



Kunnskapen du trenger



NORM-prosjektet
Rapport 2

Ble elevene bedre skrivere?

Intervensjonseffekter på elevers
skriveferdigheter og skriveutvikling

Kjell Lars Berge & Gustaf Skar

Copyright © Høgskolen i Sør-Trøndelag, Normprosjektet og forfatterne, 2015
All rights reserved

ISBN: 978-82-7877-247-8

Tittel: Ble elevene bedre skrivere? Intervensjonseffekter på elevers skriveferdigheter og skriveutvikling.

Forfattere: Professor Kjell Lars Berge (Universitetet i Oslo), k.l.berge@iln.uio.no
Dr Gustaf Skar (Høgskolen i Sør-Trøndelag, Skrivesenteret), gustaf.b.skar@hist.no

NORM-PROSJEKTET

Fullstendig tittel: *Developing national standards for the assessment of writing. A tool for teaching and learning*, omtalt som *Normprosjektet*. Nettsted med prosjektbeskrivelse og omtaler: www.norm.skrivesenteret.no.
Prosjektet er finansiert av Norges forskningsråd og Høgskolen i Sør-Trøndelag.

Prosjektledelse: Professor Synnøve Matre og førsteamanuensis Randi Solheim (Høgskolen i Sør-Trøndelag)

Forskergruppe: Professor Lars Sigfred Evensen og professor Hildegunn Otnes (NTNU), professor Ragnar Thygesen (UiA), professor Kjell Lars Berge (UiO), dr Gustaf Skar (HiST, Skrivesenteret), professor Synnøve Matre og førsteamanuensis Randi Solheim (HiST)

Prosjektkoordinator: Jorun Smemo (Høgskolen i Sør-Trøndelag/Skrivesenteret)

Stipendiater: Sindre Dagsland (HiST/NTNU), Trine Gedde-Dahl (HiST/NTNU/HiOA) og Jannike Ohrem Bakke (HBV)

Denne rapporten: Berge, K. L. & Skar, G. (2015). *Ble elevene bedre skrivere? Intervensjonseffekter på elevers skriveferdigheter og skriveutvikling*. Rapport 2 fra Normprosjektet. Trondheim: Høgskolen i Sør-Trøndelag, Avdeling for lærer- og tolkutdanning.

Innholdsfortegnelse

Prosjektledelsens forord	5
Forfatterens forord	6
Sammendrag	7
1 Normprosjektets mål	9
2 Normprosjektets forskningsdesign	10
2.1 Intervensjon og utvalg	10
2.2 Vurderingshendelser/målepunkt: tekstvurderinger av intensivutvalgene.....	13
2.3 Eksperimentdesign: prosjektskoler og kontrollskoler	18
3 Resultatpresentasjon: Har intervensjonen fra Normprosjektet hatt effekt på elevenes skriveferdigheter?	21
3.1 Elevenes skriveutvikling gjennom intervensjonen hele utvalget samlet	21
3.2 Elevenes skriveutvikling gjennom intervensjonen fordelt på skoler og trinn	24
3.3 Effekten av intervensjonen	29
3.3.1 <i>Urval</i>	29
3.3.2 <i>Metod</i>	30
3.3.3 <i>Resultat</i>	32
3.4 Variasjon og stabilitet i elevenes skriveferdigheter	38
3.4 Forholdet mellom skriveferdighetsnivå og eksterne variabler (kjønn, morsmål, målform, fødselsmåned)	41
3.4.1 <i>Forholdet mellom kjønn og skriveferdighet</i>	42
3.4.2 <i>Forholdet mellom morsmål og skriveferdighet</i>	44
3.4.4 <i>Forholdet mellom målform og skriveferdighet</i>	45
3.4.5 <i>Forholdet mellom fødselsmåned på trinnet og skriveferdighet</i>	46
4 Drøfting av og videre forskning på funnene i Normprosjektet	48
4.1 Intervensjonens læringseffekt: en oppsummering	48
4.2 Mulige forklaringer på intervensjonens læringseffekt: eksterne variabler	49
4.3 Mulige forklaringer på intervensjonens læringseffekt: videre forskning	50
Litteratur	53

Tabeller og figurer

Tabell 1. Projektskoler i Normprosjektet	10
Tabell 2. Normprosjektet: ekstensiv- og intensivutvalg og intervensjonsperiode for utvalget.....	11
Tabell 3. Intensivlevens kjønn (%)	12
Tabell 4. Intensivlevens målform (%)	12
Tabell 5. Intensivlevens morsmål (%)	12
Tabell 6. Normprosjektet: forholdet mellom pre- og posttest, og mellom intervensjonsfaser	14
Tabell 7. Reliabilitet posttest (MP4), høsten 2013.....	17
Tabell 8. Reliabilitet posttest (MP7), høsten 2014.....	17
Tabell 9. Intensivlevens aldersfordeling	18
Tabell 10. Normprosjektet: pretester for projektskoler og kontrollskoler (n elever).....	19
Tabell 11. Normprosjektet: posttester for projektskoler og kontrollskoler (n elever)	19
Tabell 12. Snitt (m) basert på sumskåre på alle mulige målepunkt, alle trinn	21
Tabell 13. ANOVA-analyser av ferdighetsutviklingen for 300- og 600-elever gjennom to års intervensjon.....	22
Tabell 14. 300-elevens resultater i sumskåre fordelt på intervensjonsskoler	24
Tabell 15. Skoler med påviselig utvikling blant 300-elevne, to intervensjonsår sett under ett.....	25
Tabell 16. Skoler med svak og/eller ingen påviselig utvikling blant 300-elevne	26
Tabell 17. 600-elevens resultater fordelt på intervensjonsskoler	27

Tabell 18. Skoler med påviselig utvikling blant 600-elevene, to intervensjonsår sett under ett	27
Tabell 19. Skoler med svak og/eller ingen påviselig utvikling blant 600-elevene	28
Tabell 20. Primært og sekundært urval og bortfall	30
Tabell 21. Riktvråden för effektmått.	32
Tabell 22. Skillnader <i>inom</i> respektive grupp mellan MP1 och MP7, parvisa t-tester.....	32
Tabell 23. Skillnader mellan grupper, ANCOVA-analyser.....	33
Tabell 24. Skillnader mellan subgrupper, 300-elever, ANCOVA-analyser.	34
Tabell 25. Skillnader mellan subgrupper, 600-elever, ANCOVA-analyser.	34
Tabell 26. Skillnader mellan kontrollskolor och särskilt urval, 600-elever.....	35
Tabell 27. Starka, moderata och svaga effektstrolekar.	36
Tabell 28. Skillnader mellan 300- och 400-elever och 600- och 700-elever	37
Tabell 29. Skillnader <i>mellan</i> respektive grupp mellan på MP7	37
Tabell 30. Starka, moderata och svaga effektstorlekar	38
Tabell 31. Förklarad varians	38
Tabell 32. Mann-Whitney U-test: kjønnsforskjeller i skriveprestasjon for 300-prosjektelever	42
Tabell 33. Mann-Whitney U-test: kjønnsforskjeller i skriveprestasjon for 300-kontrollskoleelever.....	42
Tabell 34. Mann-Whitney U-test: kjønnsforskjeller i skriveprestasjon for 600-prosjektelever.....	43
Tabell 35. Mann-Whitney U-test: kjønnsforskjeller i skriveprestasjon for 600-kontrollskoleelever.....	43
Tabell 36. Mann-Whitney U-test: språkforskjeller i skriveprestasjon for 300-prosjektelever.....	44
Tabell 37. Mann-Whitney U-test: språkforskjeller i skriveprestasjon for 300-kontrollskoleelever	44
Tabell 38. Mann-Whitney U-test: språkforskjeller i skriveprestasjon for 600-prosjektelever.....	44
Tabell 39. Mann-Whitney U-test: språkforskjeller i skriveprestasjon for 600-kontrollskoleelever	45
Tabell 40. Mann-Whitney U-test: målformforskjeller i skriveprestasjon for 300-prosjektelever	45
Tabell 41. Mann-Whitney U-test: målformforskjeller i skriveprestasjon for 300-kontrollskoleelever	45
Tabell 42. Mann-Whitney U-test: målformforskjeller i skriveprestasjon for 600-prosjektelever	46
Tabell 43. Mann-Whitney U-test: målformforskjeller i skriveprestasjon for 600-kontrollskoleelever	46
Tabell 44. Kruskal-Wallis-test: fødselsmånedforskjeller i skriveprestasjon for 300-prosjektelever.....	46
Tabell 45. Kruskal-Wallis-test: fødselsmånedforskjeller i skriveprestasjon for 300-kontrollskoleelever.....	46
Tabell 46. Kruskal-Wallis-test: fødselsmånedforskjeller i skriveprestasjon for 600-prosjektelever	47
Tabell 47. Kruskal-Wallis-test: fødselsmånedforskjeller i skriveprestasjon for 600-kontrollskoleelever.....	47
Figur 1. Skrivehjulet: skrivehandlinger og formål.....	15
Figur 2. Skrivehjulet: De semiotiske medieringsressursene	16
Figur 3. Prestasjonsutvikling for 300-elevene: sumskåresnitt og spredning (SD)	22
Figur 4. Prestasjonsutvikling for 600-elevene: sumskåresnitt og spredning (SD)	23
Figur 5. Varians för 300- och 600-elever alla målepunkter förklarad.....	39
Figur 6. Individuell variation bland 300-elever	40
Figur 7. Individuell variation bland 600-elever	40

Prosjektledelsens forord

Ei av hovudproblemstillingane i Normprosjektet handlar om kva verknad bruken av felles forventningsnormer har på elevars skrivning. Svar på dette spørsmålet finn vi i elevtekstar og i lærarars vurdering av desse tekstane, og det er resultat frå kvantitative analysar av denne empirien som vert presentert i denne rapporten. Elevtekstane som ligg til grunn, er skrivne i perioden 2012–2014 innanfor ein intervensjonstudie der lærarar på 20 skular, under rettleiing av forskarane i Normprosjektet, arbeidde med skrivning i ulike fag. Arbeidet var forankra i eit funksjonelt skrivekonstrukt og eksplisitte forventningsnormer for skrivning.

Forventningsnormene, som er ein sentral pilar i denne skriveopplæringa, blei utvikla i første fase av prosjektet (sjå Rapport 1 i Normprosjektets rapportserie og www.norm.skrivesenteret.no).

Kjell Lars Berge og Gustaf Skar har leia an i analysearbeidet med entusiastisk pågangsmot. Dei to har i fellesskap gjennomført analysane og ført rapporten i pennen. Elles har heile forskargruppa i Normprosjektet bidratt til rapporten gjennom innsamling av materiale, deltaking i omfattande diskusjonar og kritisk lesing av utkast. I tillegg har eksterne fagfellar, nasjonale og internasjonale, utfordra oss med ulike innspel i samband med presentasjonar på seminar og konferansar. Slike faglege diskusjonar har vore svært nyttige og bidratt til ytterlegare nyanseringar i analysearbeidet.

Dei kvantitative resultatane som blir presenterte i rapporten, leier til mange spørsmål og hypotesar om elevars skrivekompetanse, og dei peikar fram mot moglege forklaringar. Forskargruppa følgjer nå opp desse funna gjennom ei rekkje kvalitative studiar der vi går tett inn på utvalde elevar og skriveforløp, oppgåveformuleringar, tekstar, observasjonar frå klasserom og intervju med lærarar og leiarar, for såleis å kome nærast mogleg ei nyansert forståing av kvifor elevskrivarane utviklar seg slik dei gjer. Prosjekta til dei tre doktorgradsstipendiatane i Normprosjektet vil vere særleg viktige her. Tekstanalysar som ei stor gruppe satellitt-forskarar arbeider med, vil også supplere funna våre på verdifull måte. I tillegg er resultatane i denne rapporten interessante sett i lys av pågåande kvalitative og kvantitative studiar av læraranes vurderingskompetanse.

Trondheim 4. juni 2015

Synnøve Matre, prosjektleiar & Randi Solheim, nestleiar

Forfatternes forord

Rapporten *Ble elevene bedre skrivere? Intervensjonseffekter på elevers skriveferdigheter og skriventvikling* er skrevet av professor Kjell Lars Berge og forsker Gustaf Skar i nært samarbeid. Berge og Skar har også gjennomført de statistiske analysene som er presentert i rapporten.

Designet for studien er utarbeidet av forskergruppa i fellesskap, og de står også kollektivt bak gjennomføring av studien og innsamling av data.

Grunnlaget for de statistiske analysene i rapporten er bearbeidet av prosjektkoordinator Jorun Smemo med hjelp av Anne Holten Kvistad, Eli Muus og Pia Farstad Eriksen. Vi takker for det omfattende og grundige arbeidet de har gjennomført for å legge til rette for de statistiske analysene som er dokumentert i rapporten. Dette registreringsarbeidet har vært meget krevende, men er gjennomført med en sjelden grundighet og nøyaktighet.

Underveis har førstelektor Anne Charlotte Begnum, Lesesenteret/UiS, vært en viktig støttespiller i arbeidet med å forberede dataene for analyse. Hun hjalp til med de første utarbeidingene av datamatriser og kodebøker, og med å lage renseprotokoller og kvalitetssikre datagrunnlaget. Ikke minst spilte hun en viktig rolle som veileder for prosjektkoordinator Jorun. Vi takker Anne Charlotte for faglig solid og inspirerende assistanse.

Docent Michael Tengberg, Karlstads universitet, har kvalitetssikret de statistiske analysene i nært samarbeid med Kjell Lars Berge og Gustaf Skar. Vi takker Michael for uvurderlig hjelp med analysene.

Prosjektmedarbeiderne og -kollegene professor Synnøve Matre (HiST), førsteamanuensis Randi Solheim (HiST), prosjektkoordinator Jorun Smemo (HiST, Skrivesenteret), professor Lars Sigfred Evensen (NTNU), professor Ragnar Thygesen (UiA), stipendiat Sindre Dagsland (NTNU) og stipendiat Trine Gedde-Dahl (HiST/NTNU/HiOA) har lest, kommentert og kvalitetssikret rapporten. Vi takker de alle sammen for gode, lærerike og kvalitetsforbedrende lesninger.

Rapporten er nr. 2 i Normprosjektets rapportserie.

Oslo og Trondheim 4. juni 2015

Kjell Lars Berge & Gustaf Skar

Sammendrag

Følgende problemstillinger i prosjektet *Developing national standards for the assessment of writing. A tool for teaching and learning* («Normprosjektet», 2012–2016) er analysert, presentert og diskutert i denne rapporten:

- Om elevene på prosjektskolene etter Normprosjektets intervensjon skriver bedre målt i mulig prestasjonsendring ved ulike skrivehendelser gjennom intervensjonen.
- Om elevene på prosjektskolene etter Normprosjektets intervensjon skriver bedre enn elever fra kontrollskoler som ikke har gjennomført intervensjonen.
- Dersom det er effekt av Normprosjekt-intervensjonen, hvor sterk er effekten?
- Om variasjonen i skriveferdighet i løpet av og etter Normprosjektets intervensjon predikeres av eksterne variabler som intensivelevens individuelle egenskaper, trinntilhørighet, fødselsmåned, kjønn, målform, språk og skole – samlet eller enkeltvis.

