, cccvccv and cccvccvc. For each word group, we made a pool of 22 words. The alternate forms of the reading tests were then constructed by randomly picking a selection of six words from each group, totalling 54 items. The scoring criteria were the number of correct words read in one minute. A word was counted as correctly read if all the graphemes were articulated.

#### *Pseudowords*

The Pseudoword Test was constructed in a similar way as the regular word test, and the aim was for it to correspond to this test both in structure and level of complexity. The scoring criteria were the number of correct words read in one minute. Words read in a phonologically acceptable way were scored as correct if all the graphemes were articulated.

#### *Irregular words*

For the construction of the Irregular Word Test, we selected irregular words from a Norwegian reading test. The words were categorised based on the number of letters and syllables and separated in three groups. The first group (22 words in total) mainly consisted of high-frequency irregular words with two or three graphemes. The second group consisted of 44 one-syllable words, while the third pool consisted of 44 two-syllable words. The scoring criteria were the number of words read in an orthographically correct manner in one minute. Thus, a pronunciation that was phonologically correct but orthographically incorrect was scored as zero.

All word reading tests were administered twice, at each measurement time, with two alternate forms. The reliability of the reading tests was measured by correlating the scores

from the two alternate forms at each timepoint. The average correlations across timepoints are for the regular words .968, for the pseudowords .898 and for the irregular words .883.

### ***Reading Accuracy***

Since ‘guessing strategies’ were highly frequent within the sample, we were interested in whether the intervention could improve the students’ reading accuracy. Therefore, we calculated the reading accuracy level to separate reading accuracy from reading speed (Juul et al., 2014). Accuracy was defined as the percentage of correctly read words out of the number of total items passed in one minute.

### ***Spelling***

The construction of the alternate forms of the pseudoword spelling tests followed the same procedure as that of the reading measures, and the words were selected from the same pool as the pseudoword reading test. We picked four items for each level of difficulty, totalling 36 items in nine blocks. The time limit was four minutes.

### ***Transposition of pictographic symbols***

The transposition test was also constructed using alternate forms (13 in total) and aimed to measure progress in the use of the articulation cards. Each test consisted of 24 pseudowords scripted in the font of the articulation cards. The pseudowords were selected from the same pools, as described above. They all had a vc, cv, vcc or cvcv structure and consisted of two to four symbols (letters). The scoring criteria were the number of pictographic words transposed into a correct alphabetic script in four minutes (max score was 24). Supplementary scoring also included the number of ‘pictographic graphemes’ transposed correctly into alphabetic graphemes (max score for this was 64).

### ***Pegboard test (control task)***

The pegboard test was considered to be unrelated to the intervention and was mainly conducted to control for training effects from repeated testing. The materials used in this test were simply a pegboard and staples, and for each testing, the participants were asked to put as many staples as possible in a vertical line on the pegboard, with a time limitation of 30 seconds. For each measurement occasion, this was administered twice.

### ***Standardised pre- and post-tests***

To examine the effects of the intervention on measures not involved in the weekly measurement procedure, one word reading test (containing four subtests) and one pseudoword reading test (containing three subtests) from the STAS battery (Klinkenberg & Skaar, 2001) were administered as pre- and post-tests. This battery is standardised and normed on a Norwegian sample from the second to tenth grade. The scoring criteria were the number of correct words read out loud in 40 seconds. Composite scoring options for words and pseudowords were also standardised and normed.

### ***Social validity***

The teachers' evaluation was measured by a post-study questionnaire. This consisted of 17 questions regarding the students' motivation, general efforts during the sessions and the individual benefits of the training in respect of changes in the reading and spelling strategies.

### **Analysis**

#### **Visual inspection of data**

For the visual data inspection, we conducted a full sample overview (depicted in Figure 5) and graphic charts for each participant (supplementary material). We also utilised the protocol to evaluate the stability, trend, level changes and data overlap provided by Gast and Spriggs (2014; see also Lane & Gast, 2013).

### **Effect size measures**

In line with Lobo et al. (2017) and Wolfe et al. (2019), we report multiple effect size measures for the within-case effects, the standardised mean difference (SMD) (Busk & Serlin, 1992; Olive & Smith, 2005) and the Tau-U (Parker et al., 2011a; Parker et al., 2011b). Further, in order to get a measure of the overall effect across participants, we also calculated the between-case effect using between-case standardised mean difference (BC-SMD) estimates developed by Pustejovsky et al. (2014; Valentine et al., 2016).

The *SMD* expresses the effect in standard deviation units based on a comparison of the mean scores for each phase of the study and for each participant (Busk & Serlin, 1992; Gierut et al., 2015; Olive & Smith, 2005). The effect size was calculated by using the online calculator provided at <https://jepusto.shinyapps.io/SCD-effect-sizes>. We used the pooled *SD* across all calculations. According to suggested standards (Harrington & Velicer, 2015), an estimate lower than 1 is a small effect, 1–2.5 a medium effect and over 2.5 a large effect.

The Tau-U combines non-overlapping techniques with a baseline control. The effect size estimate was calculated by comparing all pairs of data from two phases (Parker et al., 2011). We calculated the Tau-U effect size by using the online calculator available at <http://www.singlecaseresearch.org/calculators> (Vannest et al., 2011). Based on the standards, a Tau-U estimate lower than .20 is a small effect, from .20–.60 a moderate effect, from .60–.80 a large effect and more than .80 a very large effect (Vannest & Ninci, 2015).

The BC-SMD was calculated by using the calculator available at <https://jepusto.shinyapps.io/scdhlm/>. The BC-SMD uses a two-level model with a within-case regression model at the first level and a between-case variation at the second level. The standard applies a design comparable effect size using the same benchmarks as those of Cohen's (1988) *d*, i.e. small effect = 0.10, medium effect = .30 and large effect = .50. The calculator requires a specification of both the fixed and random levels in the baseline phase

and a fixed level in the intervention phase. Further, we specified the models by using the criteria suggested by Wolfe et al. (2019): If the treatment effect across the participants differed by more than 10% of the scale on the y axis, we added a specification for a random level in the intervention phase. To determine whether to include specifications for trends in the models, we first inspected the graphic charts (Wolfe et al., 2019). If all participants had a clear and visible trend, we specified a fixed trend in that phase. For specifications of random trends, we converted the criteria from Wolfe et al. and adapted them to our sample size. They specified random trends if one out of three participants displayed a clear trend. We converted this to one-third of the sample (33.3%) and specified a random trend if four out of eleven participants showed a visibly clear trend. For all models, we choose the restricted maximum likelihood (RML) estimation method.

Since the effects on reading and spelling in this study are dependent on the transfer from learning the PAS symbols, we did not expect an immediate effect on the dependent variables from the intervention onset (Klingbeil et al., 2017). However, the results from the transposition test showed that most students automatized the use of the PAS symbols during the first two weeks of the programme, which is depicted in Figure 4. To calculate the effect from the reading and spelling outcomes, therefore, we only included data from week three of the intervention.

For the three reading measures and the pegboard test, we used two independent measures for each time point. The mean score was calculated for each pair of measures administered at the same time point and plotted as new values for each measurement occasion. The intervention and post-test phase were merged in the analysis.

In the spelling and transposition tests, some students reached the ceiling level for the time spent. For these measures, we extrapolated the scoring from the percentage of correct scores and achieved a ‘time bonus’.

## **Results**

### **Visual inspection**

#### ***Stability***

The visual inspection revealed that most outcome measures were the stability criterion of 80% of the data points within the range of +/-25% of the phase median value fulfilled across the sample (Lane & Gast, 2014). The percentage of participants with stable data across all phases of the study were as follows: 54.5% for the regular words, 72.7% for irregular words and pseudowords, 63.6% for the spelling test and 81.8% for the control. For individual results, see Table 2.

#### ***Overlap***

Overlap was evaluated by the Tau-U, which indicated only minor overlap in the data points between the two phases. Displayed in Figure 5 is a small but noticeable change associated with the intervention onset present in the full sample, most clearly depicted for reading accuracy and spelling.

#### ***Trend***

To evaluate the within-phase trends, we calculated the within-phase improvement for all the dependent variables (Gast, 2014). This was done by calculating the change in the full sample mean score from the first to last measurement occasion for each phase of the study. The results show some within-phase improvement in the baseline phase for all the dependent variables except the regular words. For the pseudowords, irregular words and spelling, the total change from the last measurement occasion in the baseline phase to the last post-measurement was between 30–42% for the regular words, i.e. 14%. The control test shows a different pattern of within-phase improvement, with most changes within the baseline phase

and smaller changes in the intervention and post-phases of the study. The results are displayed in Table 3.

### ***Level***

Descriptive results from the weekly measures show improvements in all the dependent variables from the baseline phase to the intervention and post-phase for all participants. The average level improvement calculated as a percentage from baseline to the post-phase was 37% for the regular words, 60% for the pseudowords, 45% for the irregular words, 40% for the spelling test and 23% for the control task. For individual differences, see Table 2.

### **Intervention effects**

#### ***Weekly measures***

The Tau-U results showed that most students scored above the benchmark for ‘large change’, i.e.  $\text{Tau-U} = .60$  on all the dependent variables (Vannest & Ninci, 2015). The percentage of participants with significant scores above .60 was 91 for the regular words, 75 for pseudowords, 58 for the irregular words and 66 for the spelling test. The SMD effect sizes are in line with this. According to the suggested standards from Harrington and Velicer (2015), the percentage of participants with significant scores was 91 in the category of medium or large effects for the regular words, 100 for the pseudowords, 72.7 for the irregular words and 45.4 for the spelling test (see Table 3). The most consistent within-case effects in terms of the magnitude of the effect and level changes were seen on the pseudoword reading test and the standardised pseudoword composite (Table 6). The descriptive results for all the within-case effects are shown in Table 4.

Concerning the calculation for reading accuracy, all students showed improvement in the rate of correctly read words from the baseline phase to the intervention/post-phase. The mean improvement across the reading measures from the baseline to the post-test was 10% for regular words and pseudowords and 12% for irregular words (see Table 2, Figure 5).

### ***Between-case effects***

The results from the between-case analysis show a significant positive effect on all the outcome variables, with a range of  $d = 0.31$ – $0.67$ . In the analysis, we followed the procedure described above. Details are shown in Table 6.

### **Social validity questionnaire**

The teachers provided positive evaluations of the students' benefit from the training. In the post-intervention questionnaire, the teachers evaluated the students' general benefit from the training on a scale from 1–6. All the participants scored in the 4-5 range. On a question about changes in reading and spelling strategies, the scores were in the same range. The teachers also reported the students' effort in the sessions during the intervention, and for 10 out of 11 students, this was categorised as 'better' or 'much better' compared to the teachers' experiences from prior special education sessions.

### **Standardised pre- and post-tests**

The results from the STAS test, administered prior to and in the post-intervention, show that on the pseudoword composite, most students had improved equivalent to approximately  $0.5 SD$  when converting the raw score changes from pre- to post-test to the age norms for each participant. For the word-reading measures (including both regular words and irregular words), the improvement was between  $0.23$  and  $0.85 SD$ . See Table 6 for details.

### **Control test**

The pegboard test was conducted as a control for the practice effects due to repeated measurements. According to the Tau-U, 54.5% of the sample scored significantly above the



RUNNING HEAD: Articulatory consciousness training in students with severe dyslexia 20

benchmark for large effects. For the SMD, the percentage of participants with significant scores above the benchmark for the medium effect was 91. The BC-SMD estimate was  $d = 0.32$ .

## Discussion

This study evaluated the effects of an eight-week intervention aimed at improving phonological reading and spelling strategies in 11 students with severe dyslexia. The study results include both the students' progress in automatizing the pictographic symbols and the possible transfer effects to alphabetic reading and spelling. In addition, we used a task unrelated to the intervention to control for testing effects.

The between-case results showed significant improvement on all outcome variables, with the most substantial effects for pseudoword reading and spelling and reading of regular words and only minor effects on the irregular word measure. Since the training was primarily aimed at improving phonological strategies in reading and spelling, it is not surprising that the effect on irregular words, which required other reading strategies, was in a lower range.

The within-case effect size measures showed significant improvement across the dependent variables for most participants. Although the effect size measures showed some divergent results concerning the magnitude of the effects, the measures applied seem to agree in terms of the judgement of the effect or lack thereof. This is in line with previous studies applying multiple effects size measures in SCED (Olive & Smith, 2005; Wolfe et al., 2019). An exception was the spelling test, where the SMD measure indicated a more conservative judgement than the Tau-U measure for two participants, as well as the control test where the Tau-U seemed to be more conservative than the SMD for three participants.

Further, the results showed that most students automatized the PAS symbols during the first two weeks of intervention. The teachers also reported that the students were

motivated to learn the ‘secret signs’ and generally put more effort in the training than in previous reading lessons. This suggests that the ‘articulatory way’ had some sort of appeal to this sample of reading-disabled students.

### **Interpretation of the findings**

Repeated measurements of the same variables are vulnerable to practice effects, which are a possible threat to the internal validity of single case studies (Gast 2014). For this reason, we added a non-equivalent dependent variable (Shadish et al., 2002), the pegboard test, as a control. An important issue is that we not only found effects on the outcomes targeted in the intervention but also significant and small-to-moderate related effects. However, according to the descriptive results and the visual inspection, the average level of improvement on the pegboard test was nearly half the improvement on the reading and spelling measures, respectively, 23% for pegboard and 36–60% for reading and spelling.

Moreover, the evaluation of the within-phase trends showed a different pattern of improvement for the control test compared to the other measure. Most of the improvement was in the baseline phase, but for the other dependent variables, the improvement was recorded in the intervention and post-test phases. This suggests that the five measurement occasions during the baseline were sufficient to capture the practice effects from repeated testing.

It should also be noted that we used alternate forms for all the reading and spelling tests in order to control for practice effects, which was not done for the pegboard task. A study on testing effects in neuropsychological measures showed that the use of alternate forms of tests to some extent prevents testing effects on some tests, even though continued learning occurred when an advantageous test-taking strategy could be identified (Berlinger et al., 2005). However, for verbal memory measures, it has been demonstrated that alternate forms

can eliminate practice effects (Roediger & Karpicke, 2005). Another study also indicated a general appearance of more substantial testing effects on motoric measures than on word reading (Levine et al., 2004).

Thus, it seems that the nature of the test plays a role in the magnitude of the testing effects, including when using alternate forms. The results from the pegboard test limit the conclusions that can be drawn from the study concerning the benefits on the primary outcomes. However, based on the line of arguments above, it appears that testing effects are not a likely explanation for all changes in the current study.

The results from the immediate and delayed post-tests show that, for most participants, the effect from the intervention remained after the eight weeks of training. According to the post-study questionnaire, most participants continued to use elements from the intervention programme after completing the intervention. The increased effects at post-test can, therefore, be partly explained as a continuation effect. This indicates that less intensive training was enough to maintain the achieved effect; however, with the effects from the pegboard task as a caution, this could also be interpreted as continued testing effects. Nevertheless, the within-phase analysis showed minor improvement in the post-phase for the control test compared to the other dependent variables. Therefore, the continued effect is less likely to be due to the repeated measurements.

Notably, two students (Cases 2 and 3) scored below the benchmarks for the small effects on more than one of the independent variables and across two or more effect size measures. For Case 3, poor results can probably be explained by personal situation changes due to relocation of residence during the intervention period. This case also showed weaker progress than the rest of the sample on the transposition test (Figure 4), which means that the probability of transfer effects to reading and writing was less likely. The poor results for Case 2 seemed random and could not be explained using the conditions outlined.

In conclusion, this study exemplifies that articulatory consciousness training may have positive effects on reading and spelling outcomes for students with severe dyslexia. However, the results must be interpreted with caution as significant effect sizes were recorded for the irrelevant task. Furthermore, from the perspective of generalisation, since this study only included 11 participants, future studies need to focus on group comparisons with randomisation to determine results.

## References

- Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1*(1), 91–97.  
<https://doi.org/10.1901/jaba.1968.1-91>
- Beglinger, L. J., Gaydos, B., Tangphao-Daniels, O., Duff, K., Kareken, D. A., Crawford, J., Siemers, E. R. (2005). Practice effects and the use of alternate forms in serial neuropsychological testing. *Archives of Clinical Neuropsychology, 20*(4), 517–529.  
<https://doi.org/10.1016/j.acn.2004.12.003>
- Boyer, N., & Ehri, L. C. (2011). Contribution of phonemic segmentation instruction with letters and articulation pictures to word reading and spelling in beginners. *Scientific Studies of Reading, 15*(5), 440–470. <https://doi.org/10.1080/10888438.2010.520778>
- Busk, P. L. and R. C. Serlin (1992). Meta-analysis for single-case research. Single-case research design and analysis: New directions for psychology and education. Hillsdale, NJ, US, Lawrence Erlbaum Associates, Inc: 187-212.
- Byrne, B., Olson, R. K., Samuelsson, S., Wadsworth, S., Corley, R., DeFries, J. C., & Willcutt, E. (2006). Genetic and environmental influences on early literacy. *Journal of Research in Reading, 29*(1), 33–49. <https://doi.org/10.1111/j.1467-9817.2006.00291.x>
- Castiglioni-Spalten, M. L., & Ehri, L. C. (2003). Phonemic awareness instruction: Contribution of articulatory segmentation to novice beginners' reading and spelling. *Scientific Studies of Reading, 7*(1), 25–52.  
[https://doi.org/10.1207/S1532799XSSR0701\\_03](https://doi.org/10.1207/S1532799XSSR0701_03)
- Christopher, M. E., Hulslander, J., Byrne, B., Samuelsson, S., Keenan, J. M., Pennington, B., ... Olson, R. K. (2013). Modeling the etiology of individual differences in early

- reading development: Evidence for strong genetic influences. *Scientific Studies of Reading*, 17(5), 350–368. <https://doi.org/10.1080/10888438.2012.729119>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Laurence Erlbaum.
- Duranovic, M., & Sehic, S. (2013). The speed of articulatory movements involved in speech production in children with dyslexia. *Journal of Learning Disabilities*, 46(3), 278–286. <https://doi.org/10.1177/0022219411419014>
- Eklund, K., Torppa, M., Aro, M., Leppänen, P. H. T., & Lyytinen, H. (2015). Literacy skill development of children with familial risk for dyslexia through grades 2, 3, and 8. *Journal of Educational Psychology*, 107(1), 126–140. <https://doi.org/10.1037/a0037121>
- Elbro, C., Petersen, D. K., & Borstrom, I. (1998). Predicting dyslexia from kindergarten: The importance of distinctness of phonological representations of lexical items. *Reading Research Quarterly*, 33(1), 36–60. <https://doi.org/10.1598/RRQ.33.1.3>
- Fälth, L., Gustafson, S., & Svensson, I. (2017). Phonological awareness training with articulation promotes early reading development. *Education*, 137(3), 261–276.
- Fawcett, A. J., & Nicolson, R. I. (2002). Children with dyslexia are slow to articulate a single speech gesture. *Dyslexia*, 8(4), 189–203. <https://doi.org/10.1002/dys.222>
- Gast, D. L., & Ledford, J. R. (2014). *Single case research methodology: Applications in special education and behavioral sciences* (2nd ed.). Routledge.
- Griffiths, S., & Frith, U. (2002). Evidence for an articulatory awareness deficit in adult dyslexics. *Dyslexia*, 8(1), 14–21. doi:10.1002/dys.201
- Hämäläinen, J. A., Salminen, H. K., & Leppänen, P. H. T. (2013). Basic auditory processing deficits in dyslexia: Systematic review of the behavioral and event-related potential/

field evidence. *Journal of Learning Disabilities*, 46(5), 413–427.

<https://doi.org/10.1177/0022219411436213>

Harrington, M., & Velicer, W. F. (2015). Comparing visual and statistical analysis in single-case studies using published studies. *Multivariate Behavioral Research*, 50(2), 162–183. <https://doi.org/10.1080/00273171.2014.973989>

Horner, R. D., & Baer, D. M. (1978). Multiple-probe technique: A variation of the multiple baseline. *Journal of Applied Behavior Analysis*, 11(1), 189–196.  
<https://doi.org/10.1901/jaba.1978.11-189>

Hulme, C., & Snowling, M. (2009). *Developmental disorders of language learning and cognition*. Wiley-Blackwell.

Hulme, C., Nash, H. M., Gooch, D., Lervåg, A., & Snowling, M. J. (2015). The Foundations of Literacy Development in Children at Familial Risk of Dyslexia. *Psychological Science*, 26(12), 1877–1886. doi:10.1177/0956797615603702

Joly-Pottuz, B., Mercier, M., Leynaud, A., & Habib, M. (2008). Combined auditory and articulatory training improves phonological deficit in children with dyslexia. *Neuropsychological Rehabilitation*, 18(4), 402–429.  
<https://doi.org/10.1080/09602010701529341>

Juul, H., Poulsen, M., & Elbro, C. (2014). Separating speed from accuracy in beginning reading development. *Journal of Educational Psychology*, 106(4), 1096–1106.  
<https://doi.org/10.1037/a0037100>

Kausrud, T. (2003). PAS - steget inn i alfabetisk lesning: bruk av piktografisk-artikulatoriske symboler for utvikling av fonologisk bevissthet. UiO. <http://urn.nb.no/URN:NBN:no-6248>

- Klingbeil, D., Norman, E., & Nelson, P. (2017). Precision of curriculum-based measurement reading data: Considerations for multiple-baseline designs. *Journal of Behavioral Education, 26*(4), 433–451. <https://doi.org/10.1007/s10864-017-9282-7>
- Klinkenberg, J. E., & Skaar, E. (2003). STAS. Standardisert Test i Avkoding og Staving. *Manual. PP-Tjenestens Matriellservice. Jaren.*
- Lalain, M., Joly-Pottuz, B., Nguyen, N., & Habib, M. (2003). Dyslexia: The articulatory hypothesis revisited. *Brain and Cognition, 53*(2), 253–256. [https://doi.org/10.1016/S0278-2626\(03\)00121-0](https://doi.org/10.1016/S0278-2626(03)00121-0)
- Lane, J. D., & Gast, D. L. (2014). Visual analysis in single case experimental design studies: Brief review and guidelines. *Neuropsychological Rehabilitation, 24*(3-4), 445–463. <https://doi.org/10.1080/09602011.2013.815636>
- Levine, A. J., Miller, E. N., Becker, J. T., Selnes, O. A., & Cohen, B. A. (2004). Normative data for determining significance of test-retest differences on eight common neuropsychological instruments. *The Clinical Neuropsychologist, 18*(3), 373–384. <https://doi.org/10.1080/1385404049052420>
- Lewis, A. B., Avrich, A. A., Freebairn, A. L., Taylor, G. H., Iyengar, K. S., & Stein, M. C. (2011). Subtyping children with speech sound disorders by endophenotypes. *Topics in Language Disorders, 31*(2), 112–127. <https://doi.org/10.1097/TLD.0b013e318217b5dd>
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*(1), 1–36. [https://doi.org/10.1016/0010-0277\(85\)90021-6](https://doi.org/10.1016/0010-0277(85)90021-6)
- Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences, 4*(5), 187–196. [https://doi.org/10.1016/S1364-6613\(00\)01471-6](https://doi.org/10.1016/S1364-6613(00)01471-6)



- Lindamood, P., & Lindamood, P. (1998). Lindamood Phoneme Sequencing Program for Reading, Spelling, and Speech (LiPS). <https://lindamoodbell.com/program/lindamood-phoneme-sequencing-program>
- Lobo, A. M., Moeyaert, A. M., Baraldi Cunha, A. A., & Babik, A. I. (2017). Single-case design, analysis, and quality assessment for intervention research. *Journal of Neurologic Physical Therapy, 41*(3), 187–197. <https://doi.org/10.1097/NPT.000000000000187>
- Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of Dyslexia, 53*, 1–14.
- McBride-Chang, C. (1996). Models of Speech Perception and Phonological Processing in Reading. *Child Development, 67*(4), 1836-1856. doi:10.2307/1131735
- Montgomery, D. (1981). Do dyslexics have difficulty accessing articulatory information? *Psychological Research, 43*(2), 235-243. doi:10.1007/BF00309832
- Noordenbos, M. W., & Serniclaes, W. (2015). The categorical perception deficit in dyslexia: A meta-analysis. *Scientific Studies of Reading, 19*(5), 340–359. <https://doi.org/10.1080/10888438.2015.1052455>
- Olive, M. L., & Smith, B. W. (2005). Effect Size Calculations and Single Subject Designs. *Educational Psychology, 25*, 313-313), p.313-324.
- Olson, R. K., Keenan, J. M., Byrne, B., Samuelsson, S., Coventry, W. L., Corley, R., ... Hulslander, J. (2011). Genetic and environmental influences on vocabulary and reading development. *Scientific Studies of Reading, 15*(1), 26–46. <https://doi.org/10.1080/10888438.2011.536128>
- Ottem, E., & Kausrud, T. (2001). Use of pictographic-articulatory symbols to promote alphabetic reading in a language-impaired boy: Case study. *Augmentative and Alternative Communication, 17*(1), 52. <https://doi.org/10.1080/aac.17.1.52.60>

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011a). Combining Nonoverlap and Trend for Single-Case Research: Tau-U. *Behavior Therapy, 42*(2), 284-299.

doi:10.1016/j.beth.2010.08.006

Parker, R. I., Vannest, K. J., & Davis, J. L. (2011b). Effect Size in Single-Case Research: A Review of Nine Nonoverlap Techniques. In (Vol. 35, pp. 303-322). Los Angeles, CA.

Pennington, B. F., Santerre-Lemmon, L., Rosenberg, J., Macdonald, B., Boada, R., Friend, A., ... Goodman, S. (2012). Individual prediction of dyslexia by single versus multiple deficit models. *Journal of Abnormal Psychology, 121*(1), 212-224.

<https://doi.org/10.1037/a0025823>

Preston, J. L., Hull, M., & Edwards, M. L. (2013). Preschool speech error patterns predict articulation and phonological awareness outcomes in children with histories of speech sound disorders. *American Journal of Speech-Language Pathology, 22*(2), 173-184.

[https://doi.org/10.1044/1058-0360\(2012/12-0022\)](https://doi.org/10.1044/1058-0360(2012/12-0022))

Pustejovsky, J. E., Hedges, L. V., & Shadish, W. R. (2014). Design-comparable effect sizes in multiple baseline designs: A general modeling framework. *Journal of Educational and Behavioral Statistics, 39*(5), 368-393. <https://doi.org/10.3102/1076998614547577>

Ramus, F., Marshall, C. R., Rosen, S., & van der Lely, H. K. J. (2013). Phonological deficits in specific language impairment and developmental dyslexia: Towards a multidimensional model. *Brain, 136*(2), 630-645.

<https://doi.org/10.1093/brain/aws356>

Roediger, H. L., & J. D. Karpicke (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181-210.

Sénéchal, M., Ouellette, G., & Young, L. (2004). Testing the concurrent and predictive relations among articulation accuracy, speech perception, and phoneme awareness.

*Journal of Experimental Child Psychology*, 89(3), 242–269.

<https://doi.org/10.1016/j.jecp.2004.07.005>

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology*, 72(2), 95–129.

<https://doi.org/10.1006/jecp.1998.2481>

Skjelfjord, V. J. (1987). Phonemic segmentation: An important subskill in learning to read. *Scandinavian Journal of Educational Research*, 31(1), 41–57.

<https://doi.org/10.1080/0031383870310104>

Snowling, M. J., Gallagher, A., & Frith, U. (2003). Family risk of dyslexia is continuous: Individual differences in the precursors of reading skill. *Child Development*, 74(2), 358–373. <https://doi.org/10.1111/1467-8624.7402003>

Snowling, M., Hulme, C., Wells, B., & Goulandris, N. (1992). Continuities between speech and spelling in a case of developmental dyslexia. *An Interdisciplinary Journal*, 4(1), 19–31. <https://doi.org/10.1007/BF01027070>

Swan, D., & Goswami, U. (1997). Phonological awareness deficits in developmental dyslexia and the phonological representations hypothesis. *Journal of Experimental Child Psychology*, 66(1), 18–1. [https://doi.org/10.1016/S0022-0965\(02\)00169-8](https://doi.org/10.1016/S0022-0965(02)00169-8)

Tate, R. L., Perdices, M., Rosenkoetter, U., Shadish, W., Vohra, S., Barlow, D. H., ... Wilson, B. (2016). The single-case reporting guideline in behavioural interventions (SCRIBE) 2016 statement. *Journal of School Psychology*, 56, 133–142.

<https://doi.org/10.1016/j.jsp.2016.04.001>

Torgesen, J. K., Alexander, A. W., Wagner, R. K., Rashotte, C. A., Voeller, K. K. S., & Conway, T. (2001). Intensive remedial instruction for children with severe reading

disabilities: Immediate and long-term outcomes from two instructional approaches.

*Journal of Learning Disabilities*, 34(1), 33–58.

<https://doi.org/10.1177/002221940103400104>

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Herron, J., Lindamood, P., & Torgesen, J. K.

(2010). Computer-assisted instruction to prevent early reading difficulties in students at risk for dyslexia: Outcomes from two instructional approaches. *Annals of Dyslexia*,

60(1), 40–56. <https://doi.org/10.1007/s11881-009-0032-y>

Trainin, G., Wilson, K. M., Murphy-Yagil, M., & Rankin-Erickson, J. L. (2014). Taking a

different route: Contribution of articulation and metacognition to intervention with at-risk third-grade readers. *Journal of Education for Students Placed at Risk*, 19(3-4),

183–184), 183–195. <https://doi.org/10.1080/10824669.2014.972103>

Valentine, J. C., Tanner-Smith, E. E., Pustejovsky, J. E., & Lau, T. S. (2016). Between-case

standardized mean difference effect sizes for single-case designs: A primer and

tutorial using the scdhlms web application. *Campbell Systematic Reviews*, 12(1), 1–31.

<https://doi.org/10.4073/cmdp.2016.1>

Vannest, K. J., & Ninci, J. (2015). Evaluating intervention effects in single-case research

designs. *Journal of Counseling & Development*, 93(4), 403–411.

<https://doi.org/10.1002/jcad.12038>

What Works Clearinghouse, Institute of Education Science, U.S. Department of Education.

(2008). WWC Intervention Report. <https://files.eric.ed.gov/fulltext/ED503694.pdf>

What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education.

(2015). *Beginning reading intervention report: Lindamood Phoneme Sequencing®*

(LiPS®). What Works Clearinghouse. <https://ies.ed.gov/ncee/wwc/Intervention/665>

Wise, B. W., Ring, J., & Olson, R. K. (1999). Training phonological awareness with and without explicit attention to articulation. *Journal of Experimental Child Psychology*, 271.

Wolfe, K., Dickenson, T. S., Miller, B., & McGrath, K. V. (2019). Comparing visual and statistical analysis of multiple baseline design graphs. *Behavior Modification*, 43(3), 361–388. <https://doi.org/10.1177/0145445518768723>



**Table 1**

*Sample Descriptives*

Raw Scores

Case	G	Age Y/M	Case Characteristics			Sample Inclusion tests		Pretests prior to intervention				
			Fam.Risk	Lang. Cons.	Parents Ed. Level (years)	STAS Pseudo Word Reading	WISC IV block design	WISC IV Simi lari ties	WISC IV digit span	WISC IV compre hension	WISC IV matrix reasoning	STAS Pseudo Word Dictate
2	boy	10/01	n.a.	No	12	-2.64	10	10	6	8	11	-3.2
3	boy	10/00	Yes	Yes	12	-2.29	11	7	5	7	10	-3.7
4	girl	13/02	No	No	15	-2.25	9	8	6	9	9	-0.8
6	girl	14/00	Yes	Yes	15	-2.74	10	11	3	6	9	-2.7
7	boy	11/01	Yes	No	15	-2.73	12	9	6	10	13	-2.2
8	boy	11/08	Yes	Yes	15>	-2.52	13	11	4	8	7	-2.2
9	boy	11/02	Yes	Yes	15>	-2.73	12	12	7	9	7	-1.2
10	boy	12/03	Yes	Yes	15>	-2.52	19	12	6	12	12	-0.8
11	boy	14/11	n.a.	Yes	15	-2.17	13	6	4	12	10	-2.1
12	boy	13/06	Yes	Yes	15>	-3.40	11	9	2	10	9	-4.3
13	boy	11/05	Yes	Yes	15>	-2.31	14	7	4	8	8	-1.4

Note:

G = Gender

WISC IV: Scaled scores

STAS: Norwegian standardised test for reading and spelling, standard deviations from the age mean score.

Lang.Cons = Parents' concerns about early language development

Fam.Risk = Family risk – dyslexia.

n.a. = not available

**Table 2**

*Descriptive Results*

		Outcome Measure							
ID	Phase of study	Regular words		Pseudowords		Irregular words		Spelling	Peg Board
		Median	Accuracy	Median	Accuracy	Median	Accuracy	Median	Median
2	Baseline	<b>10.5</b>	46.19	<b>9</b>	44.39	<b>9</b>	46.15	10	<b>10.0</b>
	Intervention	<b>14</b>	54.86	<b>11.5</b>	50.94	<b>8</b>	41.0	<b>17</b>	<b>12.5</b>
	Post	16	62.45	<b>13.5</b>	58.48	<b>13.5</b>	53.50	<b>16</b>	<b>13.5</b>
3	Baseline	14	75.99	<b>13</b>	86.67	3.5	27.64	<b>16</b>	<b>13.5</b>
	Intervention	<b>13.5</b>	84.93	<b>13.5</b>	88.333	4.56	38.24	<b>16</b>	16.0
	Post	<b>18.0</b>	85.05	<b>10</b>	64.84	<b>9</b>	59.0	<b>18</b>	<b>18.5</b>
4	Baseline	<b>25.5</b>	75.15	<b>14.5</b>	59.20	<b>11.5</b>	55.66	<b>22</b>	<b>8.0</b>
	Intervention	<b>31</b>	84.42	20	75.06	18	71.96	<b>28.5</b>	<b>10.0</b>
	Post	<b>32</b>	90.53	<b>26.5</b>	85.65	<b>22</b>	79.68	<b>32</b>	<b>11.0</b>



6	Baseline	30	82.43	23	75.38	23	65.22	21	15.0
	Intervention	<b>32.5</b>	86.73	<b>30</b>	87.70	<b>26</b>	76.93	<b>26</b>	<b>15.5</b>
	Post	<b>35</b>	90.25	<b>32</b>	90.98	<b>30</b>	84.27	<b>26</b>	<b>16.5</b>
7	Baseline	19	78.87	14	66.67	17	68.48	21	11.5
	Intervention	26.0	86.73	23	88.79	19	90.08	26	14.0
	Post	28	94.08	<b>25</b>	92.62	<b>23</b>	95.99	<b>26</b>	<b>15.0</b>
8	Baseline	<b>18</b>	84.52	<b>18.5</b>	87.60	9	56.06	<b>20</b>	<b>14.0</b>
	Intervention	<b>21</b>	88.36	<b>20.5</b>	87.98	<b>13</b>	78.34	<b>27</b>	<b>16.0</b>
	Post	<b>22</b>	88.69	<b>24.5</b>	91.51	<b>14</b>	78.95	<b>30</b>	<b>14.0</b>
9	Baseline	16	68.37	12	58.04	<b>12.5</b>	64.52	<b>16</b>	<b>12.5</b>
	Intervention	<b>24</b>	91.14	<b>17</b>	77.89	<b>16</b>	76.16	18	<b>13.5</b>
	Post	<b>27</b>	93.33	<b>19</b>	82.75	<b>18</b>	81.10	<b>30</b>	<b>15.0</b>
10	Baseline	<b>19.5</b>	75.22	<b>15</b>	76.13	<b>12</b>	57.19	<b>30</b>	<b>17.0</b>
	Intervention	<b>23</b>	88.75	<b>25</b>	89.85	<b>15</b>	66.34	<b>32</b>	<b>19.0</b>
	Post	<b>25</b>	92.64	<b>27</b>	94.34	<b>13</b>	62.97	<b>32</b>	<b>22.0</b>
11	Baseline	<b>29</b>	92.81	<b>29</b>	93.31	<b>27</b>	89.08	22	14.0
	Intervention	<b>32</b>	96.57	<b>32</b>	93.90	<b>28</b>	92.89	25	<b>14.5</b>
	Post	<b>34</b>	96.24	<b>33</b>	94.28	<b>27</b>	93.72	27	<b>17.0</b>
12	Baseline	<b>7</b>	22.67	<b>1.5</b>	8.7	2	8.04	5	<b>17.0</b>
	Intervention	9.5	42.30	6	25.49	3	16.15	<b>10</b>	<b>19.0</b>
	Post	<b>10.5</b>	41.64	6	20.47	4	25.9	<b>11</b>	<b>19.5</b>
13	Baseline	<b>26</b>	75.91	<b>21</b>	76.04	<b>8.5</b>	26.02	<b>29</b>	<b>14.0</b>
	Intervention	<b>32</b>	81.38	<b>26</b>	77.19	<b>14</b>	36.74	<b>33</b>	<b>15.0</b>
	Post	<b>36</b>	86.89	<b>28</b>	82.47	<b>17</b>	46.61	<b>34</b>	<b>16.0</b>

Note:

Median values in bold: Stability criteria of 80% of the datapoints within the 25% interval of the phase median value fulfilled.  
Accuracy: Percentage of correctly read words out of number of words read in one minute.

**Table 3**

*Within Phase Analysis*

Dependent variable	Improvement			
	Baseline	Intervention	Post	Total
Regular Words	0.75%	4.21%	0.32%	14.7%
Pseudowords	16.4%	19.3%	5.47%	30.5%
Irregular Words	15.5%	20.5%	8.48%	42.1%
Spelling	18.8%	12.6%	6.2%	35.8%
Control	21.0%	7.4%	1.8%	11.4%

Note: Calculation based on the full sample mean score values from the first to the last measurement occasion for each phase.  
 Total = Improvement from the last baseline measure to the last post-measure.

**Table 4**

*Within-Case Effects*

Student ID	effect size measure	Outcome Measures									
		Regular words		Pseudowords		Irregular words		Spelling		Peg Board	
		estimate	<i>p</i> value/CI	estimate	<i>p</i> value/CI	estimate	<i>p</i> value/CI	estimate	<i>p</i> value/CI	estimate	<i>p</i> value/CI
2	SMD	<b>2.15</b>	0.47-3.82	<b>2.18</b>	0.90-3.46	0.08	-0.14-0.31	0.94	-12-2.01	<b>2.87</b>	1.45-4.29
	Tau-U	<b>0.67</b>	0.04	<b>0.89<sup>#</sup></b>	0.01	0.30 <sup>#</sup>	0.33	0.49	0.13	<b>0.93</b>	0.00
3	SMD	0.50	-0.35-1.35	0.01	-0.19-0.22	<b>0.66</b>	0.11-1.20	0.54	-0.52-1.60	<b>1.89</b>	0.65-3.14
	Tau-U	0.36	0.29	0.15	0.64	<b>0.69<sup>#</sup></b>	0.03	0.24	0.46	<b>0.84</b>	0.01
4	SMD	<b>1.84</b>	0.57-3.12	<b>2.33</b>	0.85-3.81	<b>1.46</b>	0.14-2.77	<b>2.37</b>	0.88-3.86	<b>2.23</b>	0.77-3.69
	Tau-U	<b>0.97</b>	0.01	<b>0.91</b>	0.02	<b>0.81</b>	0.03	<b>0.85</b>	0.03	<b>0.87</b>	0.02
6	SMD	<b>1.65</b>	0.31-2.98	<b>3.44</b>	1.75-5.14	<b>1.62</b>	0.40-2.83	0.14	-0.02-3.1	<b>1.26</b>	0.11-2.41
	Tau-U	<b>0.83</b>	0.01	<b>1.0</b>	0.00	<b>0.80</b>	0.01	0.58	0.09	0.45 <sup>#</sup>	0.18
7	SMD	<b>1.39</b>	0.28-2.50	<b>2.10</b>	0.86-3.34	<b>1.81</b>	0.62-2.99	<b>1.43</b>	0.31-2.55	<b>2.48</b>	1.13-3.83
	Tau-U	<b>0.76</b>	0.02	<b>0.80</b>	0.01	0.58 <sup>#</sup>	0.07	<b>0.63</b>	0.05	<b>0.92</b>	0.00
8	SMD	<b>1.32</b>	0.22-2.43	<b>1.09</b>	0.02-2.15	<b>2.38</b>	1.08-3.68	<b>1.63</b>	0.48-2.78	0.50	-0.52-1.53
	Tau-U	<b>0.67</b>	0.04	<b>0.67<sup>#</sup></b>	0.04	<b>0.91</b>	0.00	<b>0.78</b>	0.01	0.28	0.39
9	SMD	<b>3.23</b>	1.81-4.65	<b>2.72</b>	1.41-4.03	<b>1.77</b>	0.64-2.90	<b>1.81</b>	0.68-2.95	<b>1.07</b>	0.03-2.10
	Tau-U	<b>0.83<sup>#</sup></b>	0.00	<b>0.99<sup>#</sup></b>	0.00	<b>0.77</b>	0.01	<b>0.83</b>	0.00	0.59	0.06
10	SMD	<b>1.89</b>	0.69-3.09	<b>4.27</b>	2.55-5.98	<b>1.81</b>	0.66-2.96	-0.43	-1.42-0.57	<b>1.90</b>	0.70-3.10
	Tau-U	<b>0.95</b>	0.00	<b>1.00<sup>#</sup></b>	0.00	<b>0.77</b>	0.01	<b>0.83</b>	0.00	<b>0.83</b>	0.00
11	SMD	<b>1.89</b>	0.69-3.09	<b>1.87</b>	0.68-3.07	<b>0.15</b>	0.08-0.22	-0.28	-1.28-0.72	<b>1.70</b>	0.54-2.87
	Tau-U	<b>0.87</b>	0.00	<b>0.85</b>	0.01	<b>0.95</b>	0.00	<b>0.75</b>	0.02	0.45	0.15
12	SMD	<b>1.50</b>	0.37-2.62	<b>1.91</b>	0.71-3.12	<b>1.30</b>	0.21-2.40	<b>2.45</b>	1.13-3.77	<b>1.68</b>	0.52-2.84
	Tau-U	0.80	0.01	<b>0.78</b>	0.01	<b>0.78</b>	0.01	<b>0.78</b>	0.01	<b>0.78</b>	0.01
13	SMD	<b>2.87</b>	1.45-4.29	<b>1.83</b>	0.64-3.01	<b>2.01</b>	0.79-3.23	0.90	-0.15-1.95	<b>1.71</b>	0.54-2.88
	Tau-U	<b>0.98</b>	0.00	<b>0.84</b>	0.01	<b>0.93</b>	0.00	0.47	0.14	<b>0.83</b>	0.00

Note:  
 SMD = Standardised mean difference  
 Significant results, *p*-value < 0.05 in bold.  
 Tau-U: *p* value (95%)  
 SMD: Confidence interval (CI) 95%  
<sup>#</sup> = Corrected for baseline trends.

**Table 5**  
*Between-Case Effects*

Outcome measure	BC-SMD	S.E.	CI	Baseline model	Intervention model
Regular Words	0.53	0.13	0.23–0.84	Level: fixed + random	Level: fixed
Pseudowords	0.67	0.20	0.22–0.94	Level: fixed + random	Level: fixed + random
Irregular Words	0.27	0.13	0.19–0.76	Level: fixed + random, Trend: random	Level: fixed + random Trend: random
Spelling	0.63	0.14	0.31–0.94	Level: fixed + random	Level: fixed
Control	0.31	0.11	0.06–0.57	Level fixed + random, Trend: random	Level: fixed

Note:  
BC-SMD = Between case–standard mean difference; CI: Confidence interval (95%); SE: standard error

**Table 6**  
*Generalisation Measure*

Case Number	2	3	4	6	7	8	9	10	11	12	13
Mean Scores											
STAS - Word Reading Composite											
Pre-test	37	22	70	73	54	28	43	36	60	11	66
Post-test	43		90	88	70	53	64	49	82	12	84
Difference ( <i>SD</i> )	6 (0.10)	n.a.	20 (0.35)	15 (0.23)	16 (0.29)	25(0.45)	21(0.38)	13(0.22)	22(0.34)	1(0.01)	18(0.32)
STAS - Pseudo Word Reading Composite											
Pre-test	9	16	36	35	26	20	21	26	45	6	29
Post-test	24	33	37	42	39	40	35	36	59	1	43
Difference ( <i>SD</i> )	15(0.55)	17(0.62)	18 0.05)	7(0.23)	13(0.48)	20 (0.74)	14(0.51)	10(0.37)	14(0.46)	-5(-017)	14(0.51)

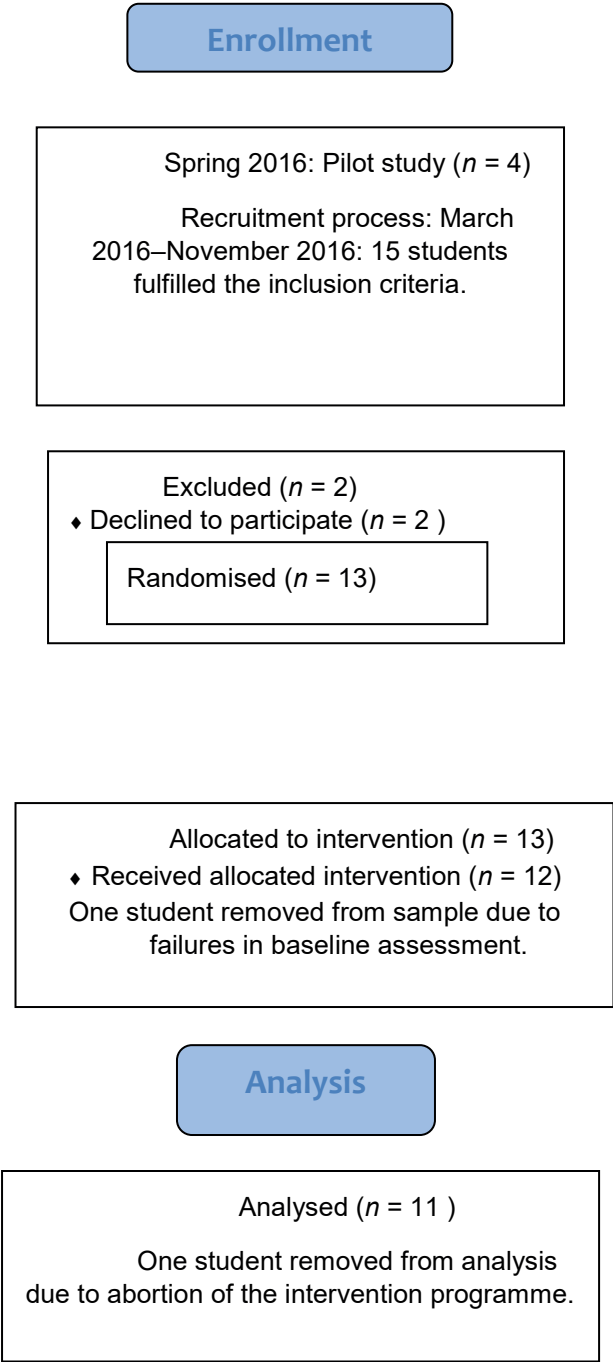
Note: All scores are raw scores.

*SD*: Improvement expressed in standard deviation units from the age mean standard.

n.a.= not available

**Figure 1**

*Flow of Participants*



**Figure 2**

*Pictographic Symbol System (PAS)*



*Figure 2.* PAS ‘spelling’ of the word ROSE [² ru:sə ] (spelled in Norwegian): Blue script for consonants and red script for vowels. Consonant cards contain indicators for voice, placement of tongue and acoustic cues. Vowel cards symbolise the shape and the opening of the mouth when pronouncing a vowel.

**Figure 3**  
*Measurement Occasions during Three Phases of Study*

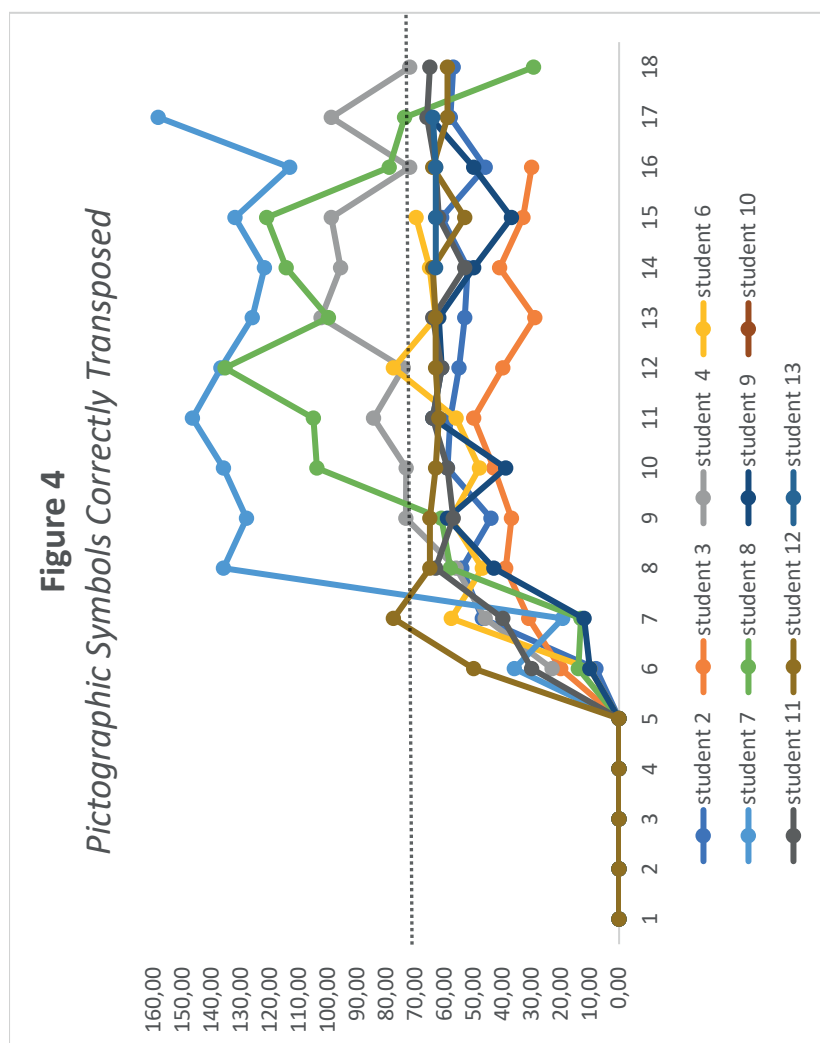
Student /Week no.	39	40	41	42	43	44	45	46	47	48	49	50	51	52	1	2	3	4	5	6	7	8	9	1	11	12	13	14	15	16	17	
1		X		X		X	X	X		X	X	X			X	X	X	X					X	X								
2		X				X	X	X		X	X	X			X	X	X	X		X	X				X	X						
3		X						X	X	X		X			X	X	X	X		X	X	X		X					X	X	X	X

Note: Red = Baseline, Black = Intervention, Blue = Post.