Det er elever som har deltatt i Normprosjektet i to undervisningsår, som er studert i denne rapporten. Det gjelder elever som begynte i 3. trinn skoleåret 2012/2013 (såkalte 300-elever), og elever som begynte i 6. trinn skoleåret 2012/2013 (såkalte 600-elever). Disse elevenes skriveferdigheter er sammenliknet med en gruppe elever fra fire kontrollskoler på tilsvarende trinn som ikke har deltatt i Normprosjektet.

Det overordnede funnet som rapporteres i denne rapporten fra Normprosjektet, er at prosjektets intervensjon gir en effekt på elevenes skriveferdighet som tilsvarer fra et halvt til to undervisningsår, avhengig av hvilken elevgruppe som undersøkes. Effektene varierer med subgruppene:

Effekten på 300-elevenes skriveferdigheter:

- På de fleste prosjektskoler har intervensjonen signifikant læringseffekt for 300-elevne.
- Effekten ved å delta i prosjektet innebærer for 300-elevne samlet minst ett års ekstra undervisning i forhold til kontrollskoleelevne ($p < .05$).
- Det er en viss variasjon i hvilke undergrupper av 300-elever som i størst grad har utviklet skriveferdighetene gjennom å delta i prosjektet. Elevene som ved pretest (MP1) presterer omkring normen, har hatt den sterkeste ferdighetsutviklingen. For denne undergruppa av 300-elever, tilsvarer effekten av å delta i prosjektet opp mot to års ekstra undervisning i forhold til kontrollskoleelevne.

Effekten på 600-elevenes skriveferdigheter:

- Elever som er fulgt gjennom 6. og 7. skoleår, har bedret egne skriveferdigheter signifikant ved å delta i Normprosjektet. For 600-elevne som gruppe betraktet, tilsvarer effekten minst ett halvt års ekstra undervisning i forhold til kontrollskoleelevne ($p > .10$).
- Det er stor variasjon i hvilke undergrupper av 600-elever som har utviklet skriveferdighetene gjennom å delta i Normprosjektet. Det er særlig ved seks skoler (Aura, Casa, Engen, Marka, Lia og Raset) at elevene har utviklet ferdighetene i betydelig

grad. For elevene på disse skolene tilsvarer effekten opp mot to års ekstra undervisning i forhold til kontrollskolelevener.

- Ved fire skoler (Dalen, Heia, Kilden, Quadra) har ingen av intensivelevener fra 600-gruppen påviselig effekt av å delta i prosjektet.

I rapporten er det undersøkt om elevenes ferdighetsutvikling kan forklares med elevenes skriveferdigheter før intervensjonen. Det er betydelig variasjon i elevenes skriveferdigheter fra skrivehendelse til skrivehendelse:

- For 300-elevener kan mellom 13 og 37 % av variansen i skriveprestasjonene gjennom intervensjonens to år forklares på grunnlag av hvordan de skrev før intervensjonen ble gjennomført.
- For 600-elevener kan mellom 15 og 23 % av variansen i skriveprestasjonene gjennom intervensjonens to år forklares på grunnlag av hvordan de skrev før intervensjonen ble gjennomført.

I rapporten er det undersøkt om eksterne variabler som elevens kjønn, fødselsmåned, målform og morsmål samlet eller enkeltvis kan forklare elevenes skriveferdighetsnivå. Følgende resultat viser seg:

- Elevenes skriveferdigheter samvarierer ikke systematisk med elevens *kjønn* bortsett fra i de fleste testsituasjoner (pretest og posttest). For 600-elevener del er det også en tendens til at jentene skriver bedre mot slutten av og etter intervensjonen.
- Elevenes skriveferdigheter samvarierer ikke systematisk med *når på året eleven er født*.
- Elevenes skriveferdigheter samvarierer ikke systematisk med elevenes *målform*.
- Elevenes skriveferdigheter samvarierer ikke systematisk med elevenes *morsmål*, unntatt ved enkelte målepunkter/vurderingshendelser.

Undersøkelsene av forholdet mellom elevenes skriveferdigheter før, gjennom og etter hele intervensjonsperioden og eksterne variabler, viser at de fire undersøkte eksterne variablene samlet eller enkeltvis ikke kan forklare effekten av å delta i Normprosjektet. Mer kvalitativt orienterte studier vil kunne gi andre og muligens mer relevante forklaringer på resultatene som er presentert i denne rapporten. Framtidige, mer kvalitativt orienterte studier ville kunne belyse blant annet følgende spørsmål:

- Hvordan er Normprosjektet blitt implementert på skolene og i klassene av skoleledelse og de deltakende lærerne?
- Hvordan har skriveopplæringen og vurderingsarbeidet på skolene blitt gjennomført og utviklet gjennom intervensjonen fra Normprosjektet?
- Hva kjennetegner skriveutviklingen i intensiveleveneres tekster?

1 Normprosjektets mål

Gjennom prosjektet *Developing national standards for the assessment of writing. A tool for teaching and learning* 2012–2016 (heretter kalt «Normprosjektet») skal det utvikles forskningsbasert kunnskap om hvilke skriveferdigheter samfunnet kan forvente at elever har utviklet og mestrer etter fire og sju års opplæring i grunnskolen. Prosjektet skal dokumentere hvordan disse forventningene kan uttrykkes som normer for tilfredsstillende skriveferdigheter. I tett samarbeid med skoler over hele landet er det gjennom Normprosjektet utviklet nasjonale *forventningsnormer* for skrivekompetanse. Disse forventningsnormene ble avdekket, formulert og utviklet gjennom et NFR-finansiert forprosjekt i 2010 (Matre et al., 2011), som inngår i fase 1 av Normprosjektet. Forventningsnormene artikulere på en systematisk måte hvilke krav som bør stilles til elever som har gjennomført henholdsvis fire eller sju års skriveopplæring. Slik uttrykker de hva elever bør mestre av skriveferdigheter ved inngangen til 5. trinn (mellomtrinnet) og 8. trinn (ungdomstrinnet).

Forventningsnormene har dannet utgangspunkt for fase 2 av Normprosjektet, der skrijving som grunnleggende ferdighet i grunnskolen er studert systematisk. Forventningsnormene er i fase 2 blitt tatt i bruk som redskap for skriveopplæring og vurdering, sammen med en rekke andre vurderingsressurser, gjennom et omfattende intervensjonsprosjekt på 20 skoler over hele landet.

I fase 2 av Normprosjektet skal det gis svar på følgende overordnede forskningsspørsmål:

1. Har bruken av felles forventningsnormer noen effekt på prosjektelevenes skriveferdigheter?
2. Har bruken av felles forventningsnormer noen effekt på prosjektlerernes vurderinger av elevtekster?

I denne rapporten presenteres resultater av statistiske analyser der det gis svar på det første av de to forskningsspørsmålene. I rapporten vil det også bli redegjort for statistiske analyser som kan bidra til å forklare resultatene av effektanalysene. Rapporten avsluttes med en diskusjon av funnene. Det presenteres også innspill til videre forskning for å kunne avdekke mulige forklaringer på resultatene som presenteres.

For lettere å kunne forstå resultatene av de statistiske analysene, gir vi i kapittel 2 av rapporten en oversikt over Normprosjektets forskningsdesign. Det vil synliggjøre under hvilke betingelser den eventuelle effekten av Normprosjektet på elevenes skrivekompetanse er studert og målt.

2 Normprosjektets forskningsdesign

2.1 Intervensjon og utvalg

I Normprosjektets fase 2 har det deltatt 20 såkalte *prosjektskoler*. Det er elever fra 3., 4., 6. og 7. trinn på de deltakende skolene som har inngått i prosjektet, og som har skrevet tekster gjennom prosjektperioden. I alt har 3088 elever deltatt i prosjektet. I tillegg har det deltatt om lag 500 lærere. Rektorene eller andre representanter fra ledelsen ved de tjue skolene har også deltatt aktivt i prosjektet.

I tillegg har det deltatt fire såkalte *kontrollskoler*. Elevene fra kontrollskolene har skrevet tekster ut fra de samme skriveoppgavene som prosjektskoleelevene, både før og etter prosjektet (se kapittel 2.2). Ut over dette bidraget har de ikke hatt kontakt med Normprosjektet eller hatt kjennskap til arbeidet på prosjektskolene.

De deltakende prosjekt- og kontrollskolene fordeler seg over hele landet. Fordelingen mellom landsdelene er oppstilt i tabell 1. Alle skoler er anonymisert og gitt et prosjektnavn. Navnene framgår av tabell 1:

Tabell 1. Prosjektskoler i Normprosjektet

Landsdel	Prosjektskoler	Kontrollskoler ^a	N skoler
Oslo	Aura, Casa, Engen, Jordet		4
Øvrige Østlandet	Dalen, Heia, Isen, Neset	Wang	5
Vestlandet/Sørlandet	Fjellet, Plassen, Quadra, Raset, Stranda	Zalen, Ytre	7
Trøndelag	Bukta, Lia, Marka, Osen, Teigen	Varden	6
Nord-Norge	Gropa, Kilden		2

^aOpprinnelig var fem skoler involvert som kontrollskoler. En av kontrollskolene trakk seg fra prosjektet før prosjektstart.

I prosjektet kalles elevene som gikk på 3. trinn da prosjektet startet *300-elever*, elevene som gikk på 4. trinn da prosjektet startet *400-elever*, elevene som gikk på 6. trinn da prosjektet startet *600-elever* og elevene som gikk på 7. trinn da prosjektet startet *700-elever*.

Prosjektet er et *intervensjonsprosjekt*, der forventningsnormene og et skrivekonstrukt (Berge, Evensen, & Thygesen, under publisering; Evensen, 2010) der skrivekonstrukt forstås som grunnleggende for læring i alle fag og på tvers av alle fag, er innført på de tjue deltakende skolene (Berge & Thygesen, 2015; Matre & Solheim, 2014). Forventningsnormene er delt opp i sju ulike vurderingsområder, og i ulike kvalitetsnivåer fordelt på en skala fra 1 til 5 (se nedenfor). De sju vurderingsområdene er *kommunikasjon*, *innhold*, *tekstopbygging*, *språkebruk*, *rettskrivning* og *formverk*, *tegnsetting* og *bruk av skriftmediet*.

Forventningsnormene er utviklet for to ulike nivå og angir hva som er rimelig å forvente av elever etter fire og sju års skriveopplæring. Normene fanger dermed opp forventninger til elevenes skriveferdighetsnivå enten ved starten av 5. trinn (mellomtrinnet) eller ved starten av 8. trinn (ungdomstrinnet).

Som en del av intervensjonen har lærerne på 3., 4., 6. og 7. trinn på de deltakende skolene blitt skolert gjennom et omfattende kursopplegg. For å få en så tett kopling til lærernes hverdagsutfordringer og -praksis som mulig, har skoleringen vært lagt til skolene og inngått i lærernes faste arbeidstid, og til samlinger. Det har vært en grunnleggende forutsetning for prosjektet at det skal oppfylle krav til økologisk validitet (Johnson & Christensen, 2000). Det innebærer blant annet at intervensjonen skal oppleves som en del av det fortløpende arbeidet med å undervise elevene etter målene i gjeldende læreplan. For mange av skolene har Normprosjektet blitt brukt som et kompetanseoppbyggende og kunnskapsutviklende prosjekt for å styrke kvaliteten på skolens skriveopplæring og vurderingsarbeid. Prosjektet har således også en utviklingsdimensjon.

I løpet av skoleringen har lærere på de deltakende skolene fått innføring i forståelsen av skrivekonstrukt og forventningsnormer. De har blitt instruert i hvordan de kan lage skriveoppgaver på grunnlag av konstruktet og gjeldende læreplaner. Prosjektlærerne har lært hvordan de kan vurdere elevenes tekster formativt/læringsstøttende og summativt/oppsummerende ved hjelp av forventningsnormene. Intervensjonen på skolene tok til høsten 2012 og ble avsluttet våren 2014.

Intervensjonsperioden varte noe mindre enn to skoleår for elevgruppa som omtales som 300- og 600-elevene (dvs. elevene som begynte på 3. og 6. trinn skoleåret 2012/13). De såkalte 400- og 700-elevene (elevene som begynte på 4. og 7. trinn 2012/13) har kun deltatt i intervensjonen i ett år. 400- og 700-elevene skal etter prosjektplanen ha skrevet minst seks tekster i prosjektregi. 300- og 600-elevene skal etter prosjektplanen ha skrevet minst tolv prosjektttekster.¹ Alle disse er samlet inn.

Tekstene som elevene har skrevet, er vurdert på skolene av trinnlærerne. Lærerne har blitt oppfordret til å vurdere tekstene både som *utkast* og som endelige tekster, såkalte *presentasjonstekster*. Gjennom intervensjonen har de blitt oppfordret til å vurdere elevenes tekster både med formativt og summativt formål (Hattie & Timperley, 2007; Matre & Solheim, 2014). I begge tilfeller har vurderingene skjedd innenfor rammene av skolens hverdagslige undervisningsoppdrag, men likevel på en slik måte som de etter skolering har blitt bedt om å gjøre. Det totale antallet elever (3088) utgjør *ekstensivutvalget*, og tekstene de har skrevet omtales som *ekstensivutvalgte* tekster.

I utvalget på 3088 elever har 379 elever blitt trukket ut for å vurderes og studeres mer inngående. Dette *intensivutvalget* utgjør 12,3% av alle deltakende prosjektelever. Det er effekten av intervensjonen for intensivutvalget vi presenterer i denne rapporten. En oversikt over omfanget av tekster fra ekstensiv- og intensivutvalg gis i tabell 2. Dessuten gis det en oversikt over antallet elever på ulike trinn fra kontrollskolene som har skrevet tekster i forbindelse med pre- og posttesten. I tabellen gis det også en oversikt over intervensjonsperiodens omfang fordelt på elevgrupper:

Tabell 2. Normprosjektet: ekstensiv- og intensivutvalg og intervensjonsperiode for utvalget

	3. trinn: 300-elever	4. trinn: 400-elever	6. trinn: 600-elever	7. trinn: 700- elever	N	N%
Ekstensivutvalg	827	748	697	816	3088	100
Intensivutvalg	97	85	97	100	379	12,3
Intervensjonsperiode	To år: 2012–2014	Ett år: 2012–2013	To år: 2012–2014	Ett år: 2012–2013		
Kontrollskoleutvalg	67	73	67	26	233	100

¹ Antallet tekster den enkelte elev har skrevet, varierer noe. Variasjonen skyldes forhold ved de enkelte skolene. Vi drøfter ikke mulige årsaker til bortfall og liknende i denne presentasjonen.