Empty fields: School holiday (marked) and scheduled stays (multiple probe design).

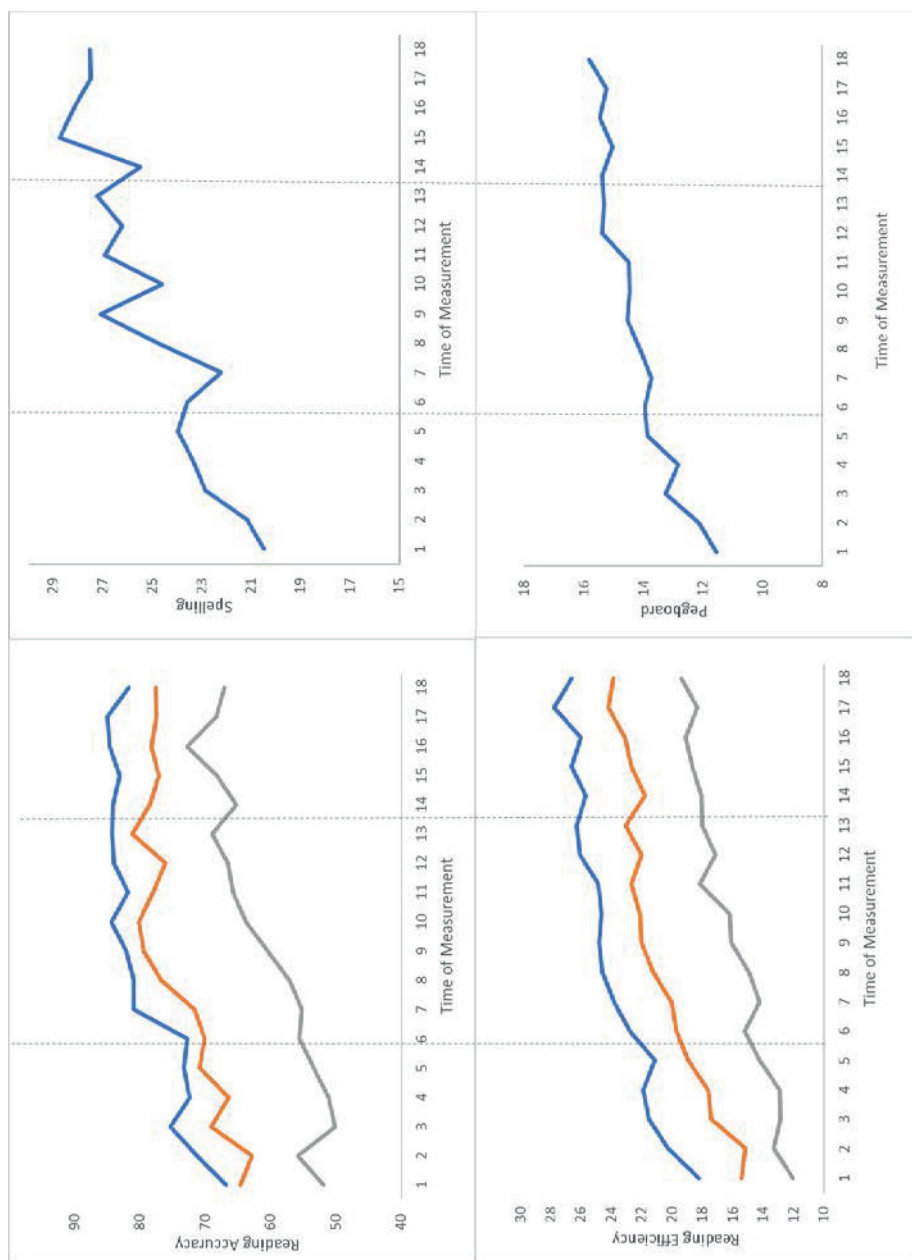
The study includes three replications across three participants (n = 9) and four single replications using the pattern of the first participant.





Note. Results from the transposition test. Pictographic symbols unknown prior to intervention onset (week 6). Horizontal line: Ceiling level (4 min.). Scoring above the line = extrapolated values (time bonus).

**Figure 5**  
*Full Sample Overview*



**Figure 5:** Graphics of mean scores across dependent variables. Dotted lines represent phase changes. Left panel: Blue line = regular words; orange line = pseudowords; grey line = irregular words.



## Appendiks





## Appendiks

Appendiks A: Godkjenning NSD

Appendiks B: Tillegg til godkjenning NSD

Appendiks C: Kort innføring til PAS systemet hentet fra veileder ( Kausrud,T., Bredtvet kompetansesenter 2005)

Appendiks N: Eksempel på transposisjonstest PAS

### Appendiks Studie 1

Appendiks D: Samtykke

Appendiks E: Informasjonsskjema

Appendiks F: Ord brukt i intervensjonsmaterialet (trente ord)

Appendiks G: Matchede utrente ord

Appendiks I: Intervensjonsprogram (Tilleggs-materiale til artikkel 1)

### Appendiks Studie 2

Appendiks I: Intervensjonsprogram (Tilleggs-materiale til artikkel 3)

Appendiks J: Eksempler på parallelle leseprøver og diktatprøver brukt til repetert kartlegging

Appendiks J: Grafikk over individuelle resultater (Tilleggs-materiale til artikkel 3)

Appendiks L: Samtykke

Appendiks M: Skjema for bakgrunnsinformasjon

Appendiks N: Spørreskjema til testleder

Appendiks O: Vurderingskjema til lærere



Harald Hårfagres gate 29  
N-5007 Bergen  
Norway  
Tel: +47-55 58 21 17  
Fax: +47-55 58 96 50  
nsd@nsd.uib.no  
www.nsd.uib.no  
Org.nr. 985 321 884

Anne Cathrine Thurmann-Moe  
Statped sørøst  
Postboks 4416 Nydalen  
0403 OSLO

Vår dato: 18.05.2015

Vår ref: 42954 / 3 / AGL

Deres dato:

Deres ref:

## TILBAKEMELDING PÅ MELDING OM BEHANDLING AV PERSONOPPLYSNINGER

Vi viser til melding om behandling av personopplysninger, mottatt 26.03.2015. All nødvendig informasjon om prosjektet forelå i sin helhet 13.05.2015. Meldingen gjelder prosjektet:

**42954**                      *Artikualtorisk bevisstgjøring av fonemene og lydmessig feedback på fonemnivå.*

*Behandlingsansvarlig*      *Statped, ved institusjonens øverste leder*

*Daglig ansvarlig*              *Anne Cathrine Thurmann-Moe*

Personvernombudet har vurdert prosjektet, og finner at behandlingen av personopplysninger vil være regulert av § 7-27 i personopplysningsforskriften. Personvernombudet tilrår at prosjektet gjennomføres.

Personvernombudets tilråding forutsetter at prosjektet gjennomføres i tråd med opplysningene gitt i meldeskjemaet, korrespondanse med ombudet, ombudets kommentarer samt personopplysningsloven og helseregisterloven med forskrifter. Behandlingen av personopplysninger kan settes i gang.

Det gjøres oppmerksom på at det skal gis ny melding dersom behandlingen endres i forhold til de opplysninger som ligger til grunn for personvernombudets vurdering. Endringsmeldinger gis via et eget skjema, <http://www.nsd.uib.no/personvern/meldeplikt/skjema.html>. Det skal også gis melding etter tre år dersom prosjektet fortsatt pågår. Meldinger skal skje skriftlig til ombudet.

Personvernombudet har lagt ut opplysninger om prosjektet i en offentlig database, <http://pvo.nsd.no/prosjekt>.

Personvernombudet vil ved prosjektets avslutning, 01.03.2019, rette en henvendelse angående status for behandlingen av personopplysninger.

Vennlig hilsen

Katrine Utaaker Segadal

Audun Løvlie

Kontaktperson: Audun Løvlie tlf: 55 58 23 07

Vedlegg: Prosjektvurdering

*Dokumentet er elektronisk produsert og godkjent ved NSDs rutiner for elektronisk godkjenning.*

*Avdelingskontorer / District Offices:*

*OSLO:* NSD. Universitetet i Oslo, Postboks 1055 Blindern, 0316 Oslo. Tel: +47-22 85 52 11. [nsd@uio.no](mailto:nsd@uio.no)

*TRONDHEIM:* NSD. Norges teknisk-naturvitenskapelige universitet, 7491 Trondheim. Tel: +47-73 59 19 07. [kyrre.svarva@svt.ntnu.no](mailto:kyrre.svarva@svt.ntnu.no)

*TROMSØ:* NSD. SVF, Universitetet i Tromsø, 9037 Tromsø. Tel: +47-77 64 43 36. [nsdmaa@sv.uit.no](mailto:nsdmaa@sv.uit.no)



Formålet er rettet mot elever med språk, lese og skrivevansker som er henvist til Statped og/eller Pedagogisk psykologisk tjeneste og omhandler systematisk utprøving av tiltak for elever med alvorlig grad av lese, skrive og språkvansker. Hensikten med prosjektet er å få sikrere kunnskap om hvordan vi kan oppnå større grad av treffsikkerhet når det gjelder funksjonelle tiltak for denne elevgruppen.

Utvalget informeres skriftlig om prosjektet og samtykker til deltakelse. Informasjonsskriv mottatt 13.05.2015 er godt utformet.

Foreldre samtykker for sine barn. Selv om foreldre/foresatte samtykker til barnets deltakelse, minner vi om at barnet også må gi sin aksept til deltakelse. Barnet bør få tilpasset informasjon om prosjektet, og det må sørges for at de forstår at deltakelse er frivillig og at de når som helst kan trekke seg dersom de ønsker det. Dette kan være vanskelig å formidle, da barn ofte er mer autoritetstro enn voksne. Frivillighetsaspektet må derfor særlig vektlegges i forhold til barn, og spesielt når forskningen foregår på eller i tilknytning til en organisasjon som barnet står i et avhengighetsforhold til, som for eksempel skole. Forespørselen må derfor alltid rettes på en slik måte at de forespurte ikke opplever press om å delta, gjerne ved å understreke at det ikke vil påvirke forholdet til skolen hvorvidt de ønsker å være med i studien eller ikke.

Det behandles sensitive personopplysninger om helseforhold.

Personvernombudet legger til grunn at forsker etterfølger Statped sine interne rutiner for datasikkerhet. Dersom personopplysninger skal lagres på mobile enheter, bør opplysningene krypteres tilstrekkelig.

Det tas høyde for at det skal anvendes ekstern transkriberingstjeneste som databehandler for prosjektet, jf telefonsamtale 13.05.2015. Statped skal inngå skriftlig avtale med denne tjenesteyteren om hvordan personopplysninger skal behandles, jf. personopplysningsloven § 15. For råd om hva databehandleravtalen bør inneholde, se Datatilsynets veileder: <http://www.datatilsynet.no/Sikkerhet-internkontroll/Databehandleravtale/>.

Forventet prosjektslutt er 01.03.2019. Ifølge prosjektmeldingen skal innsamlede opplysninger da oppbevares med personidentifikasjon i utvalgets journaler (Fabris) ved Statped etter gjeldende retningslinjer. Lydopptakk slettes etter transkripsjon jf telefonsamtale 13.05.2015.

## NSD sin vurdering

**Prosjektittel**

42954. Artikulatorisk bevisstgjøring av fonemene og lydmessig feedback på fonemnivå

**Referansenummer**

112509

**Registrert**

18.10.2019 av Anne Cathrine Thurmann-Moe - annecthu@uio.no

**Behandlingsansvarlig institusjon**

Universitetet i Oslo / Det utdanningsvitenskapelige fakultet / Institutt for spesialpedagogikk

**Prosjektansvarlig (vitenskapelig ansatt/veileder eller stipendiat)**

Anne Cathrine Thurmann-Moe, anne-cathrine.thurmann-moe@statped.no, tlf: 92610925

**Felles behandlingsansvarlige institusjoner**

Statped / Statped sørøst

**Type prosjekt**

Forskerprosjekt

**Prosjektperiode**

15.09.2015 - 30.06.2020

**Status**

17.07.2020 - Avsluttet

**Vurdering (1)****09.12.2019 - Vurdert**

## BAKGRUNN

Behandlingen av personopplysninger ble opprinnelig meldt inn til NSD 26.03.2015 (NSD sin ref: 42954) og vurdert under personopplysningsloven som var gjeldende på det tidspunktet.

18.10.2019 meldte prosjektleder inn en endring av prosjektet. Endringen består i en utsettelse av prosjektslutt til 30.06.2020. Opprinnelig prosjektslutt var 1.9.2019. NSD vil understreke at ved eventuell forlengelse utover ett år fra opprinnelig prosjektslutt må prosjektleder finne en måte å gi informasjon på eller datamaterialet anonymiseres.

Det er vår vurdering at behandlingen av personopplysninger i prosjektet vil være i samsvar med personvernlovgivningen så fremt den gjennomføres i tråd med det som er dokumentert i meldeskjemaet med vedlegg den 09.12.2019, samt i meldingsdialogen mellom innmelder og NSD. Behandlingen kan fortsette.

## MELD VESENTLIGE ENDRINGER

Dersom det skjer vesentlige endringer i behandlingen av personopplysninger, kan det være nødvendig å melde dette til NSD ved å oppdatere meldeskjemaet. Før du melder inn en endring, oppfordrer vi deg til å lese om hvilke type endringer det er nødvendig å melde:

[https://nsd.no/personvernombud/meld\\_prosjekt/meld\\_endringer.html](https://nsd.no/personvernombud/meld_prosjekt/meld_endringer.html)

Du må vente på svar fra NSD før endringen gjennomføres.

## TYPE OPPLYSNINGER OG VARIGHET

Prosjektet vil behandle særlige kategorier av personopplysninger om helseforhold og alminnelige kategorier av personopplysninger frem til 30.06.2020.

## LOVLIG GRUNNLAG

Prosjektet har innhentet samtykke fra de registrerte til behandlingen av personopplysninger.

NSD vurderer at informasjonen i hovedsak oppfyller lovens krav til form og innhold, jf. personvernforordningen art. 12.1 og art. 14. Det innhentede samtykket vurderes som frivillig, spesifikt, informert, dokumenterbart, uttrykkelig og utvetydig jf. personvernforordningen art. 4.11.

Ettersom informasjonen som er gitt til deltakerne er gitt under gammelt lovverk, oppfylles nødvendigvis ikke nye krav i personvernforordningen, som trådte i kraft 20.7.2018. Det er dermed ikke informert om

- kontaktopplysninger til personvernombudet, jf. art. 14-1 b
- rettslig grunnlag for behandlingen, jf. art. 14-1 c
- de registrertes rettigheter, jf. art. 14-2 c
- retten til å klage til Datatilsynet, jf. art. 14-2 e

Informasjon som er gitt vurderes likevel å gjøre utvalget i stand til å benytte seg av sine rettigheter etter personvernforordningen art. 15–21.



Lovlig grunnlag for behandlingen vil dermed være den registrertes uttrykkelige samtykke, jf. personvernforordningen art. 6 nr. 1 bokstav a, jf. art. 9 nr. 2 bokstav a, jf. personopplysningsloven § 10, jf. § 9 (2).

#### PERSONVERNPRINSIPPER

NSD vurderer at den planlagte behandlingen av personopplysninger vil følge prinsippene i personvernforordningen om:

- lovlighet, rettferdighet og åpenhet (art. 5.1 a), ved at de registrerte får tilfredsstillende informasjon om og samtykker til behandlingen
- formålsbegrensning (art. 5.1 b), ved at personopplysninger samles inn for spesifikke, uttrykkelig angitte og berettigede formål, og ikke behandles til nye, uforenlige formål
- dataminimering (art. 5.1 c), ved at det kun behandles opplysninger som er adekvate, relevante og nødvendige for formålet med prosjektet
- lagringsbegrensning (art. 5.1 e), ved at personopplysningene ikke lagres lengre enn nødvendig for å oppfylle formålet

#### DE REGISTRERTES RETTIGHETER

Så lenge de registrerte kan identifiseres i datamaterialet vil de ha følgende rettigheter: åpenhet (art. 12), informasjon (art. 13), innsyn (art. 15), retting (art. 16), sletting (art. 17), begrensning (art. 18), underretning (art. 19), dataportabilitet (art. 20).

Statped er felles behandlingsansvarlig institusjon. NSD legger til grunn at behandlingen oppfyller kravene til felles behandlingsansvar, jf. personvernforordningen art. 26.

NSD vurderer at informasjonen om behandlingen som de registrerte vil motta oppfyller lovens krav til form og innhold, jf. art. 12.1 og art. 13.

Vi minner om at hvis en registrert tar kontakt om sine rettigheter, har behandlingsansvarlig institusjon plikt til å svare innen en måned.

#### FØLG DIN INSTITUSJONS RETNINGSLINJER

NSD legger til grunn at behandlingen oppfyller kravene i personvernforordningen om riktighet (art. 5.1 d), integritet og konfidensialitet (art. 5.1 f) og sikkerhet (art. 32).

For å forsikre dere om at kravene oppfylles, må dere følge interne retningslinjer og/eller rådføre dere med behandlingsansvarlig institusjon.

#### OPPFØLGING AV PROSJEKTET

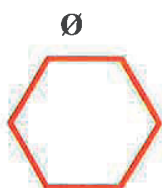
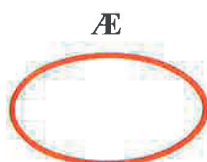
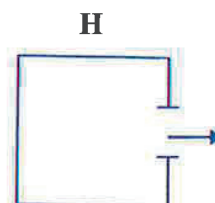
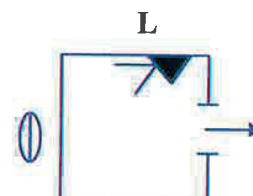
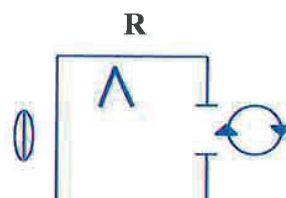
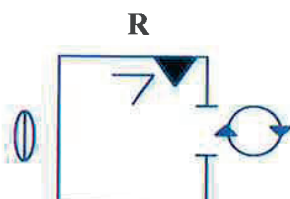
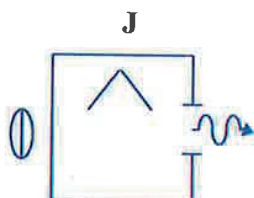
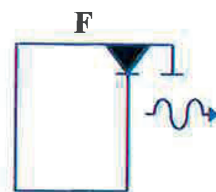
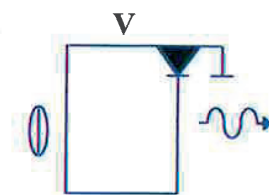
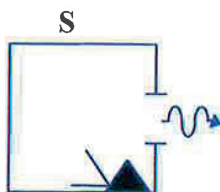
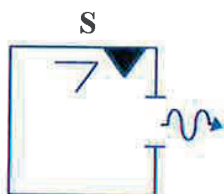
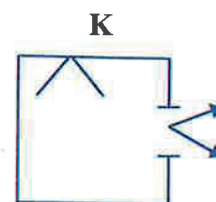
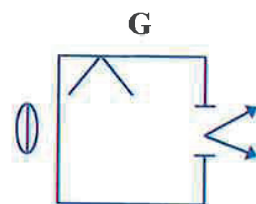
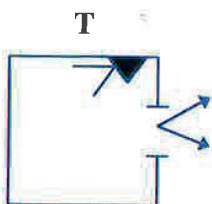
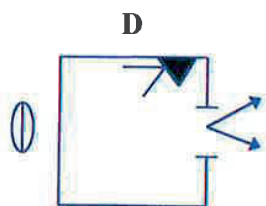
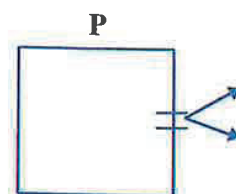
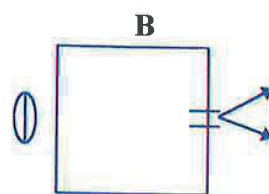
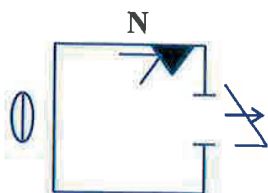
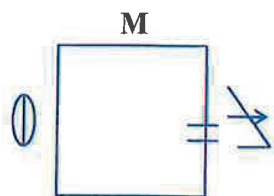
NSD vil følge opp ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet.

Lykke til med prosjektet!

Kontaktperson hos NSD: Kajsa Amundsen  
Tlf. Personverntjenester: 55 58 21 17 (tast 1)

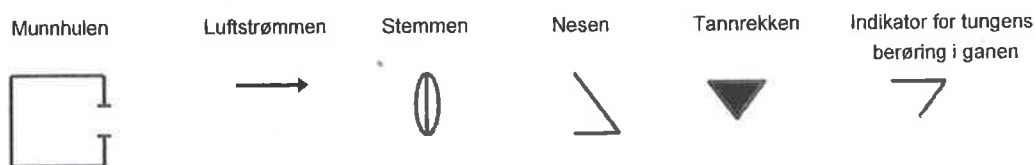
# PAS

## Piktografisk-artikulatoriske symboler



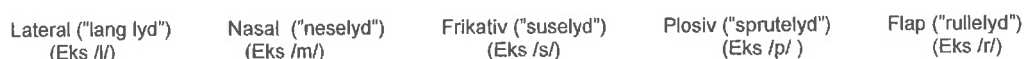
## Beskrivelse av PAS-systemet

De piktografisk-artikulatoriske symbolene (forkortet til PAS) viser på en skjematisk måte hvordan bokstavlydene uttales og høres ut. Grunnlaget for uttalen er normalisert østnorsk. Symbolene er enkle, geometriske figurer, der bare trekk som er viktig for den enkelte lyd, vises. Delelementene tilknyttet uttalen, artikulasjonen, av **konsonantene** vises i figur 1.



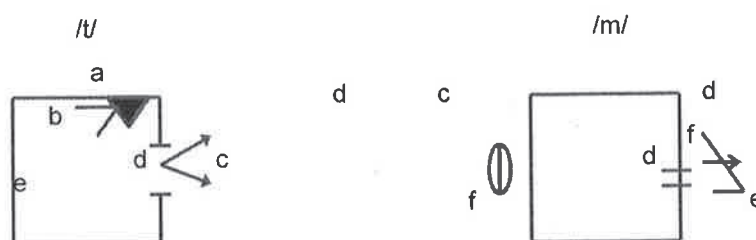
Figur 1. Delelementene i PAS

Hvordan bokstavlydene høres ut, det vil si deres akustiske trekk, framstilles gjennom ulik utforming av en pil. Se figur 2.



Figur 2. PAS-systemets framstilling av akustiske egenskaper hos lyder

Figur 3 viser symbolene for to av konsonantene.



Figur 3. PAS-symboler for /l/ og /m/.