Elevene i intensivutvalget er trukket tilfeldig. De utgjør et såkalt «randomisert» utvalg. Det innebærer at alle de deltakende 3088 elevene ved trekking av intensivutvalget har hatt like stor sannsynlighet for å komme med i intensivutvalget. Intensivutvalget ble trukket i januar 2013, etter at de fleste elevene hadde skrevet to tekster på skolen etter første intervensjonsøkt høsten 2012. Tekstene som intensivelevnene har skrevet i løpet av, og som en del av prosjektet, er blitt transkribert og gjort digitalt tilgjengelig for videre forskning på tekst- og språkkvalitet.²

Elevene i intensivutvalget er registrert med fødselsmåned, kjønn, morsmål og målform. Morsmålet er bestemt som det første språket eleven lærte som barn. Utvalget er dermed delt i norskspråklige og L2-elever³. Målform er bestemt som det opplæringspråket som brukes i undervisningen på skolen, enten nynorsk eller bokmål. Oversikter over intensivutvalgets elever fordelt på målform, kjønn og morsmål gis i tabellene 3–5:

Tabell 3. Intensivelevnes kjønn (%)

	3. trinn	4. trinn	6. trinn	7. trinn	N	N%
Gutter	41	43	49	47	180	47,5
Jenter	56	42	48	53	199	52,5

Vi ser at det er en liten overvekt av jenter i prosjektet i forhold til populasjonen.

Tabell 4. Intensivelevnes målform (%)

	3. trinn	4. trinn	6. trinn	7. trinn	N	N%
Bokmål	80	70	80	83	313	82,6
Nynorsk	17	15	17	17	66	17,4

Det er en viss overvekt av elever med nynorsk som hovedmål i utvalget, sett i forhold til populasjonen. Per 2014 er det det 12,6 % av elevene i norsk grunnskole som har nynorsk som første opplæringspråk (PROBA Samfunnsanalyse, 2014). Et slikt utvalg gjør det mulig å studere nynorskelevnes situasjon spesifikt.

Tabell 5. Intensivelevnes morsmål (%)

	3. trinn	4. trinn	6. trinn	7. trinn	N	N%
Norsk	80	73	84	84	321	84,7
Minoritetspråk (L2)	13	10	10	13	46	12,1
Begge	4	2	3	3	12	3,2

Prosentandelen L2-elever i Normprosjektet er om lag dobbelt så høy som i populasjonen, dersom vi legger prosentandelen av elever som har hatt særskilt norskopplæring i skoleårene 2011–2014 til grunn. Dette muliggjør også særskilte studier. I følge Utdanningsdirektoratet har om lag 7 % av elevene i grunnskolen særskilt norskopplæring (Utdanningsdirektoratet, 2015). Hvorvidt elevene i prosjektet er definert som L2-elever eller ikke, er bestemt på grunnlag av selvrapporing. De har blitt bedt om å krysse av for om norsk eller annet språk enn norsk var

² Tekstene for intensivutvalget er lagret nettstedet til Tekstlaboratoriet ved Institutt for lingvistiske og nordiske studier, Universitetet i Oslo: <http://www.hf.uio.no/iln/om/organisasjon/tekstlab/>

³ Andre mulige betegnelser er "minoritetsspråklige elever" eller "andrespråkelever".

det første språket de lærte. Vi ser at det er en liten gruppe elever (3,2 %) som krysser av for begge alternativer.

2.2 Vurderingshendelser/målepunkt: tekstvurderinger av intensivutvalgene

Som nevnt, skal elevene som utgjør ekstensivutvalget etter prosjektplanen ha skrevet og levert minimum seks (400- og 700-elever) eller tolv tekster (300- og 600 elever) i løpet av prosjektperioden. For intensivelevens del har to (400- og 700-elever) eller fire (300- og 600 elever) av de tekstene elevene har skrevet som en del av skriveopplæringen på skolene, blitt vurdert av prosjektledere fra alle de andre prosjektskolene. Et utvalg på om lag 80 prosjektledere fra alle tjue skoler har vært samlet på nasjonale vurderingssamlinger fire ganger i løpet av prosjektperioden: januar 2013, mai 2013, januar 2014 og mai 2014. På 2013-samlingene har tekster skrevet av elever fra alle fire involverte klassetrinn blitt vurdert. På 2014-samlingene har tekster skrevet av 300-elever og 600-elever blitt vurdert.

På de nasjonale samlingene har prosjektlederne vurdert tekstene i par. Prosjektlederne som har inngått i parene, har ikke vurdert egne elever eller elever fra skolen de selv underviser på. Tekstene skrevet av 300- og 400 elever, har blitt vurdert av lærere med erfaring og tilhørighet på småskoletrinnet (1.–4. trinn). Tekstene skrevet av 600- og 700-elever, har blitt vurdert av lærere med erfaring og tilhørighet på mellomtrinnet (6.–7. trinn). Når tekster skrevet av 300- og 400-elever er blitt vurdert, har forventningsnormene som gjelder etter fire års opplæring blitt lagt til grunn. Når tekstene skrevet av 600- og 700-elever er blitt vurdert, er forventningsnormene etter sju års opplæring brukt. Tekster skrevet av 300- og 400 elever er blitt blandet slik at de som har vurdert, ikke har kjent til hvilket trinn elevene har gått på da tekstene ble skrevet. Det samme gjelder tekster skrevet av 600- og 700- elever. Som vi skal se i kapittel 3.2 i rapporten, har denne anonymiseringen av elevens trinn da eleven skrev teksten, vært viktig for å kunne utvikle et mål på forskjellen i skrivekvalitet mellom 300- og 400-elevene på den ene sida og 600- og 700-elevene på den andre.

Både prosjektskoleelever og kontrollskoleelever har skrevet to tekster før intervensjonen startet, og én tekst etter at intervensjonen ble avsluttet. Disse tekstene utgjør inngangs- og utgangsdata i prosjektet. Slik er det mulig å studere om det kan påvises en systematisk forskjell i skriveferdighet fra før intervensjonen starter til etter at intervensjonen er avsluttet, både mellom kontrollskoleelever og prosjektskoleelever og internt blant prosjektskoleelevne.

Prosjektskoleelevne og kontrollskoleelevne har altså skrevet to tekster *før* intervensjonen tok til høsten 2012. Noen måneder *etter* at intervensjonen på skolene ble avsluttet, har intensivelevne fra prosjektskolene og kontrollskoleelevne tatt *utvalgsprøvene i skrijving*⁴. For 400- og 700-elevne, som kun har inngått i intervensjonen i ett år, innebærer det at de har skrevet utvalgsprøven høsten 2013. Det var utvalgsprøven i skrijving fra 2012 som ble brukt på posttesten (MP4) for disse elevne. For 300- og 600-elevne som har inngått i intervensjonen i to år, innebærer det at de har skrevet utvalgsprøven høsten 2014.

De to tekstene som intensivelevne har skrevet *før* de startet med prosjektet høsten 2012, og utvalgsprøvetekstene som er skrevet *etter* at prosjektet på skolene er avsluttet enten høsten 2013 (400- og 700-elevne) eller høsten 2014 (300- og 600-elevne), er alle vurdert av enten to eller ett par fra det nasjonale vurderingspanelet som er spesielt skolert for å vurdere de nasjonale

⁴ For informasjon om utvalgsprøvene i skrijving, se informasjon på Skrivesenterets nettsider: www.skrivesenteret.no/skriveprva-emneside

utvalgsprøvene i skrivning. Utvalget av tekster som har blitt vurdert av vurderingspanelet før intervensjonen på prosjektskolene, utgjør en såkalt *pretest*. Pretesten gir informasjon om nivået på prosjektelevenes skriveferdigheter *før* intervensjonen. Det vil si før de deltakende skolene utviklet en ny skriveopplæring med utgangspunkt i et funksjonelt skrivekonstrukt og eksplisitte forventningsnormer. Utvalget av tekster som har blitt vurdert av vurderingspanelet etter intervensjonen på prosjektskolene, utgjør en såkalt *posttest*. Posttesten gir informasjon om nivået på prosjektelevenes skriveferdigheter *etter* intervensjonen. Det vil si etter at de deltakende skolene har utviklet skriveopplæringen med utgangspunkt i Normprosjektets skrivekonstrukt og forventningsnormer.

Forholdet mellom pre- og posttester i Normprosjektet på den ene sida og intervensjonen på skolene fordelt på to faser gis i tabell 6.

Tabell 6. Normprosjektet: forholdet mellom pre- og posttest, og mellom intervensjonsfaser

	Pretest	Intervensjon første fase	Posttest	Intervensjon andre fase	Posttest
Periode	Høst 2012, før intervensjon	7–8 måneder, oktober 2012–juni 2013	Høst 2013 ^a	10 måneder, august 2013–juni 2014	Høst 2014 ^b
Målepunkt	MP1	MP2, MP3	MP4	MP5, MP6	MP7
Intensiv-utvalg	300-, 400-elever 600-, 700-elever	300-, 400-elever 600-, 700-elever	400-elever 700-elever	300-elever 600-elever	300-elever 600-elever
Kontroll-skoler	300, 400-, 600-, 700-elever	–	400-elever	–	300-, 600-elever
Vurderer	Tekstene er vurdert av ett par fra nasjonalt vurderingspanel	Tekstene er vurdert av to par av prosjektskolelærere	Tekstene er vurdert av ett eller to par fra nasjonalt vurderingspanel	Tekstene er vurdert av to par av prosjektskolelærere	Tekstene er vurdert av to par fra nasjonalt vurderingspanel
Vurderingsform	To tekster fra hver elev er vurdert (utkast er ikke skrevet) ^c	En tekst er vurdert på MP2. En andreversjonstekst er vurdert på MP3	En tekst er vurdert (utkast er ikke skrevet)	En første- og andreversjonstekst er vurdert på MP5 og MP6	En tekst er vurdert (utkast er ikke skrevet)

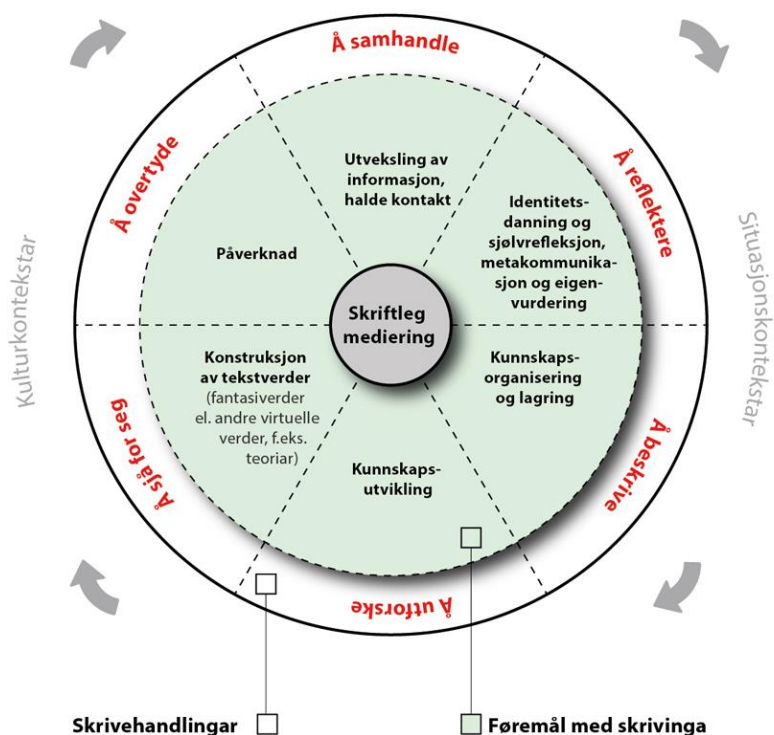
Kommentar. MP = målepunkt

^a Etter intervensjon første fase. ^b Etter intervensjon andre fase. ^c Gjelder de fleste av intensivelevnene.

Som det framgår av tabell 6, startet selve intervensjonen i oktober 2012. Den ble avsluttet i juni 2014. Intervensjonen innebar at det ble utviklet en skriveopplæring og skriveoppgaver basert på skrivekonstrukt og forventningsnormer på den enkelte skolen, på de aktuelle trinnene og i hver enkelt klasse som har deltatt i prosjektet.

Alle vurderinger av elevenes tekster er basert på skrivekonstruktet som er utviklet i forbindelse med de nasjonale utvalgsprøvene i skrivning. Dette konstruktet kalles *Skrivehjulet* (Berge et al., under publisering; Berge & Thygesen, 2015; Evensen, 2010). Skrivehjulet legger til grunn at det å ha skriveferdighet innebærer å ha tilegnet seg skriftspråklige og semiotiske ressurser til å utføre visse handlinger med ulike formål. Skrivehjulet består av seks *skrivehandlinger* med ulike skriveformål. Vi handler i skrift ved å ta relevante skriftlige redskaper i bruk. Skrivehjulet er framstilt i figur 1.

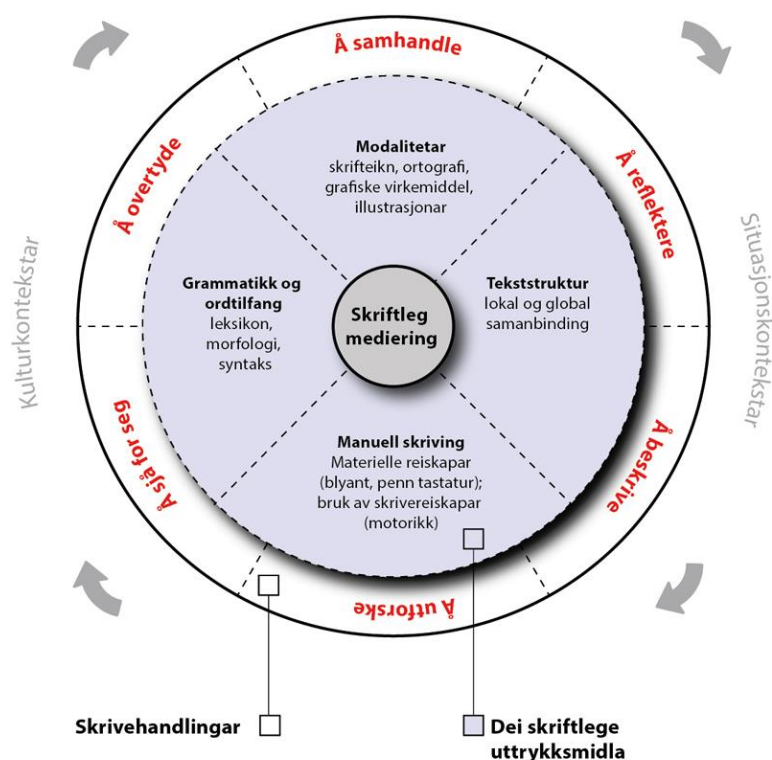
Figur 1. Skrivehjulet: skrivehandlinger og formål



Alle tekster – uansett skrivehandling og skriveformål – har blitt vurdert på de sju ulike vurderingsområdene kommunikasjon, innhold, tekstoppbygging, språkbruk, rettskriving og formverk, tegnsetting og bruk av skriftmediet ut fra en femdelte skala. Samlet utgjør vurderingsområdene kommunikasjon, innhold, tekstoppbygging og språkbruk såkalt *funksjonskompetanse*. Samlet utgjør vurderingsområdene rettskriving og formverk sammen med tegnsetting såkalt *kodekompetanse* (se ellers fotnote 5).