Ved dannelse av /l/ kommer luftstrømmen ut som en eksplosjon, vist ved de to pilene som peker i hver sin retning (c), gjennom åpen munn (d). Tungens berøringspunkt inne i munnhulen er bak tennene, vist ved indikatoren (b) mot den sorte trekanten (a) som symboliserer den øvre tannrekken. Ved dannelsen av /m/ er munnen lukket (d),

# KATEGORISERING AV BOKSTAVLYD

	FORAN I MUNNEN		MIDT I MUNNEN		BAK I MUNNEN	
	STEMT	USTEMT	STEMT	USTEMT	STEMT	USTEMT
EKSPLOSJONS- LYD						
SUSELYD						
NASALE LYD						
RULLE-LYD						
LANGE LYD						

			Antall lyder korrekt	Ord korrekt
1		al	2	1
2		<del>em</del> em	2	1
3		el	2	1
4		ma	2	1
5		<del>ir</del> ir	2	1
6		mi	2	1
7		sø	2	1
8		ek	2	1
9		fø	2	1
10		fa	2	1
11		ø	1	
12		u	1	
13		un	2	
14		aps	3	1
15		<del>eps</del> eps	3	1
16		uks	3	1

30 13





## Appendiks studie 1







## **Forespørsel om deltakelse i forskningsprosjektet «Artikulatorisk støtte i lese- og skriveopplæringen»**

### **Bakgrunn og formål**

Statped sørøst, avdeling for språk og tale, arbeider for tiden med et prosjekt der det prøves ut tiltak i den begynnende lese og skriveopplæringen. Hensikten med prosjektet er å undersøke hvorvidt undervisning med økt fokus på bokstavlydene kan føre til bedret lese og staveferdighet hos elever på 1. trinn. Det er gjort en avtale om samarbeid med xxxx kommune om dette og et utvalg skoler i kommunen vil våren 2018 prøve et nytt undervisningsopplegg som del av den ordinære oppfølgingen i lesing. Undervisningen vil følge et oppsatt program og deltakende lærere vil få veiledning.

**Som ledd i forskningsprosjektet vil ca. 200 elever bli trukket ut til å delta enten i en eksperimentgruppe eller en kontrollgruppe. Fordelingen i eksperiment og kontrollgruppe foretas ved tilfeldig utvalg.**

### **Hva innebærer deltakelse i studien?**

Alle elevene som blir trukket ut vil bli kartlagt med språk og leseprøver en gang før tiltaket starter og en gang etter at tiltaket er ferdig. Hver kartlegging vil ta ca 30 minutter og utføres enten av prosjektleder eller av studenter ved Universitet i Oslo som kommer til skolen. Det vil bli gjort lydopptak av kartleggingen og undervisningstimene.

### **Hva skjer med informasjonen som samles inn?**

Alle personopplysninger vil bli behandlet konfidensielt og vil kun være tilgjengelig for prosjektleder og veiledere i prosjektet. Data vil bli anonymisert og elever vil ikke kunne gjenkjennes ved publisering av materialet. Prosjektet skal etter planen avsluttes innen 1. september 2019. Oppbevaringen av opplysninger vil følge vanlige forskningsetiske retningslinjer og slettes ved prosjektets slutt.

### **Frivillig deltakelse**

Det er frivillig å delta i studien, og du kan når som helst trekke ditt samtykke uten å oppgi noen grunn. Dersom du trekker deg, vil alle opplysninger om deg bli anonymisert. Dersom du har spørsmål til studien, ta kontakt med prosjektleder Anne Cathrine Thurmann-Moe på tlf. xxxxxxeller på e-post [anne-cathrine.thurmann-moe@statped.no](mailto:anne-cathrine.thurmann-moe@statped.no).

*Studien inngår som ledd i prosjektlederens dr. gradprosjekt og veiledes fra Universitetet i Oslo.*

*Studien er meldt til Personvernombudet for forskning, Norsk samfunnsvitenskapelig datatjeneste AS.*

**Samtykke til deltakelse i studien**

Jeg har mottatt informasjon om prosjektet «Artikulatorisk støtte i lese- og skriveopplæringen» og gir med dette tillatelse til at mitt barn deltar i forskningsprosjektet dersom det blir trukket ut.

Barnets navn: .....født:..... skole: ..... klasse:.....

.....

Underskrift fra foresatte

**Leveres kontaktlærer.**

Oslo april 2018

**Kjære foresatte.**

Takk for at dere har gitt samtykke til deltakelse i forskningsprosjektet om artikulatorkisk støtte i leseopplæringen på 1. trinn. Det har vært av stor betydning at så mange har takket ja til å delta.

Som ledd i prosjektet ønsker vi også å samle inn noe bakgrunnsinformasjon om elevene som deltar og ber derfor dere om å fylle ut vedlagte skjema. Når skjema er fylt ut legges det tilbake i konvolutten. Denne limes igjen og leveres til kontaktlærer.

Prosjektet er meldt og godkjent av personvernombudet for forskning, Norsk Senter for forskningsdata AS (NSD) og følger vanlige forskningsetiske regler. Alle opplysninger behandles konfidensielt og skolen har ikke tilgang til informasjon gitt i skjema. Skjema er anonymt og knyttet til eleven gjennom en koblingsnøkkel som kun er kjent for prosjektleder.

**Dersom du har spørsmål til dette skjemaet eller til studien generelt, ta kontakt med prosjektleder Anne Cathrine Thurmann-Moe tlf. xxxxxeller via mail til [anne-cathrine.thurmann-moe@statped.no](mailto:anne-cathrine.thurmann-moe@statped.no)**

**SKJEMA FOR BAKGRUNNSINFORMASJON**

**1. Hvem er det som besvarer dette spørreskjemaet?**

- Mor  
 Far  
 Andre; Hvis ja, hvem: .....

**2. Hva er barnets førstespråk/morsmål?.....**

**3. Hvilket språk snakker barnet mest hjemme (maks to språk)?**

.....

**4. Kryss av for mors høyest fullførte utdanningsnivå**

- Ungdomsskole  
 Videregående skole/fagbrev  
 Bachelor  
 Master  
 Høyere (PhD o.l.)

**5. Kryss av for fars høyest fullførte utdanningsnivå**

- Ungdomsskole  
 Videregående skole/fagbrev  
 Bachelor  
 Master  
 Høyere (PhD o.l.)

**6. Har du /dere hatt bekymring knyttet til barnets utvikling?**

- Ja  
 Nei

Hvis ja, beskriv nærmere:.....

**7. Ble det registrert problemer med uttale av ord i førskolealderen?**

- Ja  
 Nei

Hvis ja, beskriv nærmere:.....

**8. Har barnet gått i barnehage?**

Ja

Nei

Hvis ja, beskriv hvor lenge:.....

**9. Har barnet byttet lærer i løpet av skolegangen ?**

Ja

Nei

**10. Har barnet byttet skole i løpet av skolegangen?**

Ja

Nei

**11. Har noen i barnets familie vansker knyttet til muntlig språk, lesing eller skriving?**

Ja

Nei

Hvis ja beskriv nærmere.....



## ORDLISTE - ord brukt i intervensjonsmaterialet - «trente ord»

- is
- ku
- to
- lue
- mus
- pus
- bål
- båt
- lås
- hus
- rev
- sol
- bie
- bær
- fem
- løk
- fot
- kåpe
- hane
- rose
- pære
- høne
- måne
- kake
- fire
- vase
- esel
- kanin
- banan
- tomat
- potet
- vask
- tre
- brus
- krus
- elg
- gris
- kniv
- kost
- brev
- ost
- saft
- okse
- sitron
- eple
- lampe
- ugle
- drage
- flue
- kråke
- vindu
- pølse
- krone
- støvel
- prins
- ekorn
- spurv
- flaske
- paraply
- undulat
- elefant
- hamburger

ORDLISTE- Ord matchet til ordene som ble brukt i intervusjonen - «utrente ord»

|

- do
- by
- ape
- mat
- gås
- bok
- lam
- hår
- duk
- nål
- gul
- mor
- øre
- far
- dag
- baby
- pute
- hale
- hage
- løve
- gave
- dame
- foto
- lege
- pose
- møte
- lage
- bade
- kose
- valp
- veps
- ulv
- står
- luft
- seks
- telt
- stol
- fant
- fast
- stor
- blir
- flere
- følte
- skole
- bruke
- store
- viste
- borte
- lenke
- beste
- burde
- prøve
- gamle
- stort
- blant
- grøft
- frukt
- sparke
- smakte
- trampoline
- fotografene



### VEILEDER for undervisning i små grupper fire timer pr uke

#### Introduksjon

Mange elever med lese, skrive og språkvansker har store problemer med å skille språklydene fra hverandre bare ved å lytte. De får dermed også problemer med presis «lagring» av bokstavlydene i minnet - de får upresise fonologiske representasjoner. Dette gjør at de har vanskelig for å «komme på» riktig lyd når de møter på bokstavtegnene i tekst. Det er ofte spesielt vanskelig for dem å skille mellom lyder som er relativt like, slik som for eksempel p/b t/d, n/m og g/k. Dette kan gjøre at lesingen blir unøyaktig og lite flytende. I dette prosjektet vil vi prøve ut om elevene ved å lære å definere bokstavlydene gjennom hvordan de uttales kan bli sikrere på koblingene mellom lyd og bokstav og dermed også mer nøyaktige og effektive når de leser. Treningen innebærer en restart av leseopplæringen med vekt på oppgaver som øker elevenes fonologiske bevissthet. Men i stedet for å «lytte ut lydene» slik man ofte øver på i tradisjonelle fonologisk bevissthets-oppgaver, har treningen her fokus på hva man konkret gjør med de forskjellige delene av tale-apparatet når man sier en lyd. Hensikten er å gi elevene et «tilleggsverktøy» for å huske koblingene mellom bokstavtegnene og bokstavlydene.

#### PAS systemet (Piktografisk Artikulatorisk System)

PAS er utviklet av Torhild Kausrud og har vært brukt på Bredtvet Kompetansesenter (nå Statped sør-øst) siden 2005. Logikken bak dette systemet er at språklydene kan kategoriseres etter hvordan de produseres artikulatorisk og at dette kan gi bedre støtte når det gjelder å skille bokstavlydene fra hverandre, særlig for de elevene som har fonologiske vansker.

PAS består av grunnelementer som settes sammen i piktografiske symboler som angir hvordan tale-apparatet fungerer når man produserer forskjellige bokstavlyder. Hvert bokstavkort i PAS inneholder informasjon om hvordan lyden skal lages med munn og stemme. Hvor setter man tungen? Hvordan kommer luften ut når man sier lyden? Er lyden stemt eller ikke? Symbolene på kortet er på en måte oppskriften på lyden. Elevene skal gjennom disse ukene med trening lære å tolke symbolene, bli bevisste på hva de gjør med tunge, stemme og luftstrøm når de sier lyden og også lære å utnytte dette som støtte ved avkoding og staving.

**NB! Det er ikke meningen at eleven skal lære seg å huske symbolene som visuelle enheter. Fokus for all trening med kortene skal være å kunne følge «oppskriften på lyden» nøyaktig slik at riktig lyd og ord kommer fram.**

#### Lærers rolle

Veiledningen inneholder et fast program med aktiviteter for hver uke og det er viktig at dette følges. Progresjonen er innebygd i programmet gjennom at det introduseres nye aktiviteter etter hvert. Det er utarbeidet en aktivitetsplan for hver uke der det også er angitt hvor lang tid

(dette er veiledende) som bør brukes på hver aktivitet for å ivareta progresjonen. Det tas utgangspunkt i 45 minutter økter.

De første to ukene legges det mest vekt på å lære å fokusere på artikulatoriske egenskaper ved lydene og tolke hvordan disse er symbolisert på kortene. De siste ukene er det mest fokus på å overføre ferdigheter i avkoding/staving med PAS til avkoding og staving med vanlige bokstaver. Lærer trenger ikke å være ekspert på å tolke symbolene, men kan gjerne innta rollen med å utforske sammen med elevene hvordan lydenes egenskaper er symbolisert på kortene. Vedlagt er også en oversikt over PAS symboler og bokstaver som lærer kan bruke som støtte. (Erfaringsmessig er tegnene lettere å lære for elevene enn for oss voksne !)

For øvrig er kortene velegnet til å bevisstgjøre elevene på at ord består av enkelttyder, til å illustrere forskjellen mellom vokaler og konsonanter og de kan brukes til å konkretisere hvordan ordene er delt inn i stavelser. Lærers rolle er ellers å lede gruppeaktivitetene og sørge for skifte av aktiviteter i løpet av timen. Det er laget et loggskjema som skal fylles ut etter hver time der lærer «kvitterer» for hvilke aktiviteter som har vært gjennomført i timen.

Aktivitet / uke	Uke  1	Uke 2	Uke 3	Uke 4	Uke5
Aktivitet 1	X				
Aktivitet 2	X	X	X		
Aktivitet 3	X	X	X		
Aktivitet 4		X	X	X	X
Aktivitet 5			X	X	X
Aktivitet 6				X	X

## Beskrivelse av aktiviteter

### Aktivitet 1 - Hvordan ser lydene ut og føles i munnen? (ca 10 min)

Lærer har en bunke med billedkort eller andre objekter og et speil (her kan hva som helst brukes, hvis ingenting annet er tilgjengelig bruk illustrasjonskort fra utlevert materiale).

Elevene trekker kort etter tur fra bunken og skal si første lyd i ordet på bildet mens de ser på sin egen munn i speilet. Lærer underviser og modellerer hvordan man setter tungen, om stemmen er på eller ikke og hvordan luften kommer ut når man sier lyden.

### Aktivitet 2 - Lære PAS symbolene (ca 25 min)

Elevene introduseres til denne aktiviteten som at de skal lære å lese og skrive med «hemmelige tegn». Fokus er å lære grunnelementene:

- Hvordan kommer luften ut ?

- Er stemmen av eller på?
- Hvor plasseres tungen?

PAS kortene (gul bakside) legges i en bunke med baksiden opp.

Elevene trekker kort etter tur og skal prøve å produsere lyden ved å fokusere på elementene på kortet - «oppskriften på lyden». Lærer viser og modellerer først ved å trekke et kort og si høyt hvordan hun tolker symbolet ved å fokusere på om det er stemme på, hvor tungen skal være, og hvordan luften kommer ut. (bruk gjerne begreper som suselyd, neselyd, rullelyd og eksplosjonslyd)

Start med lydene P, I, H og S, utvid med M, O og L i løpet av de første dagene og utvid så etter hvert til alle lyder er gjennomgått. Man kan i denne aktiviteten gjerne la elevene få prøve å stave tolydsord som IS, SOL, HI osv under veiledning.

### **Aktivitet 3 - Matching (ca 10 min)**

Billedkort (hva som helst av billedmateriale eller objekter kan brukes, gjerne flere alternative bilder for hver lyd) legges på bordet med bildesiden opp.

Utvalgte PAS kort m gul bakside (læreren velger ut lyder) legges med baksiden opp og ett kort trekkes av elevene etter tur. Eleven skal gjennom å tolke symbolet uttale lyden på PAS-kortet og matche det til riktig bilde utfra første lyd. Aktivitet fortsetter til alle elevene har fått prøve seg.

Etterpå kan man ta en ny runde, og eventuelt variere med å matche til siste lyd!

*Denne aktiviteten kan tilpasses på ulike måter, og det går også an å matche PAS-kortet til et bilde av et ord med lyden i seg, altså ikke bare først eller sist i ordet.*

### **Aktivitet 4 - Lese og skrive med PAS kort (ca 15 min)**

Elevene lukker øynene og lærer legger opp et ord med PAS brikkene (gul bakside)

Elevene leser etter tur og oppgaven gjentas til alle har fått prøve seg.

Lærer tilpasser vanskegrad etter elevenes nivå (bruk vedlagte ordliste)

**NB: Bør være kun lydrette ord, unngå ord med stumme lyder, diftonger, kj/skj/sj og o/å forvekslinger. Dobbel konsonant kan eventuelt forsøkes etter hvert.**

*Som variasjon kan elevene etter tur få lage oppgavene, skrive med kortene og lærer eller en annen elev skal lese.*

### **Aktivitet 5 - Lese PAS ord (ca 10-15 min)**

Materiale: PAS ord med illustrasjoner som «fasit» på baksiden.

Kortene legges i bunke med PAS tegnene opp og elevene får trekke etter tur.

De skal lese ordene lyd for lyd og trekke sammen, etterpå snur man kortet og ser på fasiten om det stemmer. Kortene er fargekodet etter vanskegrad i blå, grønn og rød, der blå er lettest og rød er vanskeligst.

### **Aktivitet 6 - individuelle oppgaver (ca10-15 min)**

Materiale til denne oppgaven er ark med oppgaver der elevene skal arbeide selvstendig.

Oppgavene består av to kolonner med ord, en kolonne med ord skrevet med PAS og en kolonne der elevene skal oversette dette til vanlig skrift, enten ved hjelp av silhuetter som angir bokstavenes form i bokstavhuset som støtte (enklest) eller ved at det bare er angitt med streker hvor mange bokstaver ordet skal ha.

*Materialet er fargekodet etter vanskegrad og lærer velger ut fra kjennskap til eleven hvilke oppgaver som passer best. Blå er lettest, grønn er middels og rød er vanskeligst.*

*(Silhuettoppgavene er kun tilgjengelig for de enkleste ordene (blå)).*



## Appendiks studie 2



### Veileder for intervensjon – 8 ukers individuell trening (4 timer pr uke)

#### Introduksjon

Hensikten med treningen er å vurdere om elever med avkodingsvansker kan ha nytte av å bli mer bevisste på hvordan man formulerer bokstavlydene i munnen.

Mange elever med lese, skrive og språkvansker har store problemer med å skille språklydene fra hverandre bare ved å lytte. Dette gjør at de har «vanskelig for å komme på» riktig lyd når de møter på bokstavtegnene i tekst. Det er ofte spesielt vanskelig for dem å skille mellom lyder som er relativt like, slik som for eksempel p/b t/d, n/m, g/k. Vi ønsker gjennom treningen å få vurdert hvorvidt det å definere lydene på en annen måte, altså gjennom hvordan de produseres artikulatorisk, det vil si hva man gjør med de forskjellige delene av taleorganet når man sier en lyd, kan hjelpe elevene til å få sikrere og mer automatisert bokstavkunnskap, og gjennom det også mer effektive og automatiserte avkodingsferdigheter.

Språklydene kan kategoriseres etter hvordan de produseres artikulatorisk. Hvor setter man tunga? Hvordan er det med luftstrøm? Er lyden stemt eller ikke? Grunnsymbolene i PAS angir hvordan artikulasjonsapparatet fungerer når man sier forskjellige bokstavlyder.

Hvert bokstavkort inneholder informasjon om hvordan lyden skal lages, symbolene på kortet er på en måte oppskriften på lyden. Hensikten med treningen er at eleven skal lære å tolke symbolene, bli bevisst på hva han gjør med tunge, stemme og luftstrøm når han sier lyden og lære å utnytte dette som støtte ved lesing og skriving. **Det er ikke meningen at eleven skal lære seg å huske symbolene som visuelle enheter.** Fokus for treningen skal være å kunne følge «oppskriften på lyden» på kortene nøyaktig slik at riktig lyd kommer fram.

Hver treningsøkt er på 40 minutter og består av et sett med aktiviteter, hver på 5-10 minutter.

#### Gang i treningsøkt (40 minutter)

##### 1. Gjennomgang av grunnsymbolene i PAS

Elev og lærer ser på grunnsymbolene sammen og repeterer hva de betyr og hvilke deler av taleapparatet de henviser til. Elev og lærer kan i denne økten også sortere kortene etter kategorier: Alle stemte lyder, alle nasale, alle med «vislelyd» o.s.v. Lærer viser/modellerer, speil kan gjerne benyttes.

## **2. Matche bokstaver til PAS symboler**

Elev og lærer sitter mot hverandre med bord imellom.

Kortene legges med PAS- symbolsiden opp i en bunke. Her må lærer på forhånd velge ut hvilke lyder og symboler som skal arbeides med i løpet av timen. Eleven skal trekke et kort, analysere symbolene og si lyden. Fasiten er på baksiden av kortet, slik at lærer kan se, og eventuelt også støtte og hjelpe eleven ved behov, men eleven har ikke lov til å se på baksiden av kortet før lyden er sagt.

Man kan også legge kortene med bokstavsiden opp i en bunke. Eleven trekker et kort og skal plassere kortet oppå plansjen med PAS tegnene; riktig bokstav skal matches til riktig PAS tegn.

## **3. Skrive ord med PAS kort**

Her kan mange innfallsvinkler brukes: Eleven kan skrive et ord med PAS brikker og læreren kan «lese» hvilket ord det er. Lærer kan skrive et ord og eleven lese.

Billedkort for eksempel fra Lotto eller memoryspill kan brukes:

Eleven trekker et billedkort og skal matche det til PAS symbolet for den første (eller siste) lyden i ordet på bildet

Eleven trekker et kort og skal så skrive ordet med PAS brikkene.

Arbeid med minimale par: Hensikten er å øke den fonologiske bevisstheten ved å hjelpe eleven til å bli bevisst på at *en* lyd kan endre ords betydning eks. se-le, mål-bål, hatt-katt, sel-mel o.s.v. Eleven lukker øynene og lærer legger ord med pas kortene. Eleven skal så lese hva som står, lukke øynene og lærer endrer en bokstav. Så skal eleven lese på nytt. Man kan også bytte på slik at eleven legger kortene og lærer lukker øynene, leser osv.

## **4. Tekstlesing**

Lærer velger ut en tekst som passer til elevens lesenivå. Eleven leser teksten høyt og lærer noterer ned dersom det forekommer feillesinger eller dårlig flyt i lesingen. Eleven arbeider med de ordene som har vært vanskelige å lese gjennom å bruke PAS kortene til å stave ordene. Lærer bør oppfordre til tydelig artikulasjon (ikke nødvendigvis høyt, men at eleven beveger munnen og «føler på» lydene.) Etterpå skal eleven lese samme tekst på nytt og bør igjen oppfordres til å formulere/artikulere lydene tydelig.

## **5. Sortere kortene og sette tilbake i kasse**

Tidsbruken for de ulike aktiviteten er veiledende og kan eventuelt justeres etter hvert. For eksempel er det sannsynlig at aktivitet 1 og 5 etter hvert vil kreve mindre tidsbruk enn aktivitet 2 og 3. Det er imidlertid viktig med skifte underveis i hver økt for variasjonens skyld.

## **Aktivitet 4 bør introduseres først nå elevene er blitt kjent med alle symbolene i PAS.**

### **Lærerens rolle under treningen**

Læreren bør på forhånd sette seg litt inn i PAS tegnene og hva de ulike symbolene betyr, men det er ikke nødvendig at læreren er utlært «ekspert». Lærerens holdning under treningen kan gjerne være samarbeidsorientert slik at treningen mer tar form av at lærer og elev sammen utforsker systemet med de «hemmelige» tegnene. Lærerens rolle er ellers å modellere/ vise eleven hvordan lydene lages i munnen og hjelpe eleven til å rette oppmerksomheten mot den informasjonen som ligger i symbolene.

Når arbeidet starter opp første gang kan læreren gjerne introdusere treningen på denne måten: «For de som har dysleksi er det vanskelig å lese med de vanlige bokstavene, nå skal vi prøve å lære å lese og skrive gjennom å bruke et annet system, det blir nesten som å lære hemmelige tegn.....»

### **Progresjon i innlæringen**

Det anbefales å starte med de enkleste symbolene først og så introdusere flere etter hvert. De enkleste symbolene er I S M O L. Hvor raskt man skal introdusere nye symboler er en vurdering som læreren må ta underveis og som vil avhenge av hvor raskt eleven «tar» prinsippene for PAS og lærer seg grunnsymbolene. (Erfaringsmessig er dette ofte langt enklere for elevene enn for oss voksne! )

### **Logg**

Det er laget et loggskjema som skal fylles ut etter hver økt. Dette for å bekrefte at gangen i opplegget er fulgt, men det kan også brukes til å skrive inn kommentarer og vurderinger i forhold til det som har skjedd i løpet av timen. For eksempel elevens bruk av strategier, motivasjon, grad av automatisering etc.



- |          |           |                  |
|----------|-----------|------------------|
| 1. jo    | 24. løve  | 47. flest        |
| 2. bo    | 25. lege  | 48. svart        |
| 3. be    | 26. hage  | 49. anklagene    |
| 4. se    | 27. gåte  | 50. eksemplarer  |
| 5. ni    | 28. høre  | 51. fotografen   |
| 6. ga    | 29. høne  | 52. medlidenhet  |
| 7. på    | 30. pute  | 53. desperate    |
| 8. tå    | 31. seks  | 54. ukonsentrert |
| 9. gå    | 32. vant  | 55.              |
| 10. dem  | 33. slik  | 56.              |
| 11. før  | 34. fant  |                  |
| 12. bil  | 35. helt  |                  |
| 13. tok  | 36. fram  |                  |
| 14. hus  | 37. siste |                  |
| 15. den  | 38. lenke |                  |
| 16. han  | 39. bedre |                  |
| 17. ned  | 40. burde |                  |
| 18. din  | 41. hørte |                  |
| 19. går  | 42. flere |                  |
| 20. da   | 43. grøft |                  |
| 21. løpe | 44. stund |                  |
| 22. kose | 45. frukt |                  |
| 23. mene | 46. grind |                  |

- |          |            |                 |
|----------|------------|-----------------|
| 1. is    | 24. løse   | 47. kveld       |
| 2. ha    | 25. mene   | 48. svært       |
| 3. ja    | 26. lege   | 49. tapetsere   |
| 4. te    | 27. lite   | 50. bestrebelse |
| 5. så    | 28. gave   | 51. eksemplarer |
| 6. do    | 29. pære   | 52. termometer  |
| 7. vi    | 30. dere   | 53. desperate   |
| 8. nå    | 31. seks   | 54. helikopter  |
| 9. lo    | 32. svin   | 55.             |
| 10. få   | 33. fram   |                 |
| 11. nål  | 34. tror   |                 |
| 12. far  | 35. frem   |                 |
| 13. mor  | 36. stek   |                 |
| 14. fem  | 37. flere  |                 |
| 15. hus  | 38. viste  |                 |
| 16. løk  | 39. kroner |                 |
| 17. går  | 40. lenke  |                 |
| 18. bil  | 41. bedre  |                 |
| 19. får  | 42. følte  |                 |
| 20. ser  | 43. prest  |                 |
| 21. måne | 44. smakt  |                 |
| 22. rive | 45. frukt  |                 |
| 23. sine | 46. grøft  |                 |

1.	es	24.	ams	47.	gremse
2.	ma	25.	ast	48.	smalfist
3.	ol	26.	unt	49.	skrelske
4.	al	27.	uks	50.	smilfest
5.	im	28.	afs	51.	skrylske
6.	ra	29.	efs	52.	smylrest
7.	va	30.	vryme	53.	skrensle
8.	ly	31.	trupe		
9.	fa	32.	krafe		
10.	fø	33.	gløre		
11.	ne	34.	vrise		
12.	ek	35.	vrane		
13.	ife	36.	vreg		
14.	one	37.	dims		
15.	æbe	38.	fliv		
16.	ope	39.	pløf		
17.	ike	40.	fløg		
18.	åne	41.	krot		
19.	vape	42.	trymse		
20.	råfe	43.	brimse		
21.	lipe	44.	flente		
22.	mise	45.	plaste		
23.	beke	46.	branse		

**LES-2****Leseprøve nonord 11****Appendiks J**

1. el
2. es
3. ma
4. mi
5. re
6. ar
7. ot
8. sø
9. fø
10. fa
11. lø
12. ke
13. use
14. ike
15. obe
16. ife
17. yge
18. øfe
19. luge
20. beke
21. gole
22. mise
23. våne

24. røte
25. est
26. ifs
27. eps
28. int
29. ent
30. ips
31. trupe
32. slife
33. klere
34. kripe
35. flare
36. glire
37. valm
38. flus
39. bløk
40. pløf
41. kølv
42. dalk
43. knurse
44. bleste
45. klesme
46. brimse

47. krimse
48. flaste
49. skvultert
50. skviltert
51. smelvist
52. skrilske
53. skrinkest
54. spriste
- 55.

- |           |             |              |
|-----------|-------------|--------------|
| 1. nok    | 24. kveld   | 47. hjelpe   |
| 2. for    | 25. sagt    | 48. kino     |
| 3. lov    | 26. gikk    | 49. tjære    |
| 4. meg    | 27. gjort   | 50. hjemme   |
| 5. der    | 28. ung     | 51. skjønner |
| 6. lov    | 29. orm     | 52. skyte    |
| 7. og     | 30. hund    | 53. hvilken  |
| 8. seg    | 31. sopp    | 54. kjempe   |
| 9. kun    | 32. eng     |              |
| 10. vid   | 33. stadig  |              |
| 11. knott | 34. holder  |              |
| 12. hval  | 35. hvorfor |              |
| 13. skal  | 36. kjæle   |              |
| 14. gikk  | 37. bukse   |              |
| 15. hvem  | 38. skifte  |              |
| 16. skyld | 39. gjøre   |              |
| 17. selv  | 40. loppe   |              |
| 18. sjal  | 41. skjærer |              |
| 19. frosk | 42. bukse   |              |
| 20. vers  | 43. fosse   |              |
| 21. lagt  | 44. tjene   |              |
| 22. havs  | 45. kylling |              |
| 23. hvert | 46. hvilken |              |

- |           |             |             |
|-----------|-------------|-------------|
| 1. for    | 24. hjul    | 47. hjelpe  |
| 2. den    | 25. sjø     | 48. kjenne  |
| 3. det    | 26. knott   | 49. gjemsel |
| 4. jeg    | 27. kveld   | 50. holder  |
| 5. at     | 28. vers    | 51. sperre  |
| 6. kom    | 29. vers    | 52. skyte   |
| 7. nok    | 30. godt    | 53. skjærer |
| 8. er     | 31. synd    | 54. sjelden |
| 9. ved    | 32. hval    |             |
| 10. dum   | 33. kjøre   |             |
| 11. sjy   | 34. hvilken |             |
| 12. sølv  | 35. skinne  |             |
| 13. flott | 36. kylling |             |
| 14. kjørt | 37. hjerne  |             |
| 15. ork   | 38. hvilken |             |
| 16. hva   | 39. hoppe   |             |
| 17. ung   | 40. kino    |             |
| 18. sjal  | 41. hjemme  |             |
| 19. sopp  | 42. hvorfor |             |
| 20. gikk  | 43. gjorde  |             |
| 21. gjøk  | 44. hjørne  |             |
| 22. gikk  | 45. kjæle   |             |
| 23. rødt  | 46. tjene   |             |

Blokk1 1. <b>al</b> 2. <b>el</b> 3. <b>mi</b> 4. <b>im</b>	Blokk 4 1. <b>gole</b> 2. <b>båge</b> 3. <b>bute</b> 4. <b>luge</b>	Blokk7 1. <b>krot</b> 2. <b>plåf</b> 3. <b>fløg</b> 4. <b>dims</b>
Blokk2 1. <b>ge</b> 2. <b>ek</b> 3. <b>ot</b> 4. <b>ef</b>	Blokk5 1. <b>ifs</b> 2. <b>esk</b> 3. <b>ysk</b> 4. <b>ums</b>	Blokk8 1. <b>flinte</b> 2. <b>knurse</b> 3. <b>klesme</b> 4. <b>glypte</b>
Blokk 3 1. <b>ife</b> 2. <b>ufe</b> 3. <b>ude</b> 4. <b>æme</b>	Blokk 6 1. <b>stufe</b> 2. <b>glate</b> 3. <b>glire</b> 4. <b>flare</b>	Blokk9 1. <b>skrelske</b> 2. <b>skveltert</b> 3. <b>skrinsle</b> 4. <b>skvultert</b>

Blokk1 1. im 2. al 3. re 4. il	Blokk4 1. våne 2. peme 3. sute 4. lipe	Blokk7 1. delm 2. klås 3. frag 4. pløf
Blokk2 1. ru 2. ef 3. vø 4. ly	Blokk5 1. osk 2. imt 3. unt 4. eps	Blokk8 1. bleste 2. klamse 3. glymte 4. plaste
Blokk3 1. ife 2. øle 3. ufe 4. one	Blokk6 1. klure 2. vryme 3. vrise 4. drule	Blokk9 1. smalfist 2. skrilske 3. skrinkest 4. spranke

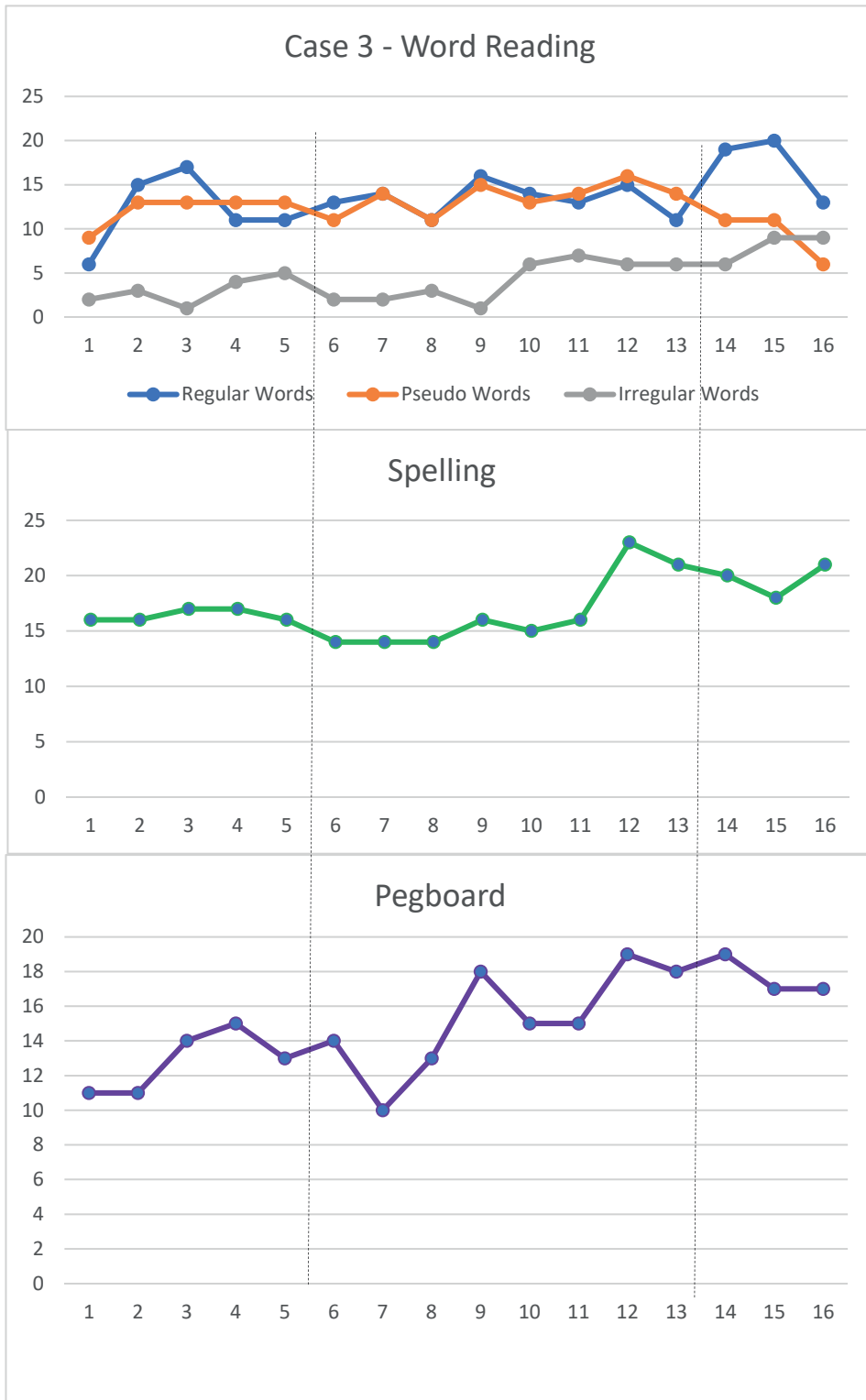


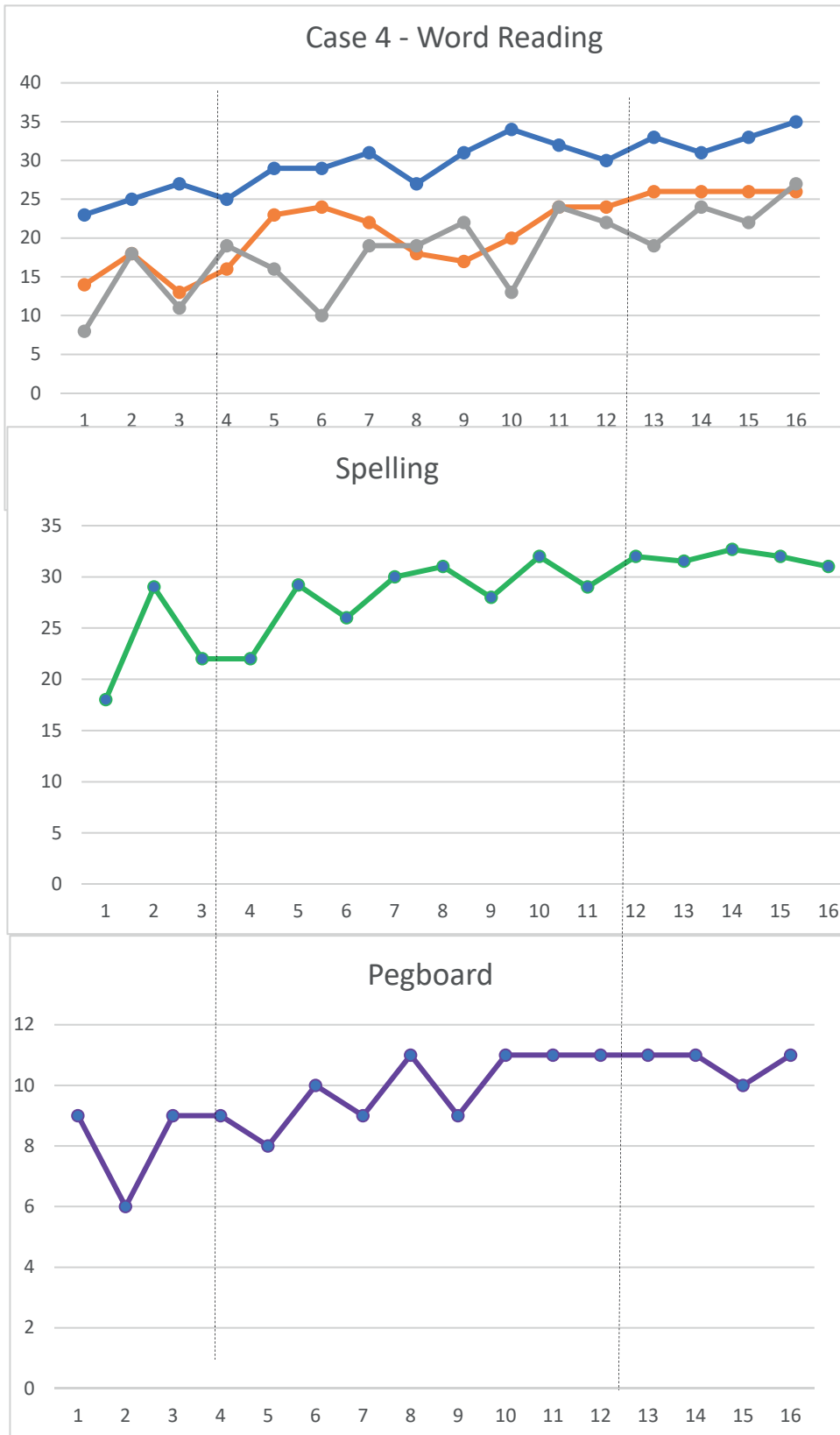
## **Supplemental material**

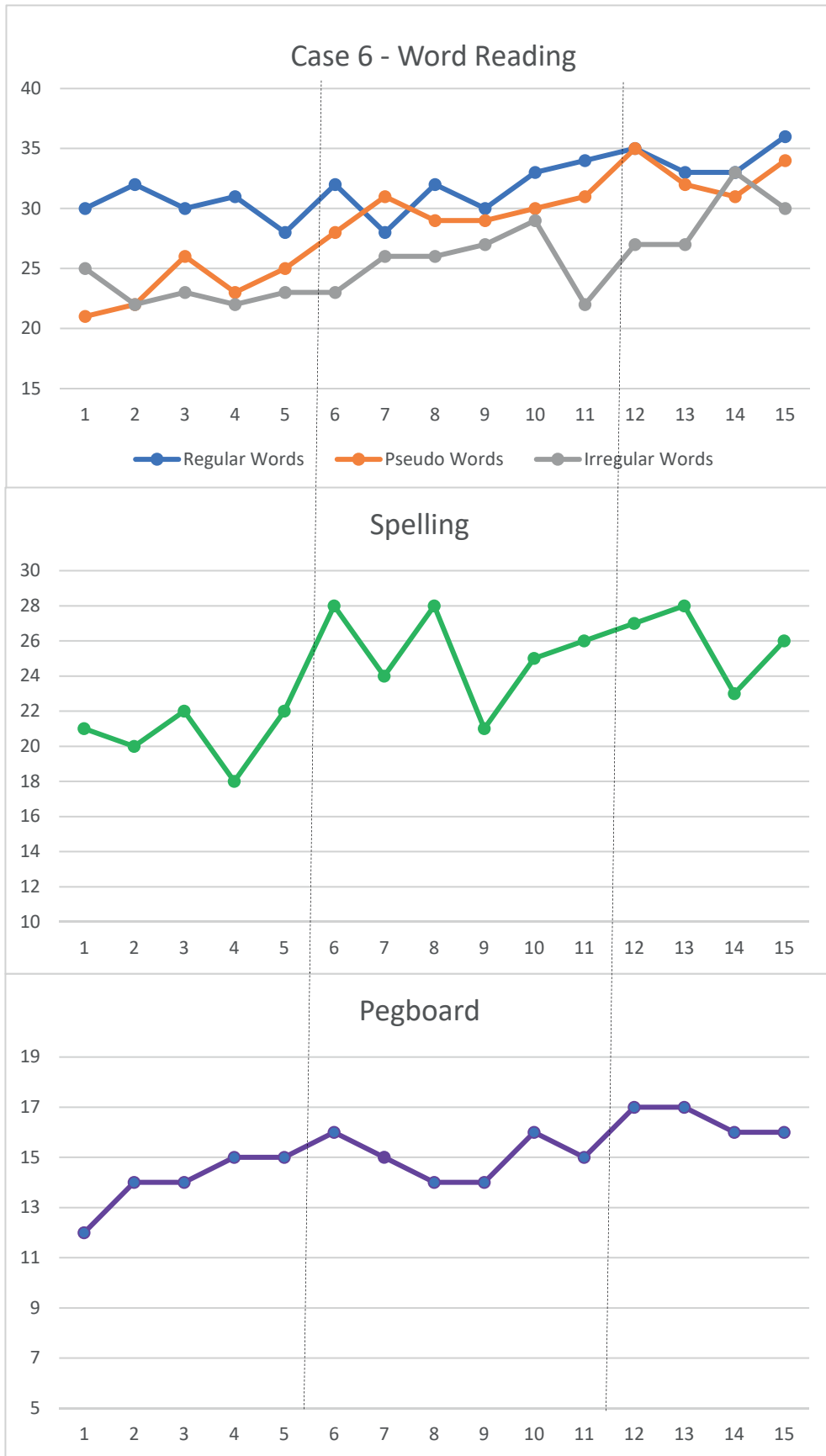
### **1. Graphic charts for each participant.**

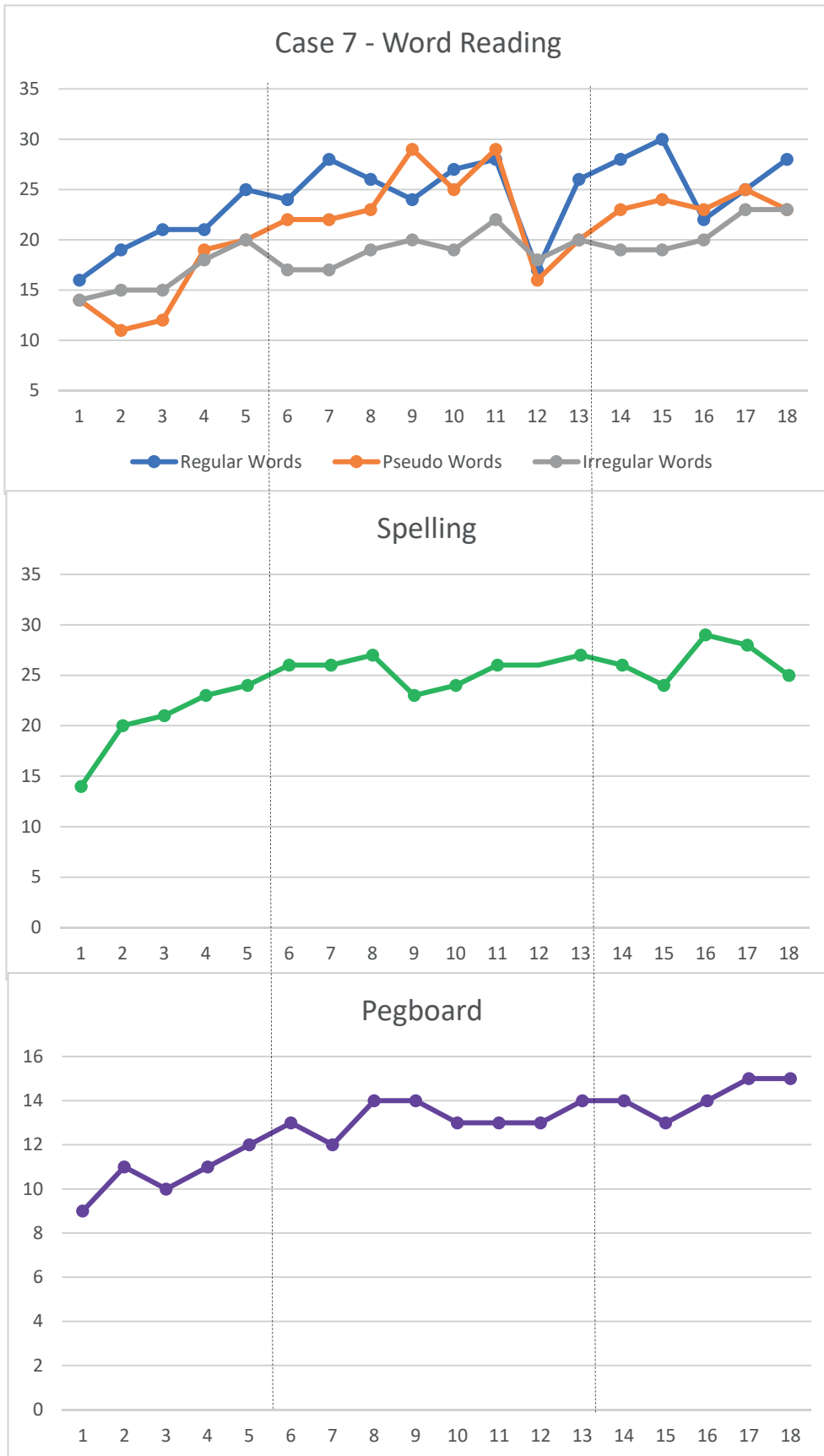
#### **1.2. Raw scores at each timepoint of measurement.**