Funksjons- og kodekompetansen og de sju vurderingsområdene som inngår i dem, artikulere hvilke dimensjoner ved den skriftlige medieringen ved elevenes tekster som skal vurderes. De fire konstituerende dimensjonene ved den skriftlige medieringen framgår av figur 2, der medieringsressursene som inngår i Skrivehjulet er synliggjort. *Manuell skriving* dekkes av vurderingsområdet bruk av skriftmediet. *Skriftens modaliteter* dekkes av vurderingsområdene rettskriving og formverk, tegnsetting og bruk av skriftmediet. *Grammatikk og ordforråd* dekkes stort sett av vurderingsområdet språkbruk, men også av vurderingsområdene kommunikasjon og innhold. *Tekststruktur* dekkes av vurderingsområdene kommunikasjon, innhold samt tekstoppbygging.

Figur 2. Skrivehjulet: De semiotiske medieringsressursene



Samlet til en såkalt *sumskåre* vil vurderingene av elevenes funksjonskompetanse og kodekompetanse gi et overordnet inntrykk av elevenes skriveferdigheter ved en gitt skrivehendelse⁵. I presentasjonen av resultatene slik de framgår av vurderingene ved de ulike målepunktene/vurderingshendelsene, er sumskåren analysegrunnlaget.⁶ Sumskåren er definert slik:

- «Sumskåre» = summen av parenes samlede vurderinger (y) av hvert enkeltkompetanseområde: $y = x_{p1} + x_{p2}$, der

y = sumskåre

x_{p1} = sumskåren for par 1

x_{p2} = sumskåren for par 2

Som eksempel ville en elevtekst som av par 1 får $3+2+4+2+3+4 = 18$ poeng og av par 2 får $2+3+3+1+4+3 = 16$ poeng, oppnå en sumskåre på 34 poeng.

Sumskåren ved et målepunkt vil kunne variere fra maksimum 60 til minimum 12 poeng. Har elevene sumskåren 36 eller bedre, vil eleven ligge på eller over forventningsnormene etter fire eller sju års opplæring. En sumskåre lik 36 er det nivået elever ifølge normen bør ligge på, det vil si at sumskåren 36 tilsvarer normen. Ved MP1 og delvis MP4 er det ikke brukt doble

⁵ Vurderingsområdet *Bruk av skriftmediet* inngår ikke i sumskåren. Dette vurderingsområdet er svært komplekst, og har vist seg vanskelig å vurdere på en valid måte. Det bør studeres som et eget kompetanseområde. I presentasjonen av resultatene fra prosjektet i denne teksten, er derfor vurderingsområdet *Bruk av skriftmediet* ikke tatt med. Det er i samsvar med vurderingspraksisen for utvalgsprøven i skriving.

⁶ Det er også mulig å undersøke elevenes prestasjoner og utvikling på utvalgte vurderingsområder, for eksempel i forbindelse med kvalitative studier av elevtekster (se kapittel 4.3).

parvurderinger. I slike tilfeller har vi brukt doblet enkle parvurderinger, og vektet de enkle parvurderingene for slik å konstruere en og samme skala (12-60). Vektingen har blitt validert ved å bruke såkalt Rasch-analyser i statistikkprogrammet FACETS (Linacre, 2014).

Når det i prosjektet er lagt vekt på at alle tekster skal vurderes av fire lærere fordelt i par, altså par mot par, er valget basert på kunnskapen om såkalt *sann vurdering* («true score»). I skrivevurdering tilsvarer sann vurdering snittet av 100 bedømmeres vurdering av en tekst, noe som med 95 % sannsynlighet nås med fem uavhengige vurderere. Fire bedømmere som inngår i og forhandler i par, anser vi å tilsvare fem uavhengige vurderere (Berge, 1993a, 1996; Borgström & Ledin, 2014).

I prosjektet har det blitt lagt vekt på at lærerne skal utvikle og utgjøre et såkalt *tolkningsfellesskap* (Berge, 1993b, 1996, 2009). Derfor er det lagt vekt på at det ikke samles inn individuelle vurderinger, men to parvurderinger som hver for seg er et resultat av to læreres forhandlinger om endelig vurdering av en eller annen tekst. I prosjektet regnes summen av de to parenes vurderinger på ulike vurderingsområder som en valid (troverdig) indeks på denne sanne vurderingen («true score»).

Sumskåren for hver enkelt vurdert tekst gir en samlet indeks på tolkningsfellesskapets sanne vurderinger. Ingen vurderinger er derfor basert på tilfeldige individuelle bedømmelser. Lærerpårene på samling har blitt rotert for å motvirke vurderingsbias knyttet til enkeltlæreres vurderingsvaner. Når det gjelder vurderingene på pretesten (se tabell 6, side 13), er hver tekst vurdert av ett par. Siden pretesten innebærer at de fleste av elevene på alle de fire trinnene har skrevet to ulike tekster som representerer tradisjonelle sjangrer eller skrivemåter i norsk skole, framgår sumskåren som summen av to ulike pars vurderinger av to tekster skrevet av en og samme elev.

I tabell 7 og 8 er reliabiliteten for posttestene høsten 2013 og 2014 presentert uttrykt i Spearmans rho (ρ) og ICC Alpha one-way random, single measure.⁷ Alle korrelasjoner er signifikante ($p < ,05$), der n = antall tekster. Bedømmersamsvaret mellom de enkle eller doble parvurderingene på posttestene er meget høyt, sammenliknet med tilsvarende undersøkelser. Bedømmersamsvaret ligger på et nivå som regnes som akseptabelt uansett hva det som vurderes (jf. Gwet, 2014; Purves, 1992). Et slik nivå er vanskelig å oppnå i vurdering av skriving (jf. Berge, 1996; Borgström & Ledin, 2014; Eckes, 2011; McNamara, 1996; Thygesen et al., 2007). Dermed vil en sumskåre basert på summen av enkle eller doble parvurderinger gi oss tilgang til et troverdig indeks på elevenes skriveferdighet.

Tabell 7. Reliabilitet posttest (MP4), høsten 2013

	ρ	ICC
400-tekster (n = 49)	,704	,699
700-tekster (n = 36)	,823	,789

Kommentar. Bedømmersamsvaret er målt med utgangspunkt i sumskåre.

Tabell 8. Reliabilitet posttest (MP7), høsten 2014

	ρ	ICC
300- tekster (n = 134)	,772	,755
600- tekster (n = 137)	,788	,807

Kommentar. Bedømmersamsvaret er målt med utgangspunkt i sumskåre.

⁷ Som det framgår av tabell 6, er tekstene vurdert på MP1 (pretest) ikke vurdert par mot par. Derfor er det ikke mulig å regne ut reliabiliteten på samme måte og på samme grunnlag som ved MP4 (posttest) og MP7 (posttest). Derfor er heller ikke korrelasjonsnivået ved MP1 presentert i rapporten.

Som det framgår av tabell 6, har 300- og 600-elevene blitt vurdert på en systematisk måte av ett eller to par ved seks ulike vurderingshendelser, eller målepunkt (MP). Ved å sammenlikne tolkningsfelleskapets sanne vurderinger fra målepunkt til målepunkt (MP2, MP3, MP5 og MP6) i intervensjonen og på pre- (MP1) og posttester (MP4 og MP7) vil det være mulig å systematisk studere den eventuelle utviklingen i intensivelevens skriveferdigheter over tid og over et visst aldersspenn. De ulike intensivelevens alder varierer normalt som oppstilt i tabell 9.

Tabell 9. Intensivelevens aldersfordeling

	300-elever	400-elever	600-elever	700-elever
Intensivelevens alder første intervensjonsår, 2012–2013	8 til 9 år: 3. trinn	9 til 10 år: 4. trinn	11 til 12 år: 6. trinn	12 til 13 år: 7. trinn
Intensivelevens alder andre intervensjonsår, 2013–2014	9 til 10 år: 4. trinn	-	12 til 13 år: 7. trinn	-

Som det framgår av tabell 9, kan skriveutviklingen etter intervensjon for en aldersutvikling fra i overkant av 8 år til i overkant av 13 år studeres systematisk. I tillegg er det i prosjektet samlet inn informasjon om elevenes fødselsmåned. Slik vil det være mulig å undersøke om elevenes skriveferdigheter varierer med elevens alder innenfor hvert trinn, eller om det er slik at trinnet bedre forklarer ferdighetsnivået.

Siden forventningsnormene etter 4. trinn er forskjellig fra forventningsnormene etter 7. trinn, er direkte sammenlikninger mellom *elev*er på småskole- og mellomtrinnet ikke mulig. Derimot er det mulig å sammenlikne skriveferdighetene til 300- og 400-elevene på den ene siden og 600- og 700-elever på den andre. Tekster skrevet av intensivelever fra henholdsvis småskoletrinnet (3. og 4. trinn) og mellomtrinnet (6. og 7. trinn) er som nevnt blandet systematisk på en slik måte at trinnet eleven tilhører, ikke kan identifiseres av vurdererne på de ulike målepunktene/vurderingshendelsene det første intervensjonsåret (MP1–3). Prosjektlerere og vurderingspanel har blitt skolert og instruert i å bruke forventningsnormene og vurderingsskalaen likt, uansett hvilken tekst som er vurdert. På denne måten kan effekten av ett og to års intervensjon måles som avstanden i sumskåre mellom MP1 og MP7 for 300- og 600-elevene (to års intervensjon) på den ene sida og MP1 og MP4 for 400- og 700-elevene (ett års intervensjon) på den andre.

Det er også samlet inn informasjon om elevenes målform, hvilken skole eleven går på, elevens kjønn og hvilket språk eleven oppgir å være det første språket eleven lærte.⁸ På denne måten vil det være mulig å studere om elevenes skriveferdigheter og den eventuelle skriveferdighetsutviklingen blant dem systematisk varierer med kjønn, skole, språk og målform – i tillegg til alder og fødselsmåned.

2.3 Eksperimentdesign: prosjektskoler og kontrollskoler

For å kunne sammenlikne skriveferdighetene til intensivelevener før og etter intervensjon med elever som *ikke* har deltatt i intervensjonen, har det i Normprosjektet blitt utviklet og brukt et *kvasiekseptimentelt* design (L. Cohen, Manion, & Morrison, 2011): Intensivelevens

⁸ Sosiokulturelle data slik som foreldres utdanning og liknende er ikke samlet inn.

skriveferdigheter, slik de vurderes av det nasjonale vurderingspanelet i pretesten høsten 2012 (alle elever) og på posttestene høsten 2013 (400- og 700-elever) og høsten 2014 (300- og 600-elever), har blitt sammenliknet med elever fra kontrollskolene. Intensivelevne og kontrollskoleelevne har blitt sammenliknet ved alle relevante målepunkt/vurderingshendelser, det vil si MP1, MP4 og MP7.

Som vi har sett, har et randomisert utvalg på fire kontrollskoler deltatt i eksperimentet. Elevene fra kontrollskolene har skrevet de samme oppgavene som intensivelevne, både ved pre- og posttestene. Tekstene fra kontrollskolene har blitt vurdert av det samme vurderingspanelet og under helt like vilkår som intensivelevne. På denne måten er betingelsene for en valid sammenlikning av vurderingene ved alle relevante vurderingshendelser (MP1, MP4 og MP7) til stede.

Av tabell 10 framgår det hvor mange elever fra de ulike trinnene på intensiv- og kontrollskoler som har fått vurdert tekster på *pretesten*.

Tabell 10. Normprosjektet: pretester for prosjektskoler og kontrollskoler (n elever)

	300-elever (n): 3. trinn	400-elever (n): 4. trinn	600-elever (n): 6. trinn	700-elever (n): 7. trinn
Intensivutvalg	93	78	90	92
Kontrollskoleutvalg	66	72	67	26

Det lave antallet kontrollskoleelever på 700-trinn skyldes praktiske utfordringer med innsamlingen av data.

Av tabell 11 framgår det hvor mange elever fra de ulike trinnene på intensiv- og kontrollskolene som har fått vurdert tekster ved posttesten høsten 2013 og 2014. Det er ikke samlet inn tekster fra kontrollskolene på 7. trinn høsten 2013 (700-elever).

Tabell 11. Normprosjektet: posttester for prosjektskoler og kontrollskoler (n elever)

	3. trinn: 300-elever (n)	4. trinn: 400-elever (n)	6. trinn: 600-elever (n)	7. trinn: 700-elever (n)
Intensivutvalg	79	77	84	83
Kontrollskoleutvalg	55	66	58	data ikke samlet inn
Posttest	MP7: Høsten 2014	MP4: Høsten 2013	MP7: Høsten 2014	MP4: Høsten 2013

Som nevnt gjør bruken av kontrollskoler det mulig å måle effekten av intervensjonen. Det er mulig fordi elevene på kontrollskolene har skrevet tekster under samme vilkår som elevene på intervensjonsskolene, samtidig som vurderingene av disse tekstene er gjort av de samme vurdererne (fra det nasjonale vurderingspanelet) og under like vilkår. Effekten måles som forskjellen i sumskåre mellom intensivelevers og kontrollskolelevers tekster vurdert på MP1 (pretest høsten 2012) og på MP7 (posttest 2014). Følgelig er det effekten av i underkant av *to års intervensjon* som måles. Det innebærer at effekten måles på elever som begynte på 3. (300-elever) eller 6. trinn (600-elever) det året som prosjektet startet høsten 2012, og som var elever på 5. eller 8. trinn høsten undervisningsåret 2014/2015. Det er ingen påviselig effekt av prosjektet etter 7-8 måneders intervensjon. Derfor er sammenlikninger mellom 400- og 700-elever fra kontroll- og prosjektskoler ikke gjennomført.

Det er viktig å legge merke til at effekten *måles på et tidspunkt der elevene ikke lenger deltar i intervensjonen*. For de fleste elevenes vedkommende innebærer det at de skriver teksten om lag et halvt år etter at de har skrevet tekster på intervensjonsskolene. For mange elevers vedkommende innebærer det at de har byttet både skole og klasse. Det gjelder særlig 600- og 700-elevne, som har gått over i ungdomsskolen. En eventuell påviselig effekt av intervensjonen kan derfor tolkes som å være vedvarende.

3 Resultatpresentasjon: Har intervensjonen fra Normprosjektet hatt effekt på elevenes skriveferdigheter?

I dette kapitlet skal vi analysere oss fram til hvilken effekt intervensjonen i regi av Normprosjektet kan ha hatt på utviklingen av elevenes skriveferdigheter. Der det er mulig, forsøker vi gjennom statistiske analyser å finne forklaringer på funnene. Er ikke det mulig, drøfter vi alternative måter å forklare resultatene på.

Innledningsvis ved hver analyse redegjør vi for resultater og hvilke statistiske metoder som er brukt i analysene.

3.1 Elevenes skriveutvikling gjennom intervensjonen hele utvalget samlet

I tabell 12 er oppstilt sumskårer i snitt (M) og spredning (SD) for alle elever (n) fordelt på de fire trinnene på alle mulige målepunkt.