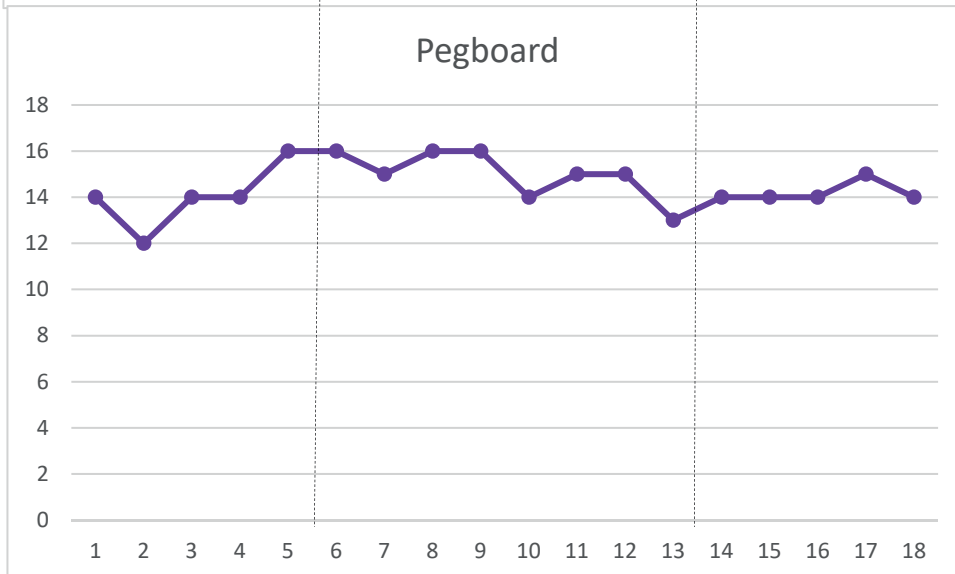
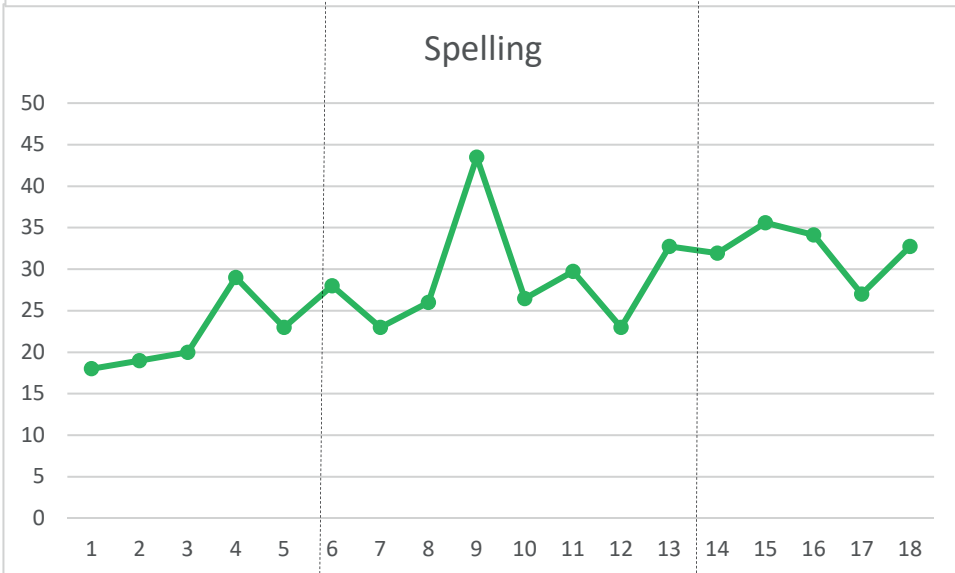
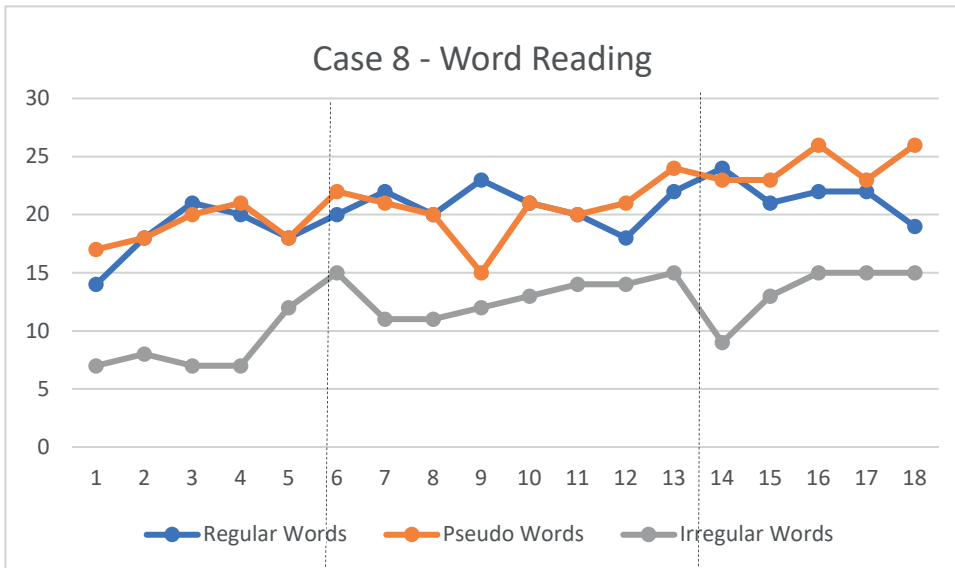
Y- axis: Raw scores, X-axis: Time of measurement. Vertical lines: Phase changes (Baseline, Intervention, Post). Note: Individual breaks from probe sessions and scheduled school holidays do not appear in the graphics-

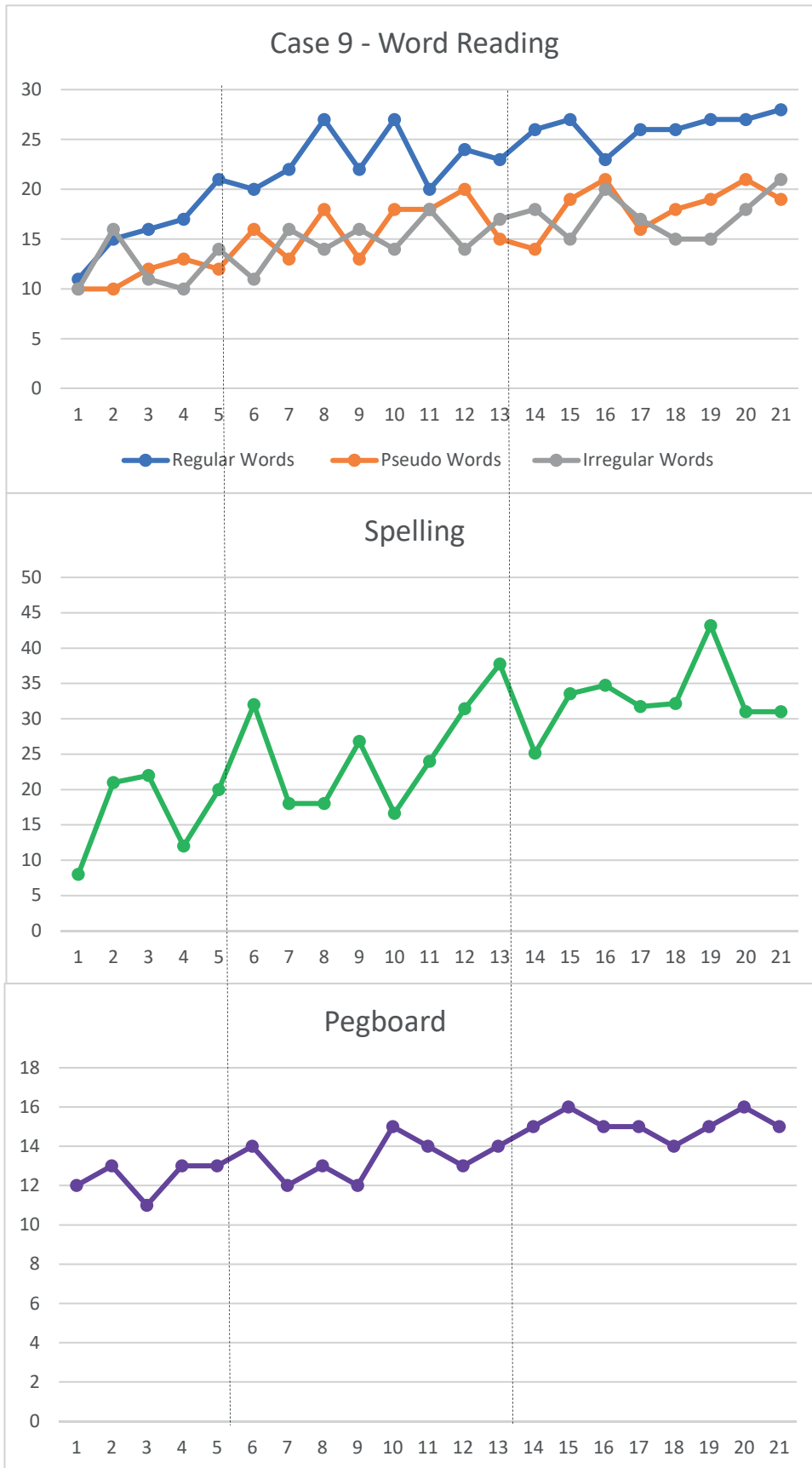


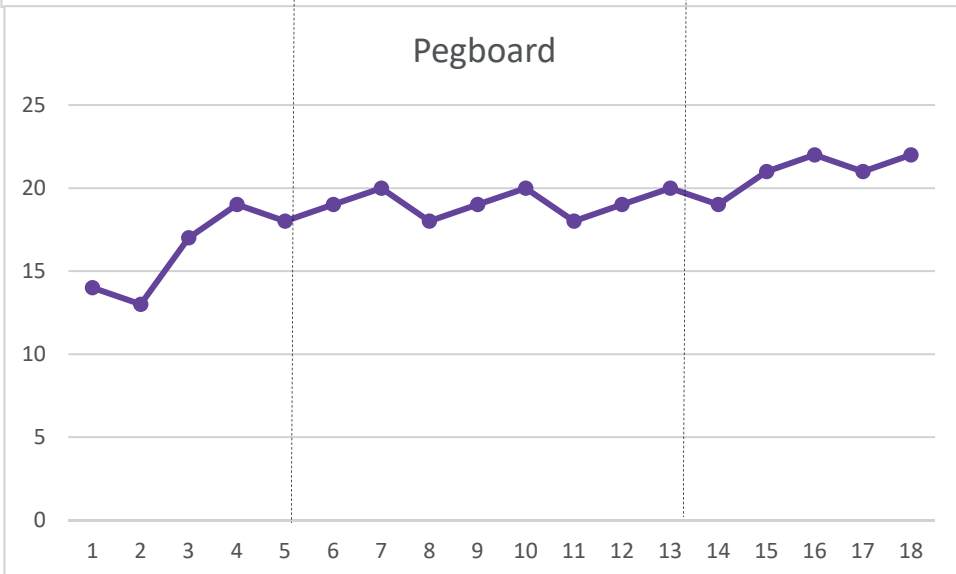
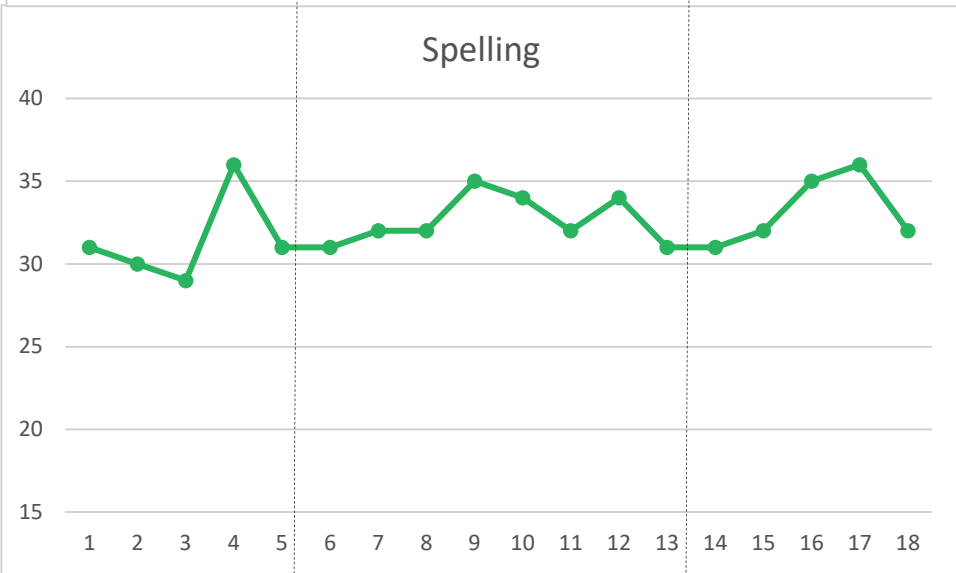
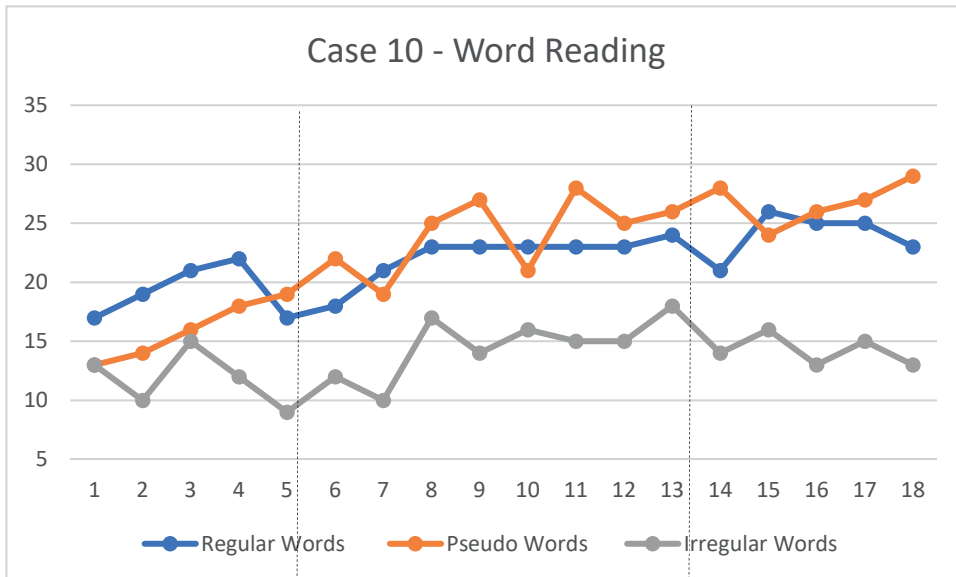




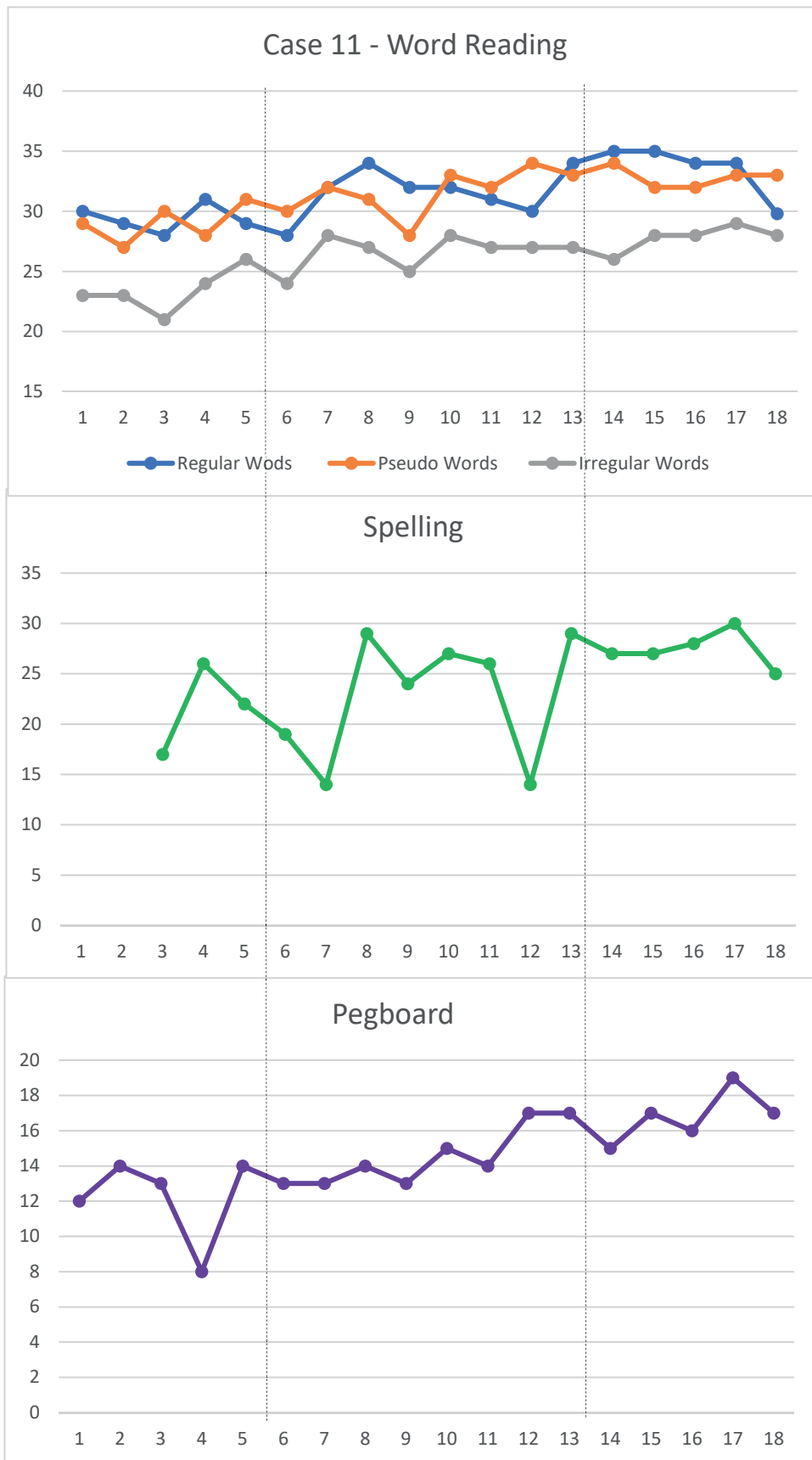


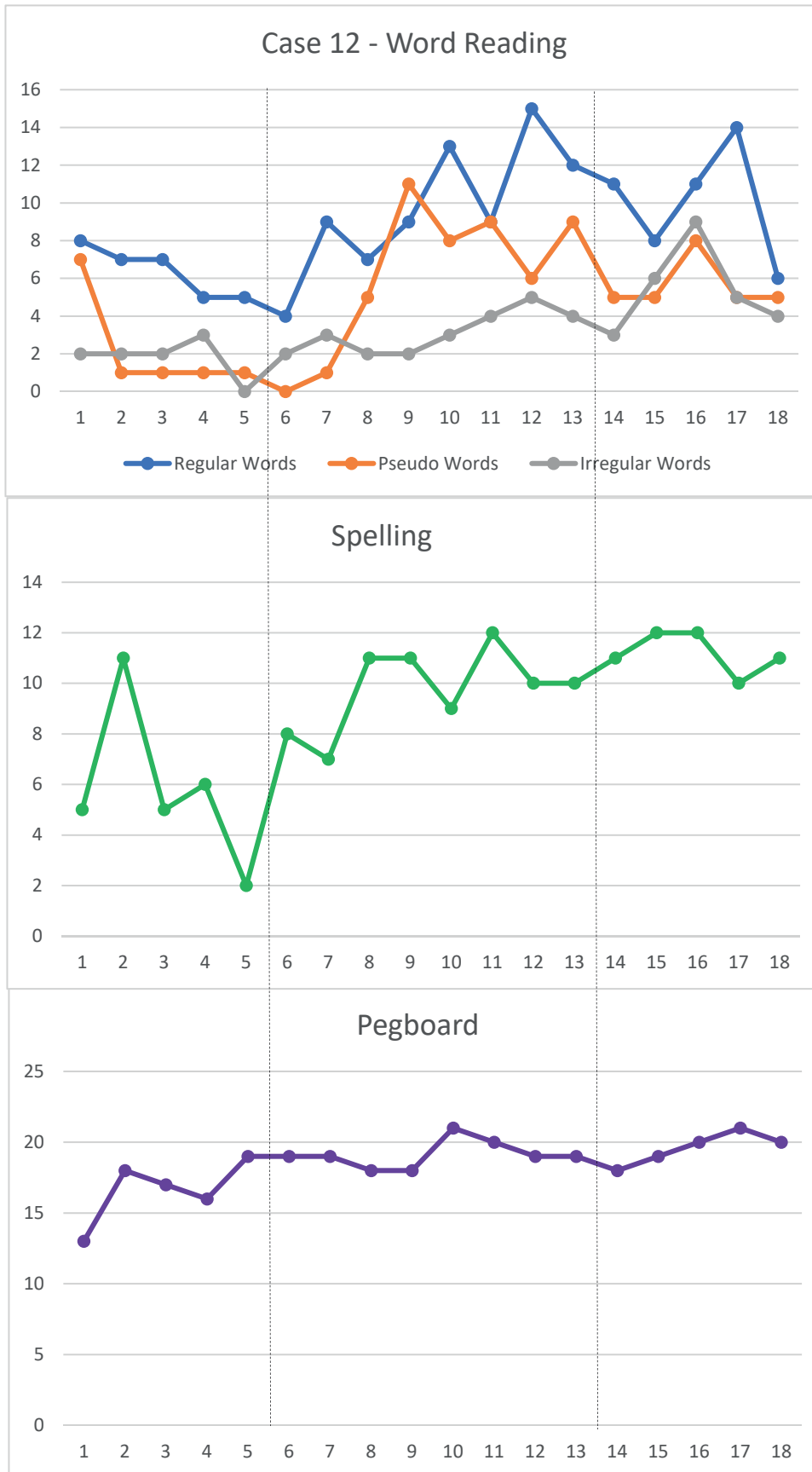


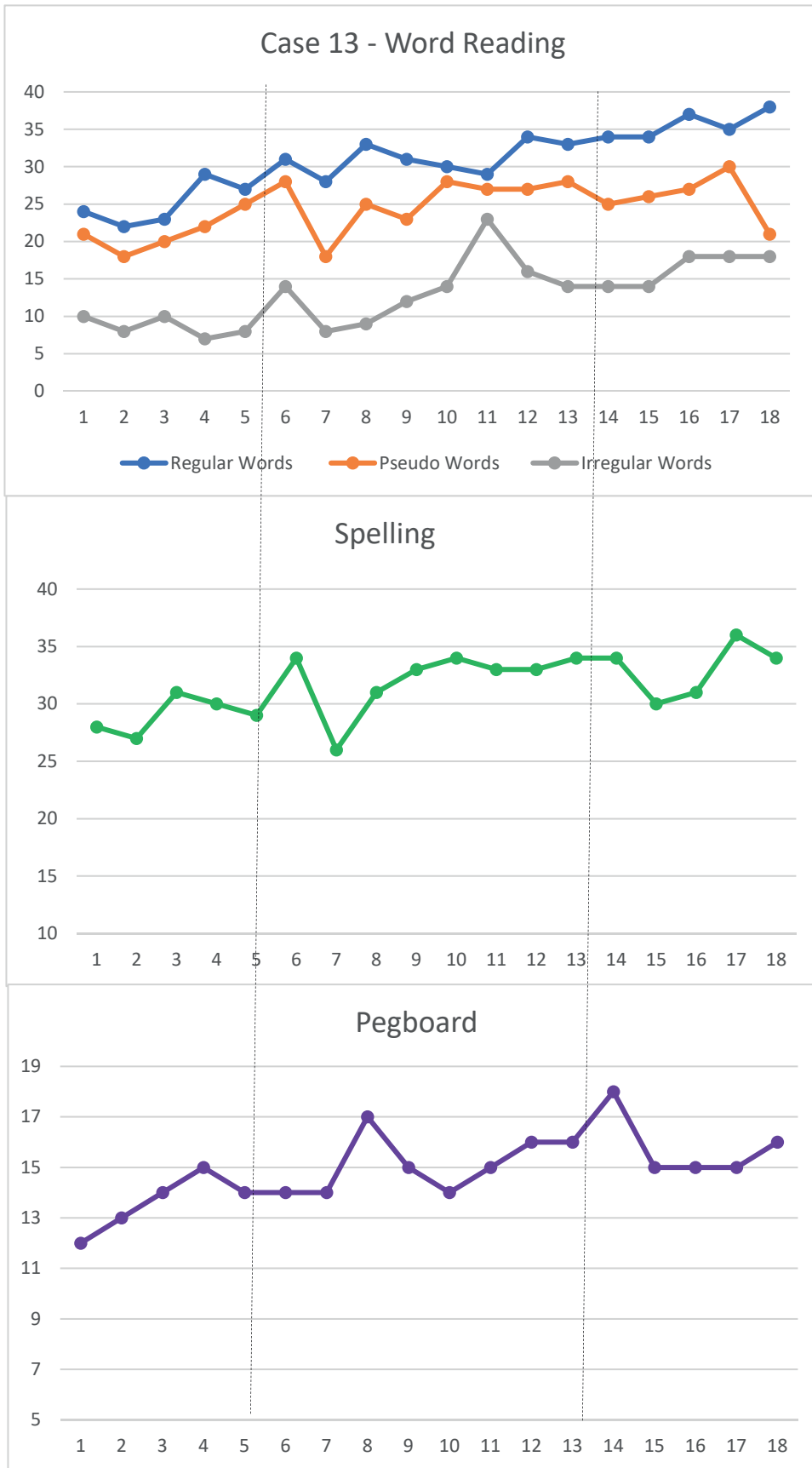




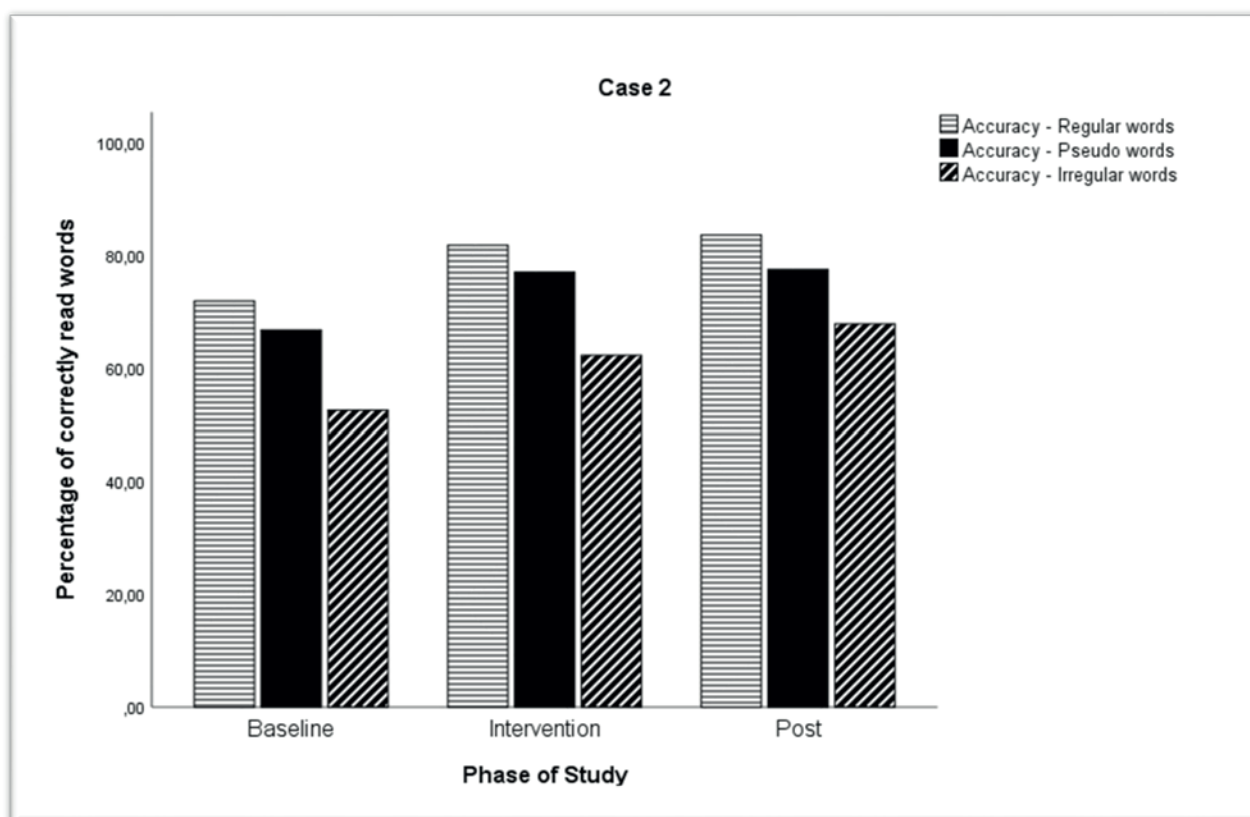


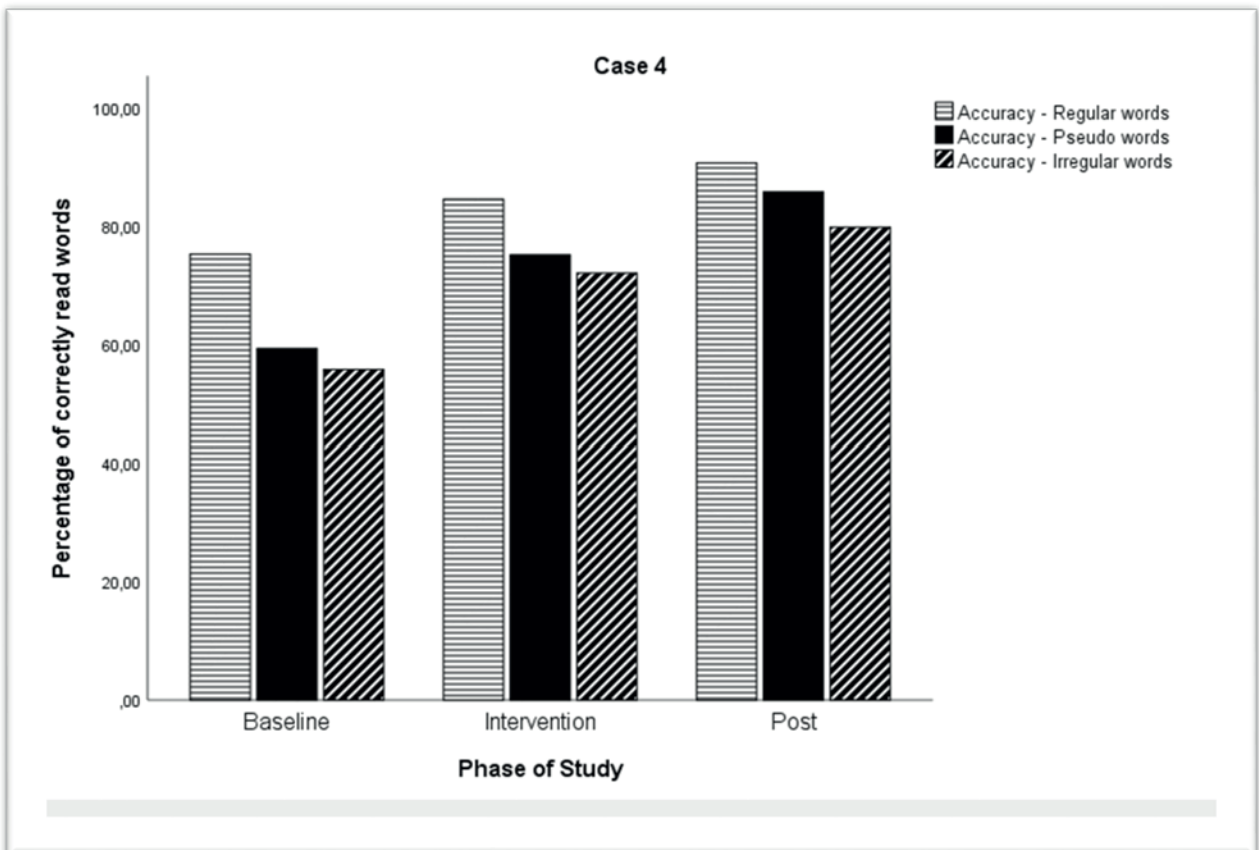
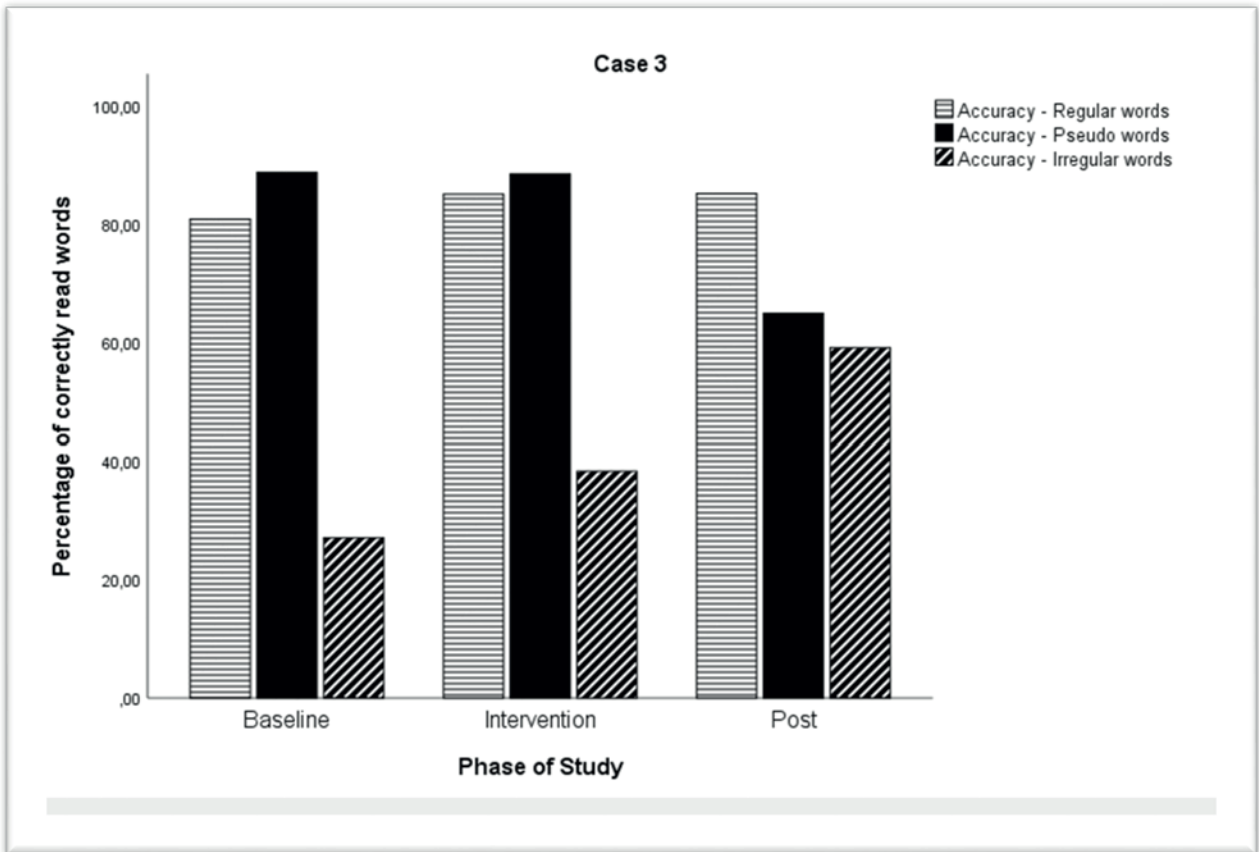


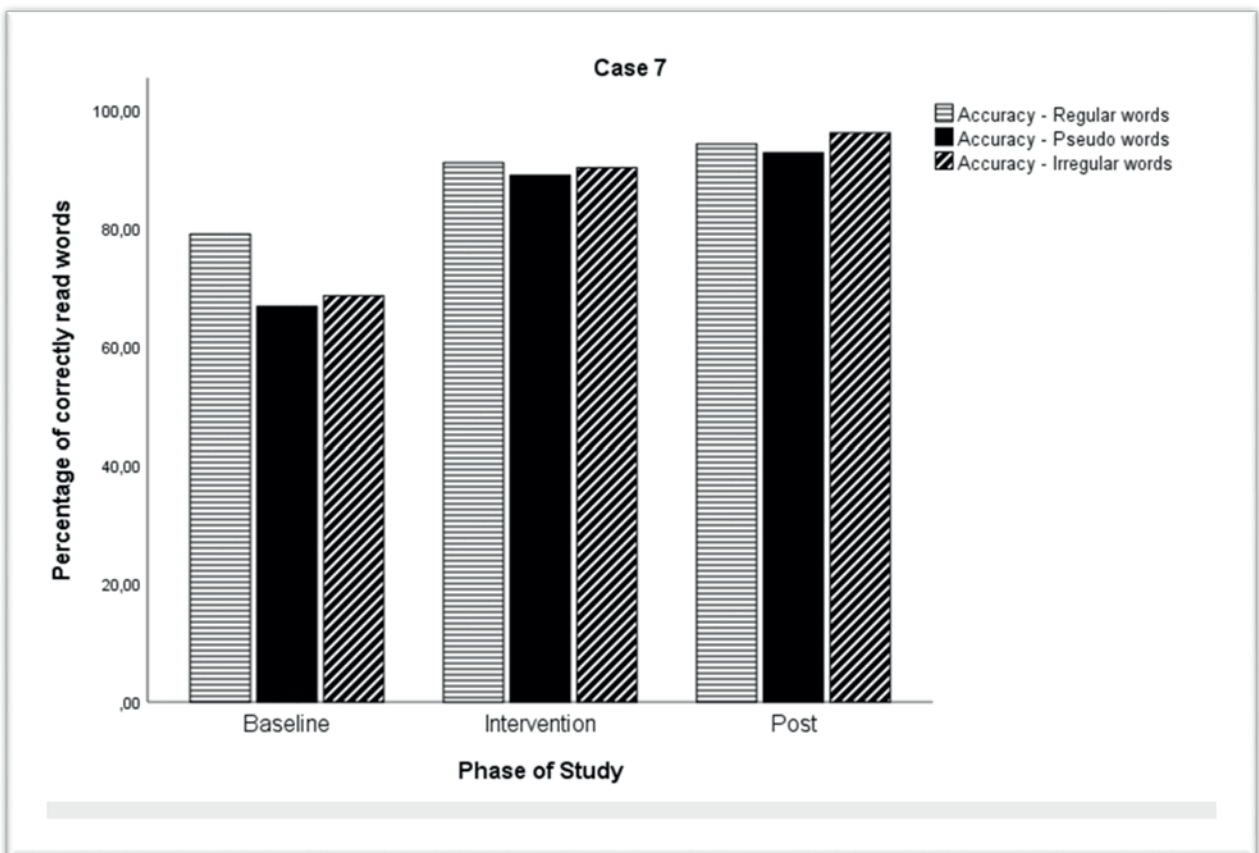
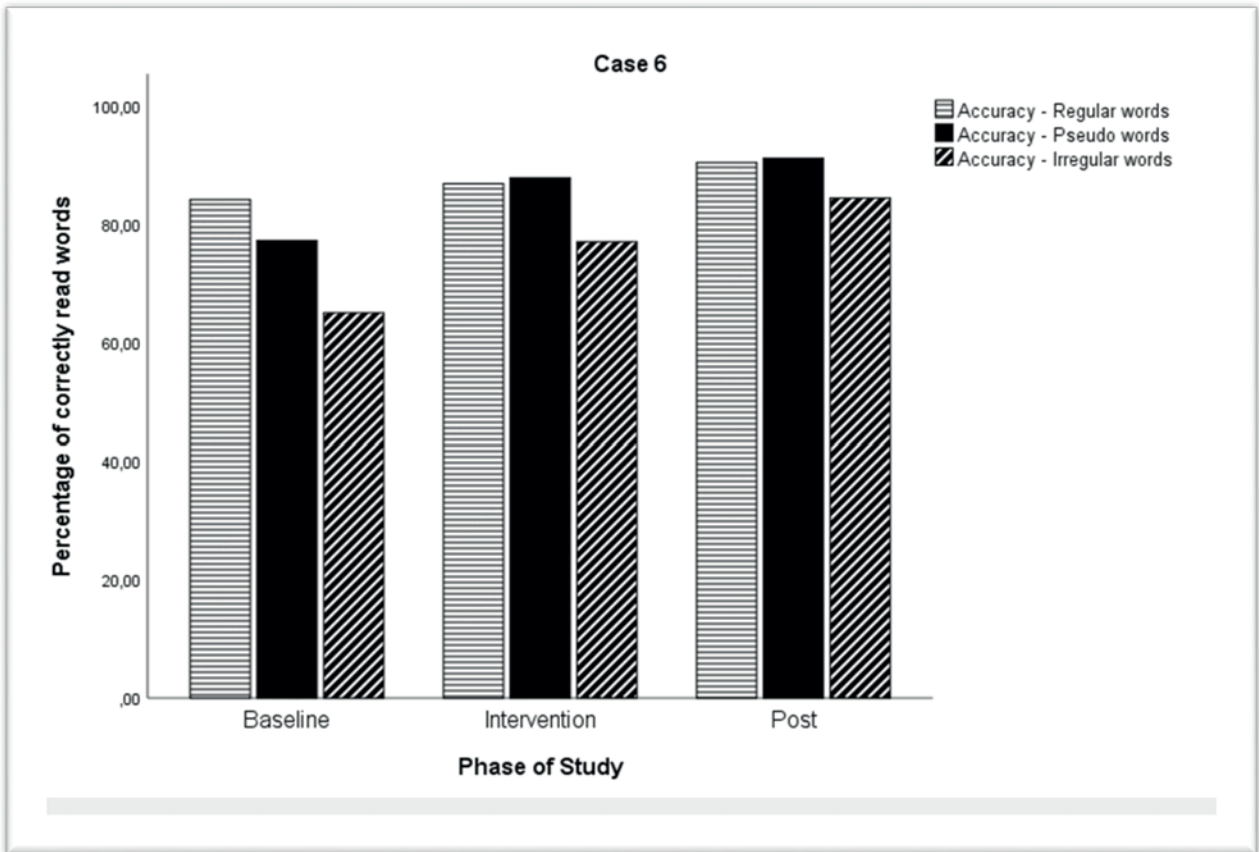


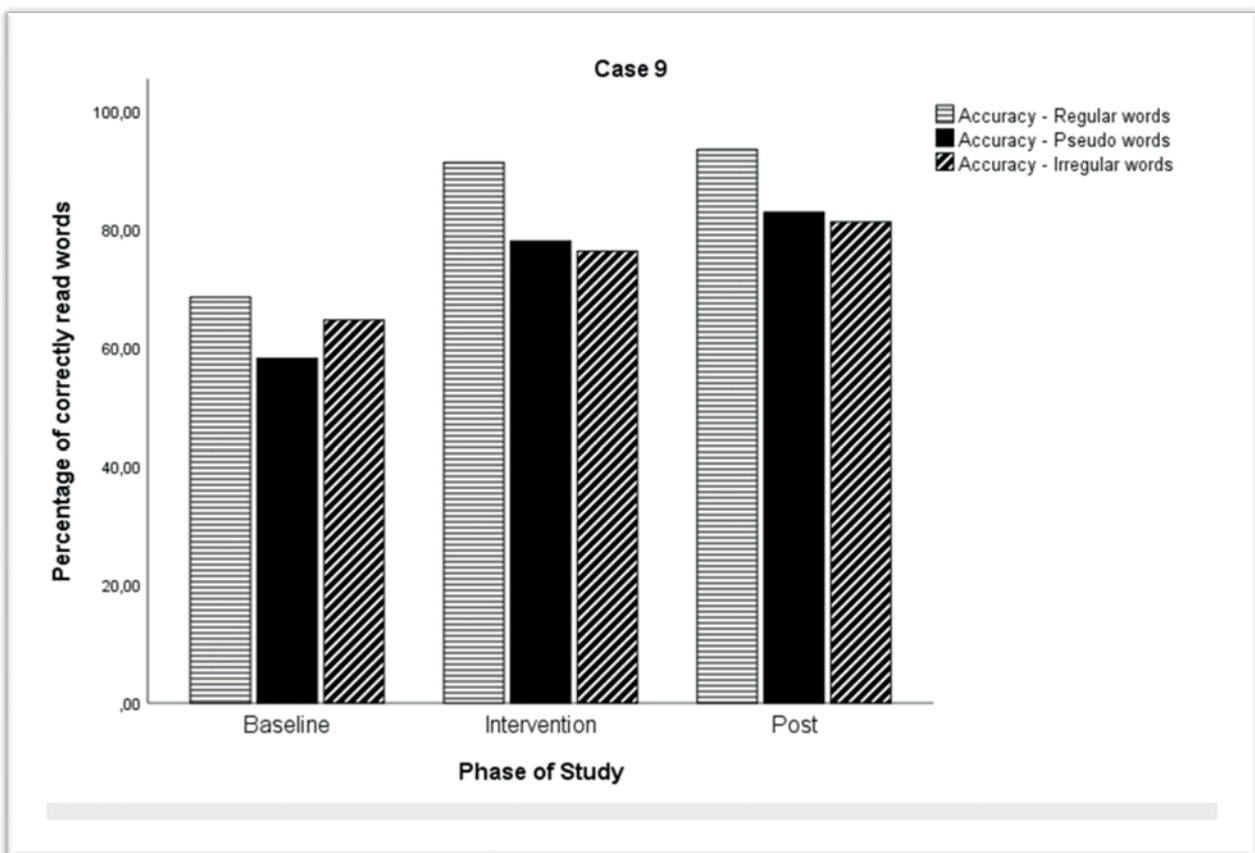
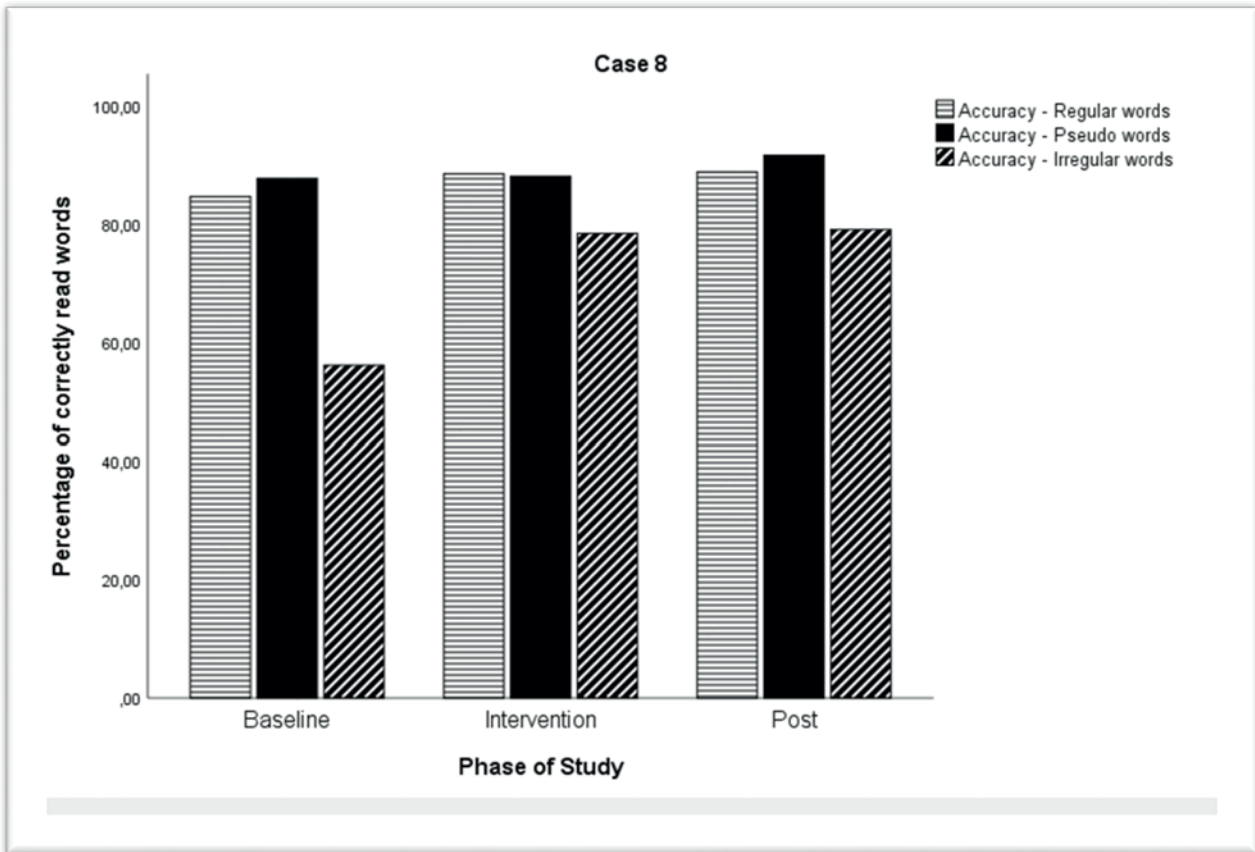