Tabell 12. Snitt (m) basert på sumskåre på alle mulige målepunkt, alle trinn

MP	300-elever			400-elever			600-elever			700-elever		
	n	M	SD	n	M	SD	n	M	SD	n	M	SD
MP1	93	29,5	7,0	78	32,3	7,6	90	28,3	6,5	92	31,6	7,1
MP2	92	29,1	7,2	85	30,3	7,7	90	28,8	6,8	93	32,6	7,8
MP3	94	32,4	6,4	83	33,6	7,6	90	31,7	6,5	96	33,8	7,5
MP4	-	-	-	77	34,4	7,6	-	-	-	83	35,6	7,8
MP5	93	33,9	8,1	-	-	-	87	34,2	8,5	-	-	-
MP6	88	33,6	6,6	-	-	-	88	34,6	9,5	-	-	-
MP7	79	40,3	8,7	-	-	-	84	33,7	8,7	-	-	-

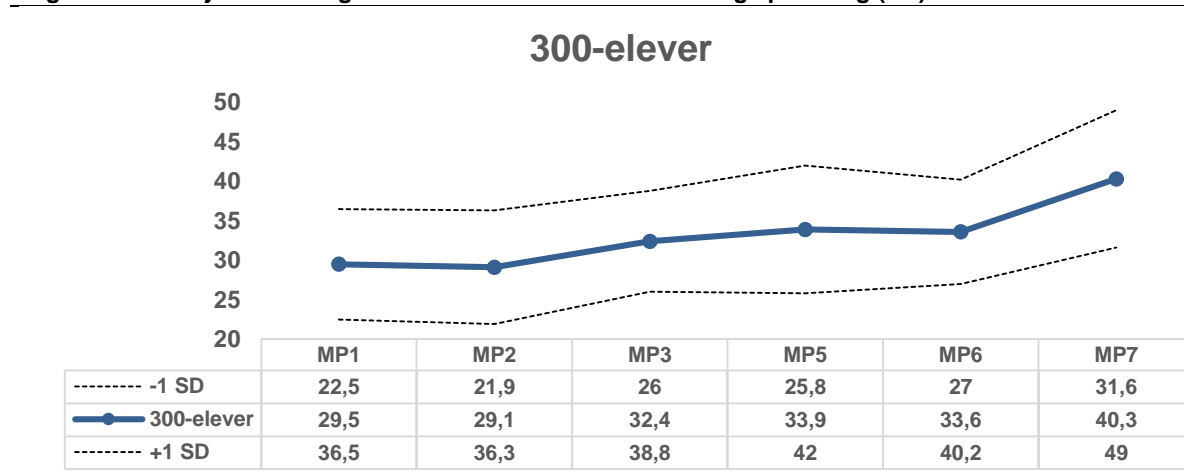
Det er flere statistiske analysemetoder som vil være interessante for å kunne avdekke om elevenes skrivekompetanse bedres gjennom prosjektperioden. For det første om det er signifikante forskjeller i prestasjonen fra målepunkt til målepunkt. Av resultatene presentert som gjennomsnittlige sumskårer (M) i tabell 12, ser det ikke ut til å være noen forskjell i skriveferdighetsnivået fra MP1 til MP2 for noen elevgrupper. For 300-elevene og 600-elevene er det ingen forbedring overhodet. For 400-elevene er det en tilbakegang, men for 700-elevene en svak framgang. Vi kan dermed trekke den slutningen at intervensjonen ikke virker kortsiktig innenfor en tidsperiode på to til tre måneder.

Undersøker vi derimot utviklingen mellom MP1 på den ene sida og de andre målepunktene i hele intervensjonsperioden på den andre, blir bildet annerledes. For alle gruppene ser det ut til å være en kvalitativ endring fra MP2 til MP3. 300-elevenes snittskåre stiger fra 29,1 til 32,5, 400-elevenes fra 30,3 til 33,7, 600-elevene fra 28,8 til 31,7 og 700-elevenes fra 32,6 til 33,8. Ved hjelp av ANOVA-analyse undersøker vi om det er signifikante forskjeller i sumskåre mellom MP1 og de andre mulige målepunktene gjennom to års intervensjon for 300- og 600-elevene. I tabell 13 presenteres resultatene av ANOVA-analysene for 300- og 600-elevenes ferdighetsutvikling gjennom to intervensjonsår:

Tabell 13. ANOVA-analyser av ferdighetsutviklingen for 300- og 600-elever gjennom to års intervensjon

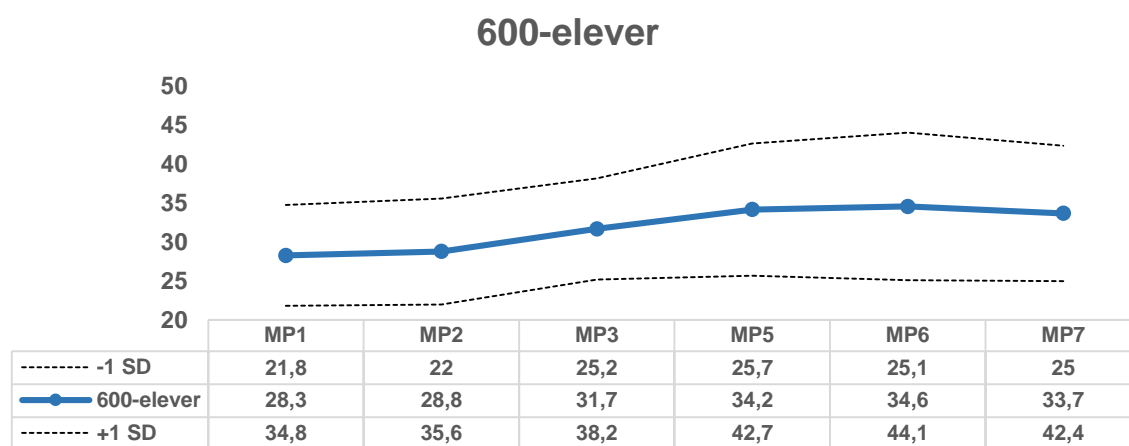
	MP1-MP3	MP1-MP5	MP1-MP6	MP1-MP7
300-elever	p = ,000	p= ,001	p= ,013	p = ,000
600-elever	p = ,068	p= ,052	p= ,001	p = ,000

ANOVA-analysene viser at det for 300-elevene skjer en kvalitativ forbedring i elevenes skriveferdigheter målt i sanne vurderinger på nasjonale vurderingssamlinger allerede etter sju til åtte måneders intervensjon om våren 2013. Det vil si fra og med MP3. Etter sommeren fortsetter tendensen til stadig bedre skrijving. I figur 3 er denne utviklingen visualisert. Vi ser at laveste nivå i spredningen (minus ett standardavvik) på MP7 ligger høyere enn snittskåren på MP1 og MP2:

Figur 3. Prestasjonsutvikling for 300-elevene: sumskåresnitt og spredning (SD)

For 600-elevene tar kvalitetsendringene i elevenes skrijving noe lenger tid enn for 300-elevene. Ikke før etter ett års intervensjon skjer det en signifikant forbedring av elevenes skriveferdighet. Figur 4 illustrerer denne noe ustabile ferdighetsutviklingen blant 600-elevene. Vi ser av figur 4 at kvaliteten på elevenes skriveferdigheter er svært spredt distribuert (målt i standardavvik) fra og med MP5. Spredningen antyder at noen 600-elever har hatt større utbytte av intervensjonen enn andre elever fra samme gruppe.

Figur 4. Prestasjonsutvikling for 600-elevene: sumskåresnitt og spredning (SD)



Som vi skal se i kapittel 3.3, skyldes noe av denne kvalitetsforbedringen som framgår av figur 3 og 4, elevenes alder. Vi skal i kapittel 3.4 studere systematisk om endringene også kan forklares med trinntilhørighet, fødselsmåned og/eller effekt av intervensjonen.

Interessant er det i så måte å sammenlikne snittet for sumskårer ved MP4 (400- og 700-elever) med sumskårer ved MP7 (300- og 600-elever), jf. tabell 10. Ved MP4 og MP7 var både 300- og 400-elevene ved begynnelsen av 5. trinn, mens 600- og 700-elevene var ved begynnelsen av 8. trinn. Forskjellen i kvalitet mellom 300- og 400-elevene er meget stor: 300-elevenes sumskåre er ved MP7 40,3 (n=79), mens 400-elevenes sumskåre ved MP4 er 34,4 (n=77). Derimot har vi motsatt tendens for 600- og 700-elevene: 600 elevenes sumskåre er ved MP7 33,7 (n=84) mens 700-elevenes sumskåre ved MP4 er 35,6 (n=83).

Det kan være flere forhold som forklarer variasjonen i de gjennomsnittlige sumskårene: For det første at de ulike gruppene er sammensatt slik at kvaliteten varierer med gruppa, for eksempel at 600-elevenes skriveferdigheter som gruppe betraktet er svakere enn 700-elevenes. For det andre at oppgaven på utvalgsprøvene 2012 og 2014 (som prosjektskolene tok høsten 2013 og 2014) stiller ulike krav til elevene, og dermed gjør direkte sammenlikninger mellom resultatene på prøvene problematiske. Forholdet mellom snittskårene på MP4 og MP7 kan derfor ikke uten videre regnes som gyldige mål for sammenlikning mellom elevgruppene. Mer kvalitative undersøkelser vil kunne gi mer informerte forklaringer. I denne rapporten legger vi til grunn at det ser ut til å være problematisk uten videre å sammenlikne utvalgsprøveresultatene for 300- og 600-elevene på den ene sida og 400- og 700-elevene på den andre. Vi kommer derfor til å legge avgjørende vekt på analysen av effekt ved å se på forholdet mellom prosjektskoleelever og kontrollskoleelever, der utgangsdata har blitt hentet inn gjennom sammenliknbare forhold. Som vi har sett av analysene over, er det først våren 2013 (i første intervensjonsår) at det ser ut til at alle 300- og 600-elevene sett under ett skriver kvalitativt bedre enn de gjorde før prosjektet startet høsten 2012. Vi har også sett at det er ytterligere signifikant kvalitetsforbedring fra og med høsten skoleåret 2013/2014.

3.2 Elevenes skriveutvikling gjennom intervensjonen fordelt på skoler og trinn

Selv om tendensen til bedre skriveferdighet er signifikant fra og med våren 2013 for alle elevgrupper, er det stor variasjon mellom skoler og/eller trinn i hvordan intervensjonen har virket, vurdert på grunnlag av sumskåresnittet på skolene. Vi skal i dette avsnittet gjennomføre en mer kvalitativ analyse, der vi abduktivt (Svennevig, 2001) og på en utforskende måte forsøker å utvikle hypoteser om hvor vidt noen skoler og/eller trinn på skolene ser ut til å bidra til effekten mer enn andre skoler og/eller trinn. Vi tar med oss hypotesene fra disse analysene når vi i neste avsnitt tester dem ved hjelp av ulike statistiske analyser.

Når vi studerer skolenes sumskåresnitt ved ulike målepunkt/vurderingshendelser, er det viktig å være oppmerksom på at sumskårene på skolene ikke er et uttrykk for kvaliteten i undervisningen på skolen generelt. Til det er elevutvalget på en rekke skoler meget lite, og for tilfeldig. Selv på skoler med forholdsvis mange elever i intensivutvalget (slik som for eksempel Aura og Engen), kan utvalget gi tilfeldige utslag som ikke kan generaliseres til skriveferdighetskvalitetene på trinnet som er undersøkt. Likevel vil vi i analysene legge til grunn at det lar seg gjøre å avdekke et mulig mønster ved å se på forholdet mellom sumskårene på alle seks målepunkt samlet og i forhold til hverandre. Mange målepunkt veier til en viss grad opp for et lite elevutvalg. Vi mener altså at vi kan trekke visse slutninger om skolens/trinnets bidrag til eventuell effekt ved å studere tendenser i sumskårene på alle seks mulige målepunkt. På skoler og eller trinn der det overhodet ikke kan avdekkes noen utvikling, regner vi det som et avvik fra den statistisk påviste tendensen til ferdighetsutvikling.

En oversikt over 300-elevenes resultater fordelt på intervensjonsskoler gis i tabell 14:

Tabell 14. 300-elevenes resultater i sumskåre fordelt på intervensjonsskoler

Skole	MP1 (n)	MP2 (n)	MP3 (n)	MP5 (n)	MP6 (n)	MP7 (n)
Aura	31.0 (12)	29.5 (12)	32.7 (12)	35.8 (12)	36.8 (11)	43.5 (11)
Bukta	29.5 (8)	29.3 (8)	34.6 (8)	32.1 (8)	33.5 (8)	38.9 (7)
Casa	32.5 (6)	31.8 (6)	34.0 (6)	34.0 (6)	30.4 (5)	41.6 (5)
Dalen	27.0 (1)	26.0 (1)	23.0 (1)	-	26.0 (1)	-
Engen	30.4 (11)	29.5 (11)	30.9 (11)	36.6 (11)	32.0 (11)	39.8 (12)
Fjellet	31.5 (2)	34.5 (2)	37.0 (2)	46.0 (2)	32.0 (2)	39.5 (2)
Gropa	25.0 (2)	22.5 (2)	30.5 (2)	32.0 (2)	33.0 (2)	46.5 (2)
Heia	15.0 (2)	26.0 (2)	24.0 (2)	24.0 (2)	-	20.0 (1)
Isen	31.5 (2)	25.5 (2)	29.5 (2)	24.5 (2)	32.5 (2)	40.5 (2)
Jordet	23.5 (6)	22.8 (6)	28.5 (6)	32.7 (6)	40.4 (5)	-
Kilden	29.6 (5)	26.2 (5)	30.8 (5)	24.5 (4)	26.3 (4)	33.5 (2)
Lia	34.0 (5)	40.0 (5)	42.8 (5)	43.5 (6)	37.7 (6)	46.2 (6)
Marka	26.7 (6)	24.2 (6)	33.0 (6)	29.6 (5)	33.4 (5)	35.3 (6)
Neset	31.8 (4)	29.0 (4)	33.0 (4)	30.6 (5)	33.2 (5)	38.3 (4)
Osen	26.5 (2)	27.0 (2)	27.0 (2)	36.0 (2)	28.5 (2)	40.0 (2)
Plassen	26.0 (4)	27.3 (4)	27.5 (4)	28.3 (4)	29.8 (4)	35.3 (4)
Quadra	18.5 (2)	23.5 (2)	23.5 (2)	26.0 (2)	24.0 (1)	20.0 (1)
Raset	31.0 (2)	29.0 (2)	35.0 (2)	34.5 (2)	35.0 (2)	39.0 (1)
Stranda	35.5 (6)	38.6 (5)	38.7 (7)	39.0 (7)	36.0 (7)	46.9 (7)
Teigen	30.6 (5)	34.0 (5)	30.8 (5)	35.0 (5)	35.0 (5)	39.3 (4)
Alle m (n)	29.4 (93)	29.1 (92)	32.5 (94)	34.0 (93)	33.7 (88)	40.3 (79)

Vi ser at det er stor variasjon i 300-elevens skriveferdighet før intervensjonen. Ved skoler som Lia og Stranda vurderes elevenes skriveferdighet høsten 2012 å ligge omtrent på normen (=36). Snittskåre for Stranda på MP1 er 35,5, mens den for Lia er 34,0. Det høye nivået før intervensjonen tar til, er oppsiktsvekkende, tatt i betraktning at elevene på dette tidspunktet har fullført om lag to måneder av 3. årstrinn, mens de er vurdert med forventningsnormer som skal gjelde etter fire års opplæring. I den andre enden av skriveferdighetsskalaen finner vi elever fra Quadra og Heia. Elevene fra disse to skolene er på et meget lavt skriveferdighetsnivå før intervensjonen høsten 2012. Snittskåre for de to Heia-elevene på MP1 er 15,0, mens den for de to Quadra-elevene i intensivutvalget er 18,5.