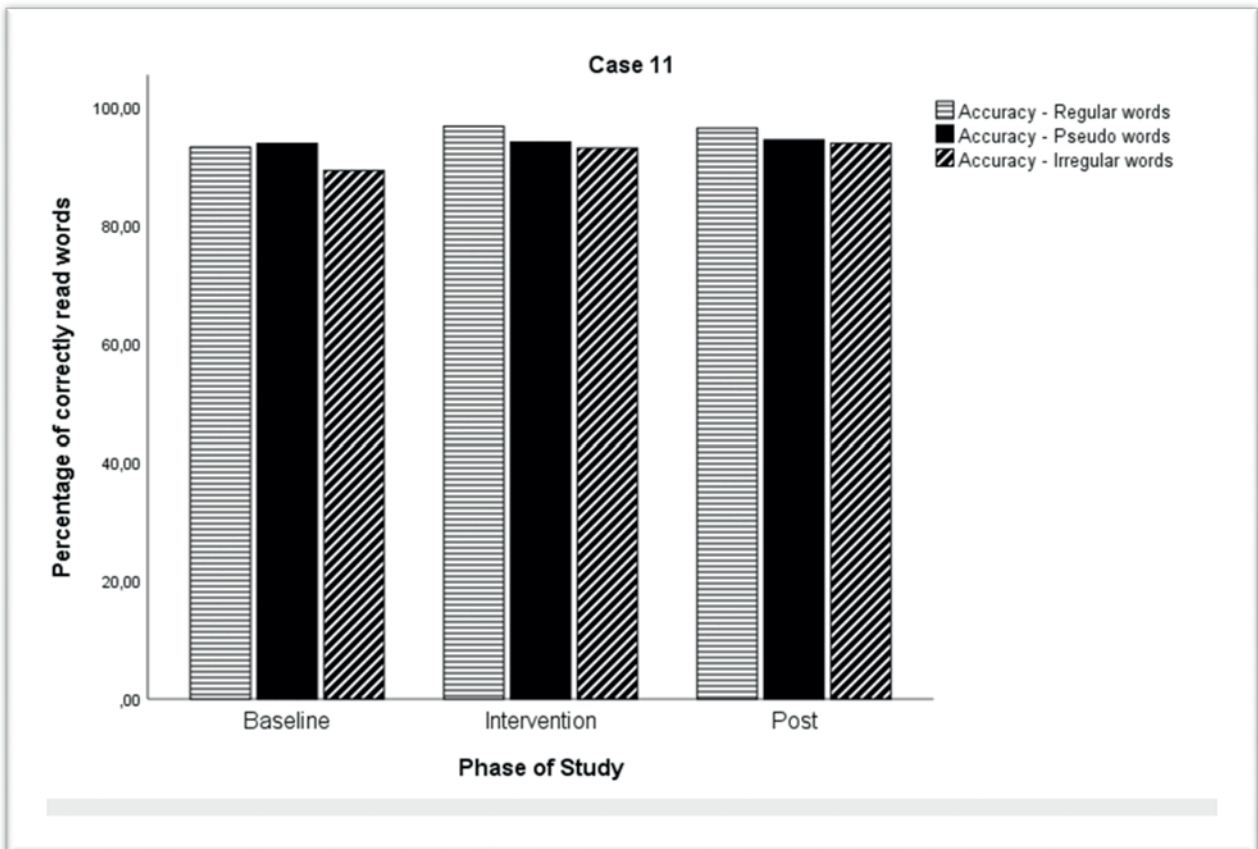
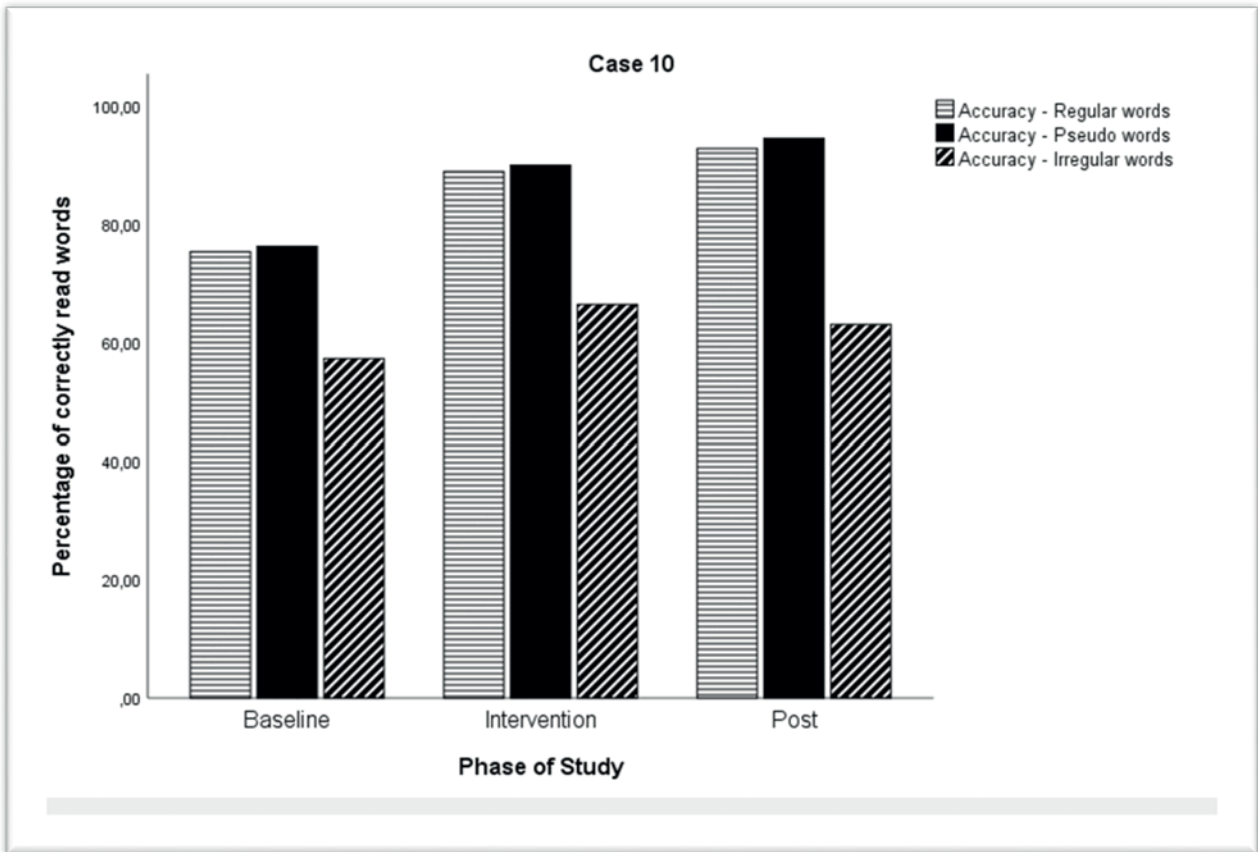
**2.2 Bar charts showing the level change of reading accuracy across three phases of study.**



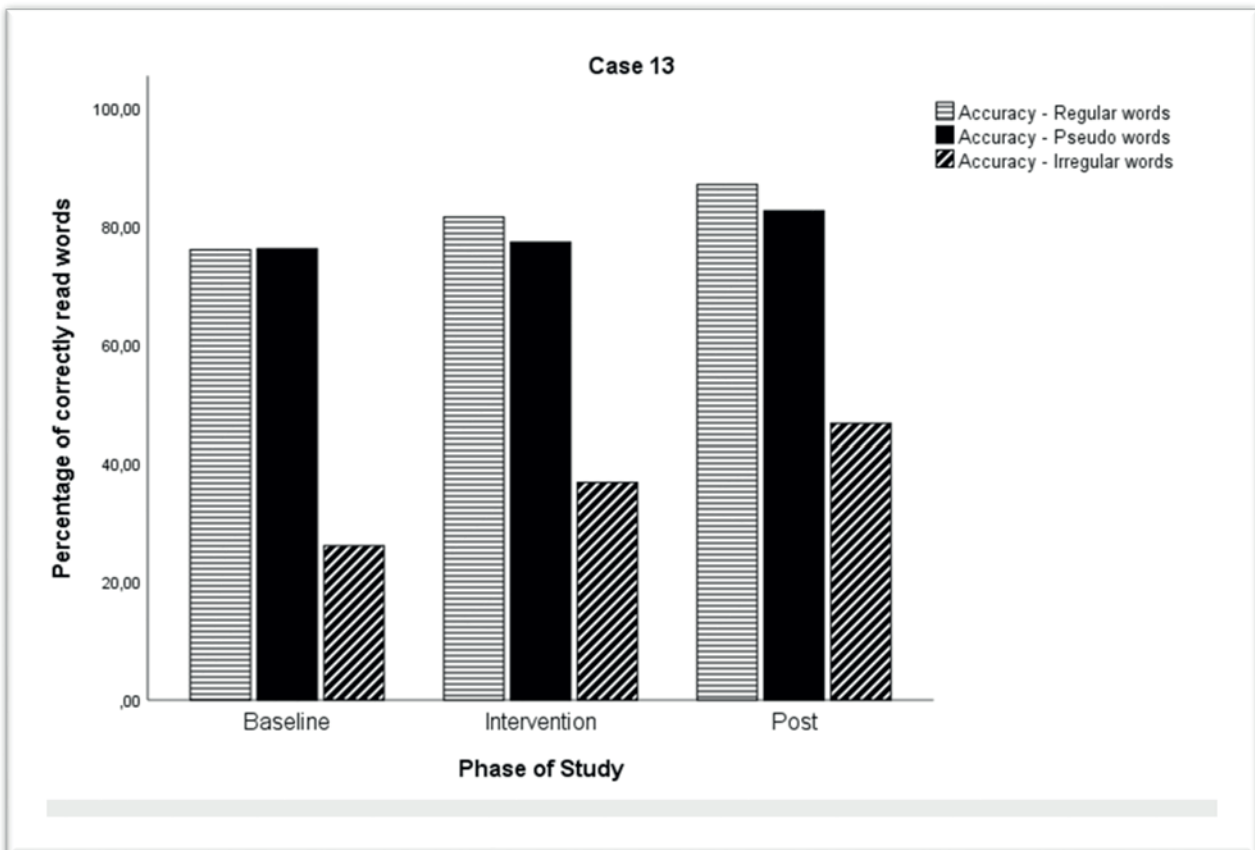
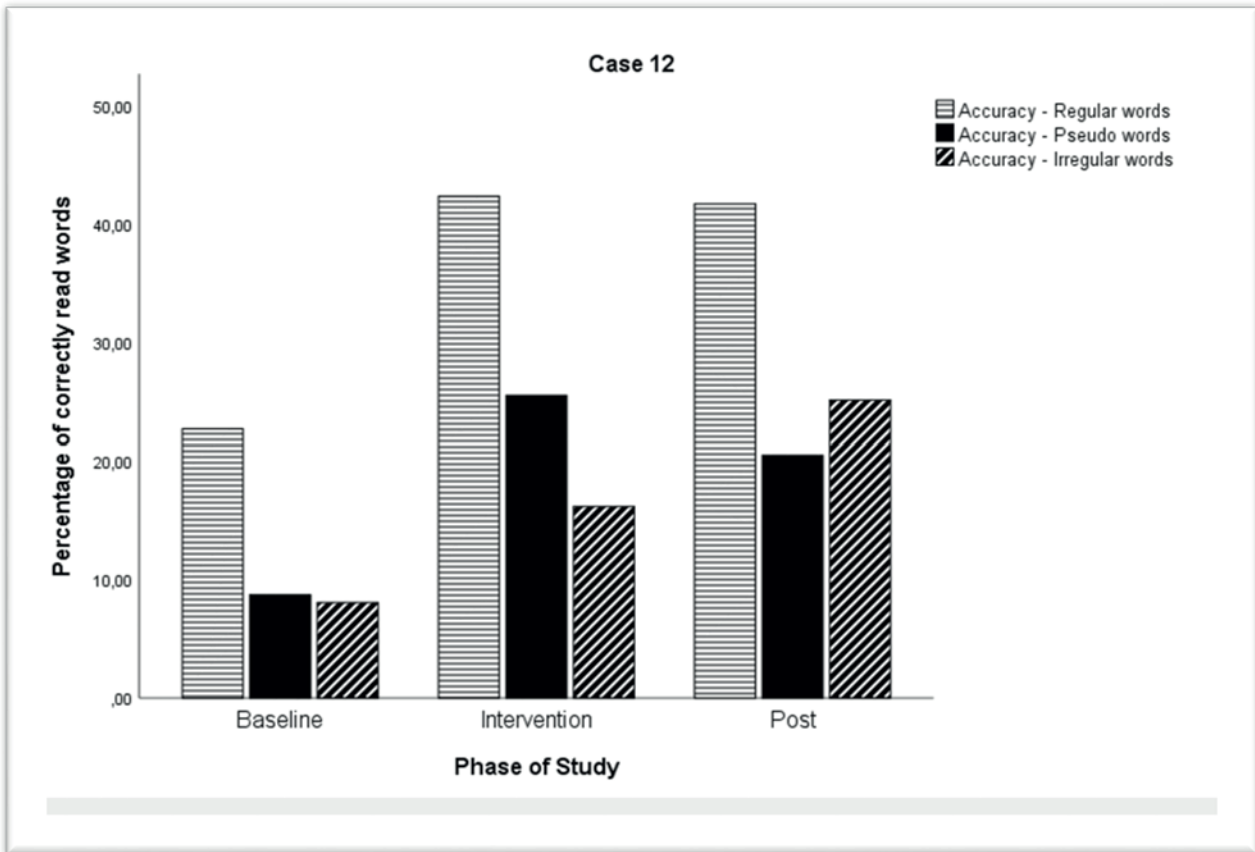














## Forespørsel om deltakelse i forskningsprosjektet

### ”Artikulatorisk bevisstgjøring av fonemene og lydmessig feedback på fonemnivå”

#### Bakgrunn og formål

Prosjektet er rettet mot elever med språk, lese og skrivevansker som er henvist til Statped og/eller Pedagogisk psykologisk tjeneste og omhandler systematisk utprøving av tiltak for elever med alvorlig grad av lese, skrive og språkvansker. Hensikten med prosjektet er å få sikrere kunnskap om hvordan vi kan oppnå større grad av treffsikkerhet når det gjelder funksjonelle tiltak for denne elevgruppen. Studien inngår som ledd i prosjektlederens dr. gradsprosjekt og veiledes fra Universitetet i Oslo.

#### Hva innebærer deltakelse i studien?

Prosjektet er en treningsstudie og tar sikte på å prøve ut to ulike innfallsvinkler til økt fonologisk bevissthet. Den første gjennom bevisstgjøring av bokstavlydenes artikulatoriske plassering og den andre gjennom økt automatisering av lyd/bokstavforbindelsene ved hjelp av digitale hjelpemidler. Tiltakene vil bli prøvd ut i to intensive perioder, hver med ca fem ukers varighet. Utprøving av tiltakene vil foregå på skolen og elevenes lærere vil motta opplæring og veiledning før og underveis i prosessen. Elevene vil bli kartlagt med språk og leseprøver både før, under og etter tiltakene. Det vil også bli gjort lydopptak av elevene under kartlegging og undervisning. Kartlegging vil delvis gjøres av lærerne og delvis av saksbehandlere fra Pedagogisk psykologisk tjeneste og Statped.

I tiltaksperiodene (seks uker x2) vil elevene motta 40 minutters opplæring 4 dager pr uke. Denne opplæringen vil i de periodene tiltaksutprøvingen foregår helt eller delvis erstatte den ordinære spesialundervisningen.

En del av formålet med prosjektet er å se på om ulikheter i diagnostisk profil hos elevene virker inn på effekten av tiltakene.. I tillegg til de kartleggingsprøvene som tas i forbindelse med utprøvingen av tiltak vil data også bli hentet fra tidligere utredninger av eleven gjennom innsyn i elevens journal i pedagogisk psykologisk tjeneste og resultater fra kartleggingsprøver tatt på skolen. Videre oppfølging av elevene etter tiltaksperioden vil ivaretas gjennom de vanlige rutinene for samarbeid mellom skole og Pedagogisk psykologisk tjeneste rundt henviste elever. Langtidseffekt vil bli målt ved innsyn i kartleggingsprøver foretatt av PPT /skolen.

#### Hva skjer med informasjonen om deg?

Alle personopplysninger vil bli behandlet konfidensielt og vil kun være tilgjengelig for prosjektleder og veiledere i prosjektet. Data vil bli anonymisert og elever vil ikke kunne gjenkjennes ved publisering av materialet. |

Prosjektet skal etter planen avsluttes innen 1. september 2019. Data vil da oppbevares i elevenes ordinære journaler/saksmapper hos Statped eller PPT og oppbevaringen vil følge ordinære retningslinjer for behandling av journalopplysninger i disse institusjonene. [ ]

### **Frivillig deltakelse**

Det er frivillig å delta i studien, og du kan når som helst trekke ditt samtykke uten å oppgi noen grunn. Dette vil i såfall ikke ha noen innvirkninger på den generelle saksbehandlingen i Statped/PPT. Dersom du trekker deg, vil alle opplysninger om deg bli anonymisert. [ ]

Dersom du har spørsmål til studien, ta kontakt med prosjektleder Anne Cathrine Thurmann-Moe tlf. 22902821/92610925 eller via mail til [anne-cathrine.thurmann-moe@statped.no](mailto:anne-cathrine.thurmann-moe@statped.no)

Studien er meldt til Personvernombudet for forskning, Norsk samfunnsvitenskapelig datatjeneste AS.

.....

### **Samtykke til deltakelse i studien**

Jeg har mottatt informasjon om prosjektet «**Artikulatorisk bevisstgjøring av fonemene og lydmessig feedback på fonemnivå**» og gir med dette tillatelse til at mitt

Barn: ..... født.....

deltar i overnevnte prosjekt.

Jeg gir samtidig tillatelse til at tidligere informasjon om barnets språk og leseutvikling dokumentert i skolens eller pedagogisk psykologisk tjenestes journaler kan benyttes som data i prosjektet. Det gis også tillatelse til innsyn i resultater fra kartleggingsprøver i den perioden elevene er under oppfølging av PPT etter at intervensjonen er avsluttet.

Jeg er inneforstått med at elevens vedtak om spesialundervisning i den perioden tiltakene prøves ut blir helt eller delvis erstattet av aktuelt tiltak.

.....

Sted

.....  
Dato

.....  
Underskrift fra foresatte

## SKJEMA FOR BAKGRUNNSINFORMASJON VED LESE OG SKRIVEVANSKER

Barnets navn..... Fødselsdato.....

I forbindelse med forskningsprosjektet «Artikulatorisk støtte i lese og skriveopplæringen....» er det behov for å samle inn noe bakgrunnsinformasjon om elevene som deltar og foresatte bes derfor om å besvare skjema. Alle personopplysninger vil bli behandlet konfidensielt og vil kun være tilgjengelig for prosjektleder og veiledere i prosjektet. Data vil bli anonymisert og elever og skoler vil ikke kunne gjenkjennes ved publisering av materialet. Oppbevaring av data følger vanlige forskningsetiske retningslinjer  
Prosjektet skal etter planen avsluttes innen 1. september 2019.

Det er frivillig å delta i studien, og du kan når som helst trekke ditt samtykke uten å oppgi noen grunn. Dersom du trekker deg, vil alle opplysninger om deg bli anonymisert.

Dersom du har spørsmål til dette skjemaet eller til studien generelt, ta kontakt med prosjektleder Anne Cathrine Thurmman-Moe tlf.92610925 eller via mail til [anne-cathrine.thurmman-moe@statped.no](mailto:anne-cathrine.thurmman-moe@statped.no)

Studien er meldt til Personvernombudet for forskning, Norsk samfunnsvitenskapelig datatjeneste AS.

## BARNETS UTVIKLING

		Ja	nei
1	Var det spesielle problemer knyttet til svangerskap eller fødsel? Hvis ja, beskriv nærmere:		
2	Når begynte barnet å snakke?		
3	Når begynte barnet å gå?		

## SKJEMA FOR BAKGRUNNSINFORMASJON VED LESE OG SKRIVEVANSKER

Barnets navn..... Fødselsdato.....

4	<b>har du /dere hatt bekymring knyttet til barnets tidlige språkutvikling?</b> <b>nei</b> Hvis ja, beskriv nærmere:	<b>Ja</b>
5	<b>Ble det registrert problemer med uttale av ord i førskolealderen?</b> Hvis ja, beskriv nærmere:	<b>Ja</b> <b>nei</b>
6		
<b>BARNEHAGE/SKOLE</b>		
7	<b>Har barnet gått i barnehage?</b> Hvis ja, beskriv hvor lenge:	<b>Ja</b> <b>nei</b>

## SKJEMA FOR BAKGRUNNSINFORMASJON VED LESE OG SKRIVEVANSKER

Barnets navn..... Fødselsdato.....

8	<b>Har barnet byttet skole i løpet av skolegangen?</b> Hvis ja, beskriv nærmere:	Ja	nei
9	<b>Har barnet byttet lærer i løpet av skolegangen ?</b> Hvis ja, oppgi når og antall lærere med ansvar for leseopplæringen:	Ja	nei
10	<b>Når ble det oppdaget at barnet hadde lese og skrivevansker?</b>		
11	<b>Har barnet fått tilrettelegging for vanskene i barnehagen/ på skolen ?</b> Hvis ja, beskriv nærmere omfang og type tilrettelegging:	Ja	nei
<b>FAMILIEFORHOLD</b>			
12	<b>Hvem bor barnet sammen med?</b>		

## SKJEMA FOR BAKGRUNNSINFORMASJON VED LESE OG SKRIVEVANSKER

Barnets navn..... Fødselsdato.....

13	Fars yrke:  Fars utdanning (antall år):
14	Mors yrke:  Mors utdanning (antall år):
15	<b>Har barnet søsken?</b> <span style="float: right;"><b>Ja</b> <b>nei</b></span> Hvis ja, oppgi antall, alder på søsken og barnets plassering i søskenflokket.
16	<b>Har andre i barnets familie vansker knyttet til muntlig språk, lesing eller skriving?</b> <span style="float: right;"><b>Ja</b> <b>nei</b></span> Hvis ja, sett ring rundt familiemedlem og beskriv nærmere  far mor søster bror onkel tante kusine fetter farmor farfar mormor morfar



## SKJEMA FOR BAKGRUNNSINFORMASJON VED LESE OG SKRIVEVANSKER

Barnets navn..... Fødselsdato.....

## Elevens opplevelse av kartleggings situasjonen i leseprosjektet

Elev nr .....

Fylt ut av.....

## 1. Elevens generelle motivasjon for å delta i kartleggingen

1	2	3	4	5	6

## 2. Elevens generelle arbeidsinnsats i kartleggingsøktene

1	2	3	4	5	6

## 3. Utvikling av innsats under kartleggingen i løpet av prosjektperioden

Jevnt svak i hele perioden	Mest innsats i starten deretter avtagende	Liten innsats i starten , deretter økende	Jevnt god i hele perioden

## 4. Utvikling av motivasjon for kartleggingen i løpet av prosjektperioden

Jevnt svak i hele perioden	Mest innsats i starten deretter avtagende	Liten innsats i starten , deretter økende	Jevnt god i hele perioden

## 5. Har kartleggingen skjedd til samme tid hver uke? (sett kryss)

ja \_\_\_\_\_ nei \_\_\_\_\_

## 6. Har eleven vegret seg for å komme til kartleggingen?(sett kryss)

ja \_\_\_\_\_ nei \_\_\_\_\_

## 7. Har kartleggingen måttet utsettes fordi eleven har nektet å komme?(sett kryss)

ja \_\_\_\_\_ nei \_\_\_\_\_

## 8. Har eleven gitt verbalt uttrykk for uvilje mot kartleggingen? (sett kryss)

ja \_\_\_\_\_ nei \_\_\_\_\_

## 9. Har eleven gitt verbalt uttrykk for positiv holdning til kartleggingen?(sett kryss)

ja \_\_\_\_\_ nei \_\_\_\_\_

**Elevens opplevelse av kartleggingssituasjonen i leseprosjektet**

**Elev nr .....**

**Fylt ut av.....**

**Egne kommentarer:**

Elevens prosjektnr.....

Utfylt av -----

Skjema spør etter lærers vurdering av elevens motivasjon og arbeidsinnsats både for skolearbeid generelt og for spesialundervisning før tiltaket og en vurdering av det samme i prosjektperioden, altså de åtte ukene eleven har hatt 40 minutters økter fire ganger pr. uke. Det er også spørsmål knyttet til din faglige vurdering av elevens utbytte etter endt tiltak. Alle spørsmålene besvares ved å krysse av i tabellene som angir en skala fra 1 til 6 der 1 er lavest og 6 er høyest. Det er også satt av plass til egne kommentarer nederst!

**1. Elevens motivasjon for skolearbeid generelt**

1	2	3	4	5	6

**2. Elevens arbeidsinnsats når det gjelder skolearbeid generelt**

1	2	3	4	5	6

**3. Fravær generelt**

1	2	3	4	5	6

**4. Elevens motivasjon i forhold til lesing og skriving (før tiltaket)**

1	2	3	4	5	6

**5. Elevens arbeidsinnsats i forhold til lesing og skriving (før tiltaket)**

1	2	3	4	5	6

**6. Elevens motivasjon for spesialundervisningstimen ( før tiltaket)**

1	2	3	4	5	6

**7. Elevens arbeidsinnsats i timene med spesialundervisning(før tiltaket)**

1	2	3	4	5	6

Elevers prosjektnr.....

Utfylt av .....

**8. Elevers motivasjon for å delta i leseprosjektet**

1	2	3	4	5	6

**9. Elevers arbeidsinnsats i leseprosjekttimene**

1	2	3	4	5	6

**10. Elevers arbeidsinnsats i prosjekttimene sett i forhold til arbeidsinnsats i tidligere spesialtimer.**

Veldig mye dårligere	Mye dårligere	Litt dårligere	Omtrent som før	Litt bedre	Mye bedre	Veldig mye bedre

**11. Arbeidsinnsats i løpet av prosjektperioden**

Jevnt svak i hele perioden	Mest innsats i starten deretter avtagende	Liten innsats i starten , deretter økende	Jevnt god i hele perioden

**12. Generelt utbytte av opplæringen med PAS kortene?**

1	2	3	4	5	6

**13. I hvilken grad har opplæringen med PAS kortene ført til endring av elevens lesestrategi?**

1	2	3	4	5	6

**14. I hvilken grad har opplæringen med PAS kortene ført til at eleven leser med færre feil og mindre gjettinger?**

1	2	3	4	5	6

Elevens prosjektnr.....

Utfylt av -----

**15. I hvilken grad har opplæringen med PAS kortene ført til at eleven har fått bedre leseflyt?**

1	2	3	4	5	6

**16. I hvilken grad har opplæringen med PAS kortene ført til at eleven skriver med færre stavefeil?**

1	2	3	4	5	6

**17. Har flere enn en lærer deltatt i undervisningen med PAS kortene? sett kryss**

Nei

Ja

Hvis ja , angi hvor mange lærere og i hvor mange økter. ....

**Egne kommentarer :**

**Vurderingsskjema for elever i leseprosjekt**

**Appendiks O**

Elevens prosjektnr.....

Utfylt av -----