Går vi litt grundigere inn på variasjonen mellom skolene for 300-elevene, ser vi at det er hele 15 av skolene som vi på grunnlag av snittskårene på ulike målepunkt, vil kunne hevde har en solid kvalitetsutvikling gjennom hele intervensjonen. Med unntak av to skoler⁹ har hele tretten skoler en sumskåre på posttesten (MP7) som ligger på eller over snittskårenormen på 36. Noen skoler har sumskårer for sine elever som er betydelig over normen. Det gjelder følgende skoler: Aura, Casa, Gropa, Isen, Lia, Osen og Stranda. Men mange av de andre skolene har en snittskåre tett opp til normen. En oppsummerende oversikt over disse tendensene gis i tabell 15.

Tabell 15. Skoler med påviselig utvikling blant 300-elevene, to intervensjonsår sett under ett

Sumskåre \geq 36	Til en viss grad
Aura	Isen
Bukta	Marka
Casa	Osen
Engen	
Fjellet	
Gropa	
Lia	
Neset	
Raset	
Stranda	
Teigen	

På to av skolene tenderer utviklingen blant intensivelevene til å være ustabil. Det gjelder skolene Marka og Osen.

Fem av skolene i Normprosjektet har, når det gjelder 300-elevene, en svært svak og/eller ingen påviselig utvikling blant intensivelevene gjennom to intervensjonsår. For å skape et tydeligere bilde av elevene på disse skolene har vi sammenliknet sumskåren på pretesten (altså før intervensjonen) med snittskårene for alle mulige målepunkt i andre intervensjonsår, i tillegg til sumskåren på posttesten. Slik unngår vi at sumskåren på utvalgsprøven høsten 2014, tatt etter intervensjonen, får for stor betydning for vår analyse av de aktuelle fem skolenes utbytte av intervensjonen de har deltatt i. Tabell 16 er en presentasjon av sumskårene før intervensjon og etter to års intervensjon på de fem skolene med svakest presterende intensivelever:

⁹ Tekstene fra den ene av de to skolene, Jordet, er ikke vurdert av utvalgsprøvepanelet. Det skyldes en misforståelse på skolen om hvordan tekstene skulle skrives. Dermed kunne ikke MP7-data fra Jordet kunne regnes som gyldige.

Tabell 16. Skoler med svak og/eller ingen påviselig utvikling blant 300-elevene

Skole	Sumskåre MP1	Snittskåre MP5+MP6+MP7
Dalen	27.0	26,0
Heia	15.0	22,0
Kilden	29.6	28,1
Plassen	26.0	31,1
Quadra	18.5	23,3

Kommentar. Etter to intervensjonsår.

Vi ser at de fire elevene ved Kilden har lavere snittskåre på MP5, MP6 og MP7 samlet enn de hadde på pretesten. Også den ene eleven ved Dalen får i andre intervensjonsår en skåre som er lavere enn på pretesten. Elevene skriver altså *dårligere* etter intervensjonen enn det de skrev før intervensjonen. Det svake resultatet er enda mer oppsiktsvekkende siden elevene det angår, har blitt to år eldre i løpet av prosjektet. De i alt fem elevene dette resultatet gjelder for, skriver da også betydelig under normen. Et svakt avvik fra dette mønsteret finner vi på Kilden, der de to av elevene (av fem mulige) som tok prøven, skåret noe bedre enn forventet ut fra sumskårene til elevene samlet på andre målepunkt andre intervensjonsår.

Når det gjelder de meget svake elevene fra skolene Heia og Quadra, skjer det en viss forbedring av skriveferdighetene fra MP1 til andre intervensjonsår sett under ett. Likevel ligger de fire intensivelevnene fra disse skolene som gruppe betraktet, langt under sumskårenormen på 36.

En skole det kan være mer problematisk eller kontroversielt å plassere blant skoler der intervensjonen ikke ser ut til ha virket på de deltakende elevene, er Plassen. De fem elevene som inngår i intensivutvalget fra denne skolen, har så å si ingen påviselig forbedring av skriveferdighetene. Likevel lykkes fire av elevene fra Plassen på utvalgsprøven høsten 2014 (MP7) å oppnå en sumskåre som i snitt ligger like under normen. Vi kan derfor ikke utelukke at elevene fra denne skolen har hatt læringseffekt av intervensjonen. For de andre fire skolene kan det settes spørsmålsteget ved om elevene har deltatt i eller har hatt utbytte av intervensjonen. Den svake resultatutviklingen gjelder elever på følgende skoler: Dalen, Heia, Kilden og Quadra. Det er viktig igjen å understreke at det lave antallet elever ikke uten videre gir grunnlag for å hevde noe generelt om hvorvidt intervensjonen har virket på alle de deltakende elevene ved skolen. Mer kvalitative undersøkelser av skriveopplæringen på skolene i løpet av intervensjonen vil kunne gi en forklaring på det påfallende avviket disse 300-elevnene utgjør i intensivutvalget.

Trass i at det er elever fra fem projektskoler som kan se ut til å ha hatt lite utbytte av å delta i intervensjonen, er det allmenne mønsteret blant 300-elevnene at det har vært en sterk kvalitetsutvikling gjennom intervensjonen. Hvor sterk denne utviklingen har vært, skal vi analysere oss fram til i neste avsnitt. Men hva med 600-elevnene? Følger de mønsteret fra 300-elevnene der det store flertallet av skoler, klasser og elever har utviklet skriveferdighetene betydelig gjennom prosjektet?

Tabell 17 gir en oversikt over sumskårer for 600-elevnene fordelt på skoler på alle relevante målepunkt:

Tabell 17. 600-elevenes resultater fordelt på intervensjonsskoler

Skole	MP1 (n)	MP2 (n)	MP3 (n)	MP5 (n)	MP6 (n)	MP7 (n)
Aura	30.6 (11)	27.9 (12)	35.6 (12)	33.6 (11)	39.2 (11)	39.1 (9)
Bukta	25.5 (4)	28.0 (4)	31.8 (5)	30.0 (5)	34.6 (5)	32.6 (5)
Casa	34.0 (7)	33.9 (7)	36.2 (6)	46.2 (6)	36.6 (7)	41.8 (6)
Dalen	20.0 (3)	27.0 (3)	30.3 (3)	-	25.0 (3)	27.0 (3)
Engen	28.6 (9)	29.6 (9)	27.3 (9)	37.8 (9)	33.2 (9)	36.3 (6)
Fjellet	23.0 (2)	34.0 (2)	37.0 (2)	39.5 (2)	34.5 (2)	34.5 (2)
Gropa	32.0 (2)	25.5 (2)	28.5 (2)	33.0 (2)	41.5 (2)	22.0 (2)
Heia	22.0 (2)	25.5 (2)	32.5 (2)	24.0 (2)	24.0 (2)	40.0 (1)
Isen	28.3 (3)	27.3 (3)	30.3 (3)	29.5 (2)	30.0 (3)	33.0 (3)
Jordet	29.3 (7)	23.5 (8)	31.8 (8)	35.4 (8)	30.5 (6)	26.6 (7)
Kilden	20.3 (3)	-	29.0 (3)	29.3 (3)	30.0 (3)	37.0 (3)
Lia	24.5 (6)	39.3 (6)	35.4 (5)	34.2 (6)	40.7 (6)	34.5 (6)
Marka	33.0 (5)	34.5 (2)	30.0 (2)	37.8 (4)	41.5 (4)	40.3 (4)
Neset	35.3 (6)	36.8 (6)	33.0 (6)	40.2 (5)	39.8 (5)	33.4 (5)
Osen	25.3 (3)	23.3 (3)	33.3 (3)	26.0 (3)	22.7 (3)	25.0 (3)
Plassen	26.5 (4)	26.8 (5)	35.5 (4)	31.8 (5)	32.0 (5)	34.3 (4)
Quadra	-	31.5 (2)	24.5 (2)	20.5 (2)	-	24.0 (2)
Raset	30.0 (2)	38.0 (2)	36.0 (2)	39.5 (2)	35.0 (2)	35.0 (2)
Stranda	22.8 (5)	28.4 (5)	27.0 (5)	33.4 (5)	34.8 (5)	32.2 (5)
Teigen	28.3 (6)	25.4 (7)	25.0 (6)	27.8 (5)	33.6 (5)	32.0 (6)
Alle m (n)	28.3 (90)	28.8 (90)	31.7 (90)	34.2 (87)	34.6 (88)	33.7 (84)

Vi ser at også for 600-elevene varierer ferdighetene målt ved pretesten ganske mye. Høyest ligger elevene fra Casa (34 i snitt), Marka (33 i snitt) og Neset (35,3 i snitt). Et forholdsvis svakt skriveferdighetsnivå før intervensjon finner vi særlig på Dalen (20 i snitt), Kilden (20,3 i snitt) og Heia (22 i snitt).

600-elevene framstår i tabell 17 som en mer sammensatt gruppe enn 300-elevene. Likevel finner vi at elever fra fjorten av de tjue prosjektskolene har en påviselig positiv skriveutvikling, to intervensjonsår sett under ett. Likevel, som vi har påpekt tidligere, når ikke 600-elevene i snitt en sumskåre på utvalgsprøvene høsten 2014 (MP7) som ligger på eller over normen. Som omtalt tidligere, gjelder det spesielle forholdet for 600-elevene at de har tatt prøven på et tidspunkt da de fleste av dem har byttet både klasse, skole og skoleslag. Da posttesten ble tatt høsten 2014, var de begynt på ungdomskolen, som kan ha en noe annen læringskultur enn den vi finner i barneskolen. En interessant skole i så måte er Neset. De seks eller fem elevene fra denne skolen ligger på et ferdighetsnivå i andre intervensjonsår som er meget høyt, men elevene lykkes ikke med å bekrefte det høye nivået på posttesten.

Oversikten over hvilke skoler som ser ut til å ha hatt en påviselig skriveutvikling, målt i sumskåre på MP7, framgår av tabell 18:

Tabell 18. Skoler med påviselig utvikling blant 600-elevene, to intervensjonsår sett under ett

Sumskåre \geq 36	Til en viss grad
Aura	Fjellet
Casa	Gropa
Engen	Jordet
Marka	Lia
	Neset
	Raset
	Stranda
	Teigen

Av de fjorten skolene er det fire skoler som etter intervensjon oppnår en sumskåre over normen på 36 eller bedre. Det er følgende skoler: Aura, Casa, Engen og Marka. For flere av disse skolene er sumskårene på posttesten meget høye og fullt på høyde med de beste skårene for 300-utvalget. Som vi ser, er disse skolene også skoler som utmerker seg med en solid skriveutvikling også blant 300-elevene. Ett unntak fra dette mønsteret er Marka skole. Noe av den samme utviklingen finner vi på følgende skoler: Lia og Raset. Disse i alt seks skolene ser ut til å utgjøre en gruppe skoler med spesielt sterk ferdighetsutvikling. Vi skal studere disse seks skolenes utvikling statistisk i kapittel 3.2.

Vender vi oss til skolene med svak og/eller ingen påviselig utvikling blant 600-elevene, to intervensjonsår sett under ett, finner vi fire skoler som også hadde svak utvikling blant 300-elevene. Det gjelder skolene Dalen, Heia, Kilden og Quadra. En oversikt over skoler med svak og/eller ingen påviselig utvikling blant 600-elevene, to intervensjonsår sett under ett, er oppstilt i tabell 19.

Tabell 19. Skoler med svak og/eller ingen påviselig utvikling blant 600-elevene

Skole	Sumskåre MP1	Snittskåre MP5+MP6+MP7
Dalen	20,0	26,0
Heia	22,0	29,3 (høyt på MP7: en elev)
Isen	28,3	30,8
Kilden	20,3	32,1 (høyt på MP7: tre elever)
Osen	25,3	24,6
Quadra	ikke vurdert	25,2

Kommentar. Etter to intervensjonsår.

Trass i svake resultater fra intensivelevnene på disse skolene, kan sumskårene variere på måter som sannsynligvis krever kvalitative forklaringer. Oppsiktsvekkende nok skårer den ene eleven fra Heia som tok posttesten høsten 2014, meget høyt her (snittskåre på MP7=40). Det er et resultat som avviker sterkt fra den aktuelle elevens ferdighetsutvikling gjennom prosjektperioden. Det er også oppsiktsvekkende at de tre elevene fra Kilden på posttesten hadde en så høy snittskåre som 37, mens de på alle andre målepunkt ikke dokumenterer annet enn ganske svake skriveferdigheter.

Når det gjelder de seks intensivelevnene fra 600-gruppa på skolene Isen og Osen, ser de ikke ut til å utvikle skriveferdighetene overhodet i løpet av intervensjonsperioden. Også i disse tilfellene er det grunn til å undersøke gjennom kvalitative metoder hva den manglende skriveferdighetsutviklingen kan skyldes. Dersom forhold ved skolene gjør at det i realiteten ikke har vært gjennomført en intervensjon etter hensikten og målet med Normprosjektet, kan det gis grunner for å ekskludere intensivelevnene fra 600-elevene på disse skolene fra de videre kvantitative analysene.

Som en oppsummering av de mer kvalitativt vinklede analysene av skriveferdighetsvariasjon mellom skolene og/eller mellom trinn på eller på tvers av skolene, kan følgende mønster skimtes:

- Følgende fire skoler har en tydelig skriveferdighetsutvikling uansett trinn: **Aura, Casa, Engen, Lia**. Til sammen utgjør elevene fra disse skolene 34,5 % av intensivutvalget (67 av 194 intensivelever) i siste intervensjonsfase skoleåret 2013/2014.
- Følgende tre skoler har en påfallende skriveferdighetsvariasjon mellom ulike trinn på skolen: **Stranda, Marka og Raset**. Denne variasjonen på skolene kan selvsagt skyldes

elevutvalget, men likevel er forskjellene påfallende. 300-elevene på Stranda skole holder et betydelig høyere skriveferdighetsnivå enn 600-elevene på samme skole. For Marka og Raset gjelder det motsatte forhold. Der er det 600-elevene som ser ut til å hatt påviselig utbytte av intervensjonen. Kvalitative studier kan avdekke om variasjonen på samme skole skyldes ulikheter i undervisningen på ulike trinn. Til sammen utgjør elevene fra disse skolene 13,9 % av intensivutvalget (27 av 194 intensivelever) i siste intervensjonsfase skoleåret 2013/2014.

- Følgende fire skoler er generelt skoler med svak eller ingen skriveutvikling blant intensivelevnene, uansett trinn: **Dalen, Heia, Kilden, Quadra**. Her er skriveferdighetsnivået så svakt og skriveferdighetsutviklingen uansett målepunkt og gjennom hele intervensjonen så liten, at det kan være grunn til mer kvalitativt å undersøke om intervensjonen er gjennomført etter intensjonen, og om elevene faktisk har hatt utbytte av den. Til sammen utgjør elevene fra disse skolene 10,3 % av intensivutvalget (20 av 194 intensivelever) i siste intervensjonsfase skoleåret 2013/2014.

For de øvrige ni skolene gjelder det forhold at intensivelevnene på skolene utvikler skriveferdighetene påviselig, med unntak av Plassen skole for 300-elevene, og Isen og Osen skoler for 600-elevene. Vi kan altså konstatere at for nesten 90 % av elevene som gruppe betraktet, har det vært en påviselig skriveutvikling gjennom intervensjonen. For noe i overkant av 10 % av elevene, har utbyttet vært meget svakt eller ikke påviselig.

På bakgrunn av disse analysene skal vi nå studere effekten av prosjektet målt som forskjellen i skriveutvikling mellom prosjektskoler og kontrollskoler. Dette gjør det mulig å kontrollere i hvilken grad den påviste utviklingen skyldes intervensjonen, eller om den kan tilskrives alder og modning. Vi skal ta med oss observasjonene fra dette kapitlet i drøftingen av de statistiske analysene og funnene vi presenterer der. I den følgende avgjørende delen av analysen skal vi gå grundigere inn på det statistiske resonnementet som ligger til grunn for analysene.

3.3 Effekten av intervensjonen

I den här sektionen skall vi utvidga den tidigare analysen och statistiskt testa hypoteserna utvecklade i kapitel 3.1. Vi kommer i huvudsak att svara på två frågor:

- 1) Hur stark är den kvalitetsutveckling som framkom i förra kapitlet?
- 2) Är det någon skillnad i utveckling mellan projektskolor och kontrollskolor?

3.3.1 Urval

Det primära urvalet för denna del av undersökningen är de elever som deltagit i projektet i två år, och som dessutom har genomfört både MP1 och MP7 (se tabell 20). Detta innebär 129 elever på 300-nivån och 136 elever på 600-nivån, vilket i sin tur innebär ett bortfall på mellan 13,4 % och 22,6 %. Mot bakgrund av resultaten som presenterades ovan, har vi också konstruerat ett sekundärt urval. I detta har eleverna från Dalen, Heia, Kilden och Quadra uteslutits (jfr diskussionen av dessa fyra skolor i avsnitt 3.1).

Tabell 20. Primärt och sekundärt urval och bortfall

	N totalt	Prim. urval (n)	Bortfall (%)	Sek. urval (n)	Bortfall (%)
300 _i = projektsk.	97	75	22,7	71	26,8
300 _k = kontrollsk.	67	54	19,4	54	19,4
600 _i = projektsk.	97	78	19,6	71	26,8
600 _k = kontrollsk.	67	58	13,4	58	13,4

Som synes är det ett någorlunda jämnt bortfall i tre av fyra grupper med omkring 20 %. Bortfallet för kontrollskoleelever på 600-nivån är något mindre (13.4 %).

Skillnaden mellan det primära och sekundära urvalet är liten (4 respektive 7 elever på 300- respektive 600-nivån). Vi kan därmed inte vänta oss några större skillnader i resultat. Analyser på grundval av det sekundära urvalet kan ändå vara av intresse, eftersom Dalen, Heia, Kilden och Quadra systematiskt avviker från övriga projektskolor (se ovan). Kommande kvalitativa undersökningar kan närmare granska vad som kännetecknar interventionen på dessa skolor och därmed i större utsträckning avgöra vilket av urvalen som är det rimligaste att inkludera.

3.3.2 Metod

För att svara på frågorna har vi genomfört en rad analyser. För det första har vi genomfört parvisa t-tester för att undersöka skillnaden mellan MP1 och MP7 för samtliga grupper.¹⁰ Vi har i samband därmed också skattat effektstorleken (ES) inom respektive trinn och inom och mellan respektive urval (projekt- och kontrollskolor), för att få en indikation på styrkan i skillnaden mellan pre- och posttest för de olika grupperna.

För det andra har vi genomfört ANCOVA-analyser för att möjliggöra direkt jämförelse mellan projekt- och kontrollskolorna. Även här har vi skattat effektstorleken för att få en indikation på styrkan i skillnaden.

För det tredje har vi specialstuderat projektskolorna genom att undersöka i vilken utsträckning resultat på MP1 kan förklara eller användas för att förutspå resultat på övriga mätpunkter.

I de kommande avsnitten presenterar vi ANCOVA och ES mer ingående. Därefter presenterar vi resultaten.

3.3.2.1 ANCOVA-analyserna

ANCOVA (*analysis of covariance*) är ett lämpligt redskap vid experiment eller experimentliknande upplägg med pre- och posttest. En ANCOVA-analys innebär möjligheter att kontrollera för initiala skillnader mellan grupper, vilket gör jämförelsen av resultat på posttestet rimligare. Denna möjlighet är viktig, eftersom skillnader i posttest-resultat kan hänga samman med skillnader i pretest-resultat (se mer om ANCOVA i Glass & Hopkins, 1996).

ANCOVA-analyser innebär att vi tar hänsyn till en så kallad covariate när vi tolkar resultaten på en beroende variabel (dvs. posttest-resultat). En covariate kan jämföras med en oberoende variabel (x), som kan användas för att förutsäga resultatet på den beroende variabeln (y). Ett konkret exempel är att elevers skrivförmåga i starten av en intervention (pretest, x) sannolikt kan användas för att förutsäga resultaten i slutet av interventionen (posttest, y). När vi då har att göra med grupper vars skrivförmåga initialt skiljer sig åt gör vi ett estimat av i vilken grad pretest-resultaten förutsäger posttest-resultaten för att kompensera för detta när vi jämför posttest-resultaten.

¹⁰ Vi har genomfört parametriska tester, eftersom dataunderlaget är att betrakta som robust; i de allra flesta fall bygger elevresultaten på dubbel parbedömning.

För att kunna göra meningsfulla tolkningar av ANCOVA-analyser, måste vissa kriterier mötas: covariate måste vara mätt före interventionsstart, relationen mellan covariate och beroende variabel måste vara linjär, och det får inte finnas någon interaktion mellan covariate och beroende variabel. Vi har undersökt detta och materialet lever upp till dessa kriterier.

I vår undersökning har vi låtit MP1 utgöra covariate och MP7 beroende variabel. Det innebär att vi kan undersöka effekten av att vara en del av interventionen, samtidigt som vi kompenserar för initiala skillnader mellan experiment- och kontrollgrupper i starten av projektet. Dessa undersökningar har vi dels genomfört på gruppnivå, dels på subgruppnivå.

Subgrupperna har definierats på två sätt. I det första fallet har vi på elevnivå tagit utgångspunkt i resultat på MP1 och skapat tre grupper. I den första gruppen ingår elever vars resultat låg under förväntningsnormen. I den andra gruppen ingår elever vars resultat låg omkring förväntningsnormen (0,5 skalsteg under eller över). I den sista gruppen ingår elever vars resultat låg över förväntningsnormen.

I det andra fallet har vi närmare studerat de 600-skolor som utmärkte sig särskilt positivt, nämligen Aura, Casa, Engen, Lia, Marka, Raset (jfr diskussionen av dessa skolor i avsnitt 3.1). Dessa skolor bryter med den allmänna trenden på 600-nivå, och når genomgående höga resultat. Som vi tidigare konstaterat, finns anledning att utveckla som hypotes att undervisningen i dessa skolor varit kvalitativt annorlunda än i övriga skolor. Av den orsaken är det viktigt att få en uppfattning om hur dessa skolor presterar i förhållande till kontrollskolorna.

ANCOVA-analysen innebär två typer av intressanta resultat. För det första kan vi undersöka om en eventuell skillnad mellan grupper är statistiskt signifikant. För det andra får vi nya medelvärden som justerats med utgångspunkt i vår covariate.

3.3.2.2 Effektstorlek

ANCOVA-analysen anger om det finns en statistiskt signifikant skillnad mellan experiment- och kontrollgruppen. Den anger däremot inte *hur stark* denna skillnad är. För detta finns det dock andra mått, nämligen effektstorlekar. Två vanliga mått på effektstorlek är *partial eta squared* och Cohen's *d*.

Partial eta squared (som också skrivs η_p^2) bygger på ANCOVA-analysen, som ju i sin tur innebär att kontrollera för skillnader i pretest-resultat. Partial eta squared anger hur många procent av variansen i den beroende variabeln som kan förklaras av den oberoende variabeln, det vill säga tillhörighet i antingen projektskola eller kontrollskola. Vedertagna riktlinjer för tolkning av partial eta squared anger att ,01 är en liten effektstorlek, ,06 är en moderat och ,14 är en stor effektstorlek (J. Cohen, 1988). Se också tabell 21 nedan.

Cohen's *d* är ett uttryck för den standardiserade skillnaden mellan två grupper (J. Cohen, 1988; L. Cohen et al., 2011). Det är ett mycket vanligt mått på effektstorlek, vilket gör det möjligt att jämföra en effekt i det här projektet med effekter i andra, liknande projekt.

Ett värde på $d = 1,0$ innebär att skillnaden i medelvärden mellan grupperna är 1 standardavvikelse. Statistiskt anger därför d -värden hur distributionen för grupper överlappar varandra.

Värden på d kan också användas för att göra inferenser. Låter vi elever från olika årskurser genomföra samma test, kan vi mäta effektstorleken för att, allt annat lika, få en indikation på vilken effekt det har att gå i skolan ett år.

Även för d finns riktlinjer som kan användas för att tolka effektens storlek (se tabell 21). Dessa går tillbaka till Cohen själv (J. Cohen, 1988) och anger att effekter omkring $d = 0,20$ kan betraktas som små, $d = 0,50$ som mellanstarka och $d = 0,80$ som starka.

Ett annat sätt att förstå vad ett d -värde representerar, är att konsultera andra undersökningar. En vanlig approximation är att $d = 0,40$ motsvarar ett års undervisning. Detta vill säga att om vi

mäter elever i skolar x och elever i skolar y kommer skillnaden att motsvara $d = 0,40$. Detta tal bygger på olika undersökningar i olika kontexter. Exempelvis så nämner Hattie (2009) att en effektstorlek på 1,0 ”påskyndar” elevers lärande med 2–3 år eller innebär en ökning med 50 %, vilket korresponderar med detta riktvärde. Emellertid är effekten för äldre elever ofta något lägre. Exempelvis rapporterar Eriksen och Roe (2011) lägre skillnader mellan elever i 8. trinn och elever i 9. trinn. Jan-Eric Gustafsson (personlig kommunikation, den 7 mars 2015) betonar att en viss del av effekten också kan tillskrivas den undervisningsberoende faktorn ökad ålder. Detta har i vissa sammanhang visat sig uppgå till en tredjedel av effektstorleken.

För att ha ett så robust jämförelsevärde som möjligt har vi utnyttjat det faktum att 300- och 400-elever, liksom 600- och 700-elever har skrivit samma prov (MP1). Vi har därför beräknat effektstorleken baserat på skillnaden mellan 300- och 400-elever och mellan 600- och 700-elever. Detta återkommer vi till längre fram.

Tabell 21. Riktvärden för effektmått.

	Svag effekt	Moderat effekt	Stark effekt
η_p^2	,01	,06	,14
d	0,20	0,50	0,80

3.3.3 Resultat

3.3.3.1 Hur stark är utvecklingen? Resultat av t-tester

För att undersöka utvecklingen från MP1 till MP7 för såväl projekt- som kontrollskolor har vi genomfört parvisa t-tester. Dessa avslöjar om skillnaden inom grupperna mellan MP1 och MP7 är signifikant. Vidare kan resultaten användas för att skatta den eventuella styrkan i förändringen, genom att beräkna effektstorleken (d). Analysen presenteras i tabell 22.

Tabell 22. Skillnader inom respektive grupp mellan MP1 och MP7, parvisa t-tester.

	MP1		MP7		r^a	df^b	t^c	p^d	ES (d)
	M	SD	M	SD					
300 _k (n = 54)	28,0	5,8	34,7	6,8	,58	53	8,5	,000	1,16
300 _i (n = 75)	30,4	7,1	40,2	8,8	,68	74	13,0	,000	1,54
600 _k (n = 58)	27,8	6,9	31,7	8,2	,50	57	3,8	,000	0,47
600 _i (n = 78)	28,0	6,5	34,1	8,8	,47	77	6,6	,000	0,78

^a Korrelationen (r) mellan MP1 och MP7. ^b Frihetsgrader. ^c t-värde. ^d Sannolikhet (p)

Som vi kan se i tabellen, är det för samtliga grupper en signifikant skillnad mellan MP1 och MP7. Denna kunskap hade vi redan om projektskolorna (se tabell 11), och här kan vi alltså se en tydlig progression också bland kontrollskolorna. Sammantaget får alltså den genomsnittlige eleven ett betydligt bättre resultat vid MP7 än MP1. När vi jämför med data från de nationella utvalgsprövena i skrivning, tycks kontrollskolegrupperna vid MP7 prestera något bättre eller motsvarande den genomsnittlige eleven på 5. respektive 8. trinn.¹¹

¹¹ De nationella utvalgsprövena i skrivning genomförs varje höst på ett nationellt representativt urval elever på 5. respektive 8. trinn. På de nationella utvalgsprövena i skrivning hösten 2014 var medelvärdena – efter linjär transformering till vår skala – följande för 5. och 8. trinn: 31,4 respektive 32,4 poäng. Det betyder att kontrollskolorna på 300-nivå presterade över den nationella nivån 2014, medan kontrollskolorna på 600-nivå presterade tämligen likt nationell nivå 2014. De nationella utvalgsprövena i skrivning finns beskrivna på Skrivsenterets hemsida (<http://www.skrivesenteret.no/skriveprva-emneside/>) och i tekniska rapporter (t.ex. Skar, Evensen, & Iversen, 2015).

De olika effektstorlekarna indikerar emellertid att styrkan på skillnaden mellan MP1 och MP7 varierar.¹² Vi kan se att för projektskolorna på 300-nivån är effekten mycket stor, liksom för kontrollskolorna på samma nivå. Effekten på 600-nivå är inte lika stor. Projektskolorna närmar sig stark effekt och kontrollskolorna ligger strax under vad som betraktas som en moderat skillnad.

Resultaten indikerar stora rörelser för elever som började i 3. trinn (300-eleverna) vid projektets start, oavsett deltagande i intervention eller ej. För eleverna på 6. trinn (600-eleverna) är rörelsen mindre i båda grupper. Den jämförbart låga effektstorleken för kontrollskolor på 600-nivå antyder att den lägre effekt – i termer av måluppfyllelse – som vi sett i föregående kapitel möjligen är att vänta. Eller, annorlunda uttryckt: på 300-nivå är effekten stark även projektmedverkan förutan, något som indikerar en stor förändringspotential i skrivförmågan för denna grupp generellt, oberoende av intervention.

I de kommande avsnitten skall vi närmare granska skillnaderna mellan projekt- och kontrollskolor, liksom skillnader mellan trinn.

3.3.3.2 Har interventionen effekt?

Genom ANCOVA-analyser (Glass & Hopkins, 1996) har vi undersökt om det finns skillnader också mellan grupperna. I tabell 21 ovan ser vi att projektgrupperna har högre medelvärden än kontrollgrupperna på MP7.

Vi kan i tabellen också utläsa tillsynes små skillnader mellan projekt- och kontrollgrupper vid MP1. På 300-nivån är den dock signifikant [$t = 2,0$, $df\ 127$, $p < ,05$]¹³, vilket den inte är på 600-nivån [$t = 0,1$, $df\ 134$, $p = ,108$]. I ANCOVA-analysen tar vi hänsyn till denna initiala skillnad när vi undersöker skillnader på MP7. Resultaten presenteras i tabell 23.

Tabell 23. Skillnader mellan grupper, ANCOVA-analyser.

		M _j MP7 ^b	df	F	p	η_p^2
300: alla	300 _k ^a (n = 54)	35,8				
	300 _{i_a} (n = 75)	39,4	1, 126	10,13	,002	,074
300: urval	300 _k (n = 54)	36,0				
	300 _{i_u} (n = 71)	39,9	1, 122	11,74	,001	,088
600: alla	600 _k (n = 58)	31,7				
	600 _{i_a} (n = 78)	34,0	1, 133	3,13	,079	,023
600: urval	600 _k (n = 58)	32,0				
	600 _{i_u} (n = 71)	33,9	1, 126	2,15	,145	n/s

^a "k" = kontrollskolor, "i_a" = samtliga projektskoleelever, "i_u" = urval av projektskoleelever. ^b Justerat medelvärde vid MP7.

I tabell 23 kan vi se att det vid MP7 är en signifikant skillnad mellan projekt- och kontrollgrupperna på 300-nivå (se "300: alla", [$F(1, 126) = 10,13$, $p < ,002$]). Detta resultat gäller även när vi uteslutit skolorna Dalen, Heia, Quadra och Kilden (se "300: urval", [$F(1, 122) = 11,74$, $p < ,001$]) (jfr kapitel 3.2). Vi kan också notera att medelvärdet justerats, efter att vi har

¹² När man undersöker effektstorleken baserat på två medelvärden som kommer från samma grupp, bör man korrigera för medelvärdenas avhängighet av varandra. För detta används värdet på korrelationen.

¹³ Vi redovisar för intresserade läsare t-värden, F-värden, frihetsgrader (df) och signifikansnivåer. Se Glass och Hopkins (1996) för information om notationer.

kontrollerat för resultat på MP1. Vi ser då att när vi tar det mindre urvalet 300-elever är skillnaden mellan projekt- och kontrollskolor större.

Skillnaden mellan projekt- och kontrollskolor vid MP7 för 600-eleverna är också signifikant, men på 10 %-nivån (se ”600: alla”, [$F(1, 133) = 3,13, p < ,079$]). Sannolikheten att skillnaden är reell är alltså något mindre. Ännu mindre blir den när vi utesluter de fyra svagpresterande skolorna (se ”600: urval”, [$F(1, 126) = 2,15, p < ,145$]) och vi ser också att de justerade medelvärdena ligger något närmre varandra. Att så blir fallet har att göra med oväntat goda resultat från elever från de fyra lågt presterande skolorna på MP7 (se tabell 17 ovan).

Som grupper betraktat, kan vi konstatera en signifikant skillnad mellan projekt- och kontrollskolor på 300-nivå, men inte samma resoluta skillnad på 600-nivå. Tar vi bort de fyra skolorna är skillnaden på 600-nivå inte signifikant ens på 10 %-nivå, vilket indikerar mycket högre resultat än förväntat för eleverna på dessa skolor. Det gäller Heia och Kilden, vars elever presterar betydligt bättre på MP7 än i övrigt.

Samlat sett betyder resultatet att projektet på grupp-nivå verkar ha spelat stor roll för 300-eleverna. Vi skall dock se närmare på subgrupper inom trinngrupperna för att undersöka om resultaten är desamma oberoende av vilken utgångspunkten var vid MP1. Vi skall också se närmare på den grupp 600-skolor som varit högpressterande.

3.3.2.2 Resultat på subgrupp-nivå

Som vi nämnde inledningsvis, har vi också genomfört två typer av analyser som tagit utgångspunkt i en klassificering av MP1-resultaten som under, omkring eller över förväntningsnormerna. Elever som i genomsnitt presterat under 2,5 på alla bedömningsområden, har placerats i kategori 1, elever som presterat mellan 2,5 och 3,5 i kategori 2 och elever som presterat över 3,5 i kategori 3. Urvalet har här varit det mindre, dvs. utan de fyra svagpresterande skolorna. Resultaten presenteras i tabell 24 och 25.

Tabell 24. Skillnader mellan subgrupper, 300-elever, ANCOVA-analyser.

		M_j MP7 ^b	df	F	p	η_p^2
<i>Under</i>	300 _k ^a (n = 29)	31,6				
	300 _{i_u} (n = 24)	34,3	1, 50	2,9	,092	,056
<i>Norm</i>	300 _k (n = 24)	39,0				
	300 _{i_u} (n = 41)	43,6	1, 62	8,9	,004	,125
<i>Över</i>	300 _k (n = 1)	38,4				
	300 _i urval (n = 6)	46,4	1, 4	,5	,535	n/s

^a ”k” = kontrollskolor, ”i_u” = urval av projektskoleelever. ^b Justerat medelvärde vid MP7.

Tabell 25. Skillnader mellan subgrupper, 600-elever, ANCOVA-analyser.

		M_j MP7 ^b	df	F	p	η_p^2
<i>Under</i>	600 _k ^a (n = 33)	28,6				
	600 _{i_u} (n = 38)	30,2	1, 68	,7	,390	n/s
<i>Norm</i>	600 _k (n = 22)	36,0				
	600 _{i_u} (n = 30)	37,2	1, 49	,3	,564	n/s
<i>Över</i>	600 _k (n = 3)	36,3				
	600 _{i_u} (n = 3)	50,7	1, 3	26,0	,015	,897

^a ”k” = kontrollskolor, ”i_u” = urval av projektskoleelever. ^b Justerat medelvärde vid MP7.

Som vi kan se, innebär denna indelning att gruppstorlekar uppgår till mellan 20–30 elever för grupperna ”norm” och ”under”. Detta är i underkant av vad som är eftersträvansvärt när jämförelser av detta slag skall göras. Samtidigt vet vi att även en indelning baserad på exempelvis percentiler skulle leda till små grupper.

Vi kan också se att indelningen resulterar i mycket små grupper av elever som presterade över normen. Gruppernas storlek gör att resultaten varken blir säkra eller generaliserbara, varför vi inte intresserar oss för dem i den vidare analysen. Däremot redovisar vi de värden som erhållits vid beräkningarna.

När vi, med dessa förbehåll, studerar dessa subgrupper ser vi att den enda signifikanta skillnaden återfinns bland de 300-elever som på MP1 hade resultat omkring förväntningsnormen [$F(1, 62) = 8,9, p < ,004$]. Interventionen verkar alltså ha spelat störst roll för dessa elever. Analysen visar också skillnader på 10 % signifikantnivå för 300-gruppen som låg under förväntningsnormen [$F(1, 50) = 2,9, p < ,092$].

För övriga grupper finns ingen signifikant skillnad. Vad gäller eleverna på och under normen på 600-nivå är resultaten tankeväckande. När vi studerar gruppen som presterade omkring normen på 600-nivå ser vi att det skiljer 1 poäng mellan projekt- och kontrollskolor, när medelvärdet justerats. Detta innebär antingen att det för denna grupp inte tycks ha spelat någon roll att medverka i projektet, eller – som vi tidigare diskuterat – att resultaten på MP7 inte fullt ut går att använda som valida indikatorer på skrivförmåga. Så skulle vara fallet om tillräckligt många elever varit omotiverade att genomföra provet, sedan MP7 för många elever genomfördes under andra omständigheter än de tidigare målepunkterna.¹⁴ Vissa elever hade t.ex. bytt skola. Samtidigt innebär en sådan förklaring att kontrollskolegruppens elever inte varit omotiverade. Rimligheten i ett sådant antagande kan därför diskuteras.

Mot bakgrund av de resultat som presenterades ovan har vi också studerat ytterligare en subgrupp bland 600-eleverna. Det gäller elever på skolorna Aura, Casa, Engen, Lia, Marka och Raset, vilka som subgrupp är intressanta eftersom de bryter mönstret. Tillsammans utgör de en grupp skolor som presterar över genomsnittet på 600-nivå. För fyra av skolorna (Aura, Casa, Engen och Lia) gäller dessutom att de presterar väl på 300-nivå.

Totalt har 32 elever från dessa skolor deltagit på både MP1 och MP7. När vi studerar resultaten på MP1 avslöjar sig ännu en intressant aspekt. Vid MP1 är skillnaden mellan detta särskilda urval och kontrollskolorna inte signifikant [$t = 1,3, df = 88, p = ,204$]. Det betyder att detta urval skolor *inte* är att betrakta som extra högpresterande vid MP1.

I tabell 26 återfinns vi medelvärden för MP1 och MP7 för kontrollskolorna och detta särskilda urval projektskolor. Vi återfinns också skattningar av effektstorlekar.

Tabell 26. Skillnader mellan kontrollskolor och särskilt urval, 600-elever.

	MP1		MP7		M_j MP7	ES	
	M	SD	M	SD		η_p^2	d^b
600 _k (n = 58)	27,8	6,9	31,7	8,2	32,1	,124	0,82
600 _{su} ^a (n = 32)	29,7	6,5	38,3	7,9	37,5		

^a 600_{su} = särskilt urval (Aura, Casa, Engen, Lia, Marka, Raset). ^b Beräknat på oberoende t-test.

ANCOVA-analysen visar att skillnaden vid MP7 är signifikant [$F(1, 87) = 12,3, p < ,001$]. Här har vi alltså att göra med en grupp skolor som startar på ungefär samma resultat som kontrollskolorna, som vi vet har följt interventionen och som vid MP7 presterar signifikant

¹⁴ En del projekt lärare har, oberoende av varandra, rapporterat att enskilda elever var omotiverade att skriva i samband med posttestet.

bättre. Vi ser också att effektstorlekarna är betydande ($\eta_p^2 = 0,124$; $d = 0,82$), när vi jämför med de riktlinjer som presenterades ovan.

Att skillnaden mellan detta särskilda urval skolor och kontrollgruppen är signifikant och att effektstorlekarna är betydande är inte förvånande med tanke på urvalet. Självfallet kommer vi att se stora skillnader när vi jämför kontrollskole-elever med elever som representerar skolor som i allmänhet får goda resultat. Det som dock är iögonenfallande här är att de båda jämförelsegrupperna presterar tämligen likt vid projektets start. Det kan därmed finnas stora värden i att mer kvalitativt undersöka vad som har utmärkt interventionen/projektet vid Aura, Casa, Engen, Lia, Marka och Raset.

Analysen av resultat på subgrupp-nivå har givit en mängd intressanta svar. Bland 300-eleverna är det de som presterar omkring normen vid MP1 som verkar ha störst nytta av interventionen, men vi ser också att elever som presterar under normen får bättre resultat än motsvarande kontrollskolegrupper. Här väntar viktiga uppgifter med att genom kvalitativa analyser närmare förklara vilket slags undervisning som är bedriven och hur projektet på både makro- och mikronivå har realiserats på projektskolorna.

Resultaten på 600-nivå indikerar inte motsvarande skillnad mellan projekt- och kontrollgrupper. När vi emellertid gör ett särskilt urval, på skol- och inte individnivå, ser vi att det finns en grupp 600-skolor som presterar signifikant bättre än kontrollskoleeleverna. Även här finns, som sagt, anledning att mer kvalitativt studera projektets realisering.

Vi har till nu huvudsakligen intresserat oss för signifikanstester, som visat att skillnader finns, men inte hur stora de är. I nästa sektion skall vi se närmare på styrkan mätt i effektstorlekar.

3.3.2.3 Effekten av intervensionen

Som vi nämnde inledningsvis, rapporterar vi två effektmått, *partial eta squared* (η_p^2) och *d*. η_p^2 bygger på ANCOVA-analyserna, i vilka vi ju kontrollerat för initiala skillnader mellan jämförelsegrupper. η_p^2 anger hur stor del av variansen som kan förklaras med grupptillhörighet och ju större del som kan det, desto starkare är effekten. I ANCOVA-tabellerna har vi redan kunnat se värden för η_p^2 och att dessa varierar kraftigt. När det gäller effektstorlekar relaterade till signifikanta skillnader mellan projekt- och kontrollskolor, kan vi klassificera dem som starka, moderata eller svaga (jfr ovan). Vi får då resultaten som presenteras i tabell 27.

Tabell 27. Starka, moderata och svaga effektstorlekar.

Stark ES		Moderat ES		Svag ES	
Grupp	η_p^2	Grupp	η_p^2	Grupp	η_p^2
300-elever som presterade runt normen på MP1: Värdet närmar sig ett starkt värde. Resultaten betyder att cirka 12,5 % av variansen i MP7 kan förklaras med grupptillhörighet.	0,125	300-eleverna som helhet: Vi noterar moderata effektstorlekar för 300-gruppen som helhet, oaktat om vi tar det större eller mindre urvalet. Mellan 7,4 % och 8,8 % av variansen kan förklaras med grupptillhörighet i dessa fall.	0,074, 0,088	600-eleverna som helhet: För hela gruppen 600-elever med signifikant skillnad i resultat är effektstorleken svag och 2,3 % av variansen kan förklaras med grupptillhörighet.	0,023
600-elever, särskilt urval: Ett liknande värde får vi när vi jämför det särskilda urvalet av 600-skolor med kontrollskolorna.	0,124				

Dessa siffror förstärker bilden av att hela 300-gruppen tydligt har gynnats av deltagande i projektet, och att det finns subgrupper i 300- och 600-grupperna som gynnats särskilt mycket.

Vi har också beräknat effektstorlekarna i *d*-värden. *D*-värdet uttrycker den standardiserade skillnaden mellan två grupper och kan därmed jämföras med andra *d*-värden. Även för *d* finns riktvärden, men för att skapa en säker referenspunkt har vi undersökt effektstorleken mellan 300- och 400 elever och mellan 600- och 700-elever vid MP1. På så sätt får vi, allt annat lika, en indikation på effekten av ett års skrivundervisning (jfr resonemang ovan). I tabell 28 återfinns skillnaden mellan 300- och 400- samt 600- och 700-elever.

Tabell 28. Skillnader mellan 300- och 400-elever och 600- och 700-elever

	<i>M</i>	<i>SD</i>	<i>M diff.</i>	<i>df</i>	<i>t</i>	<i>p</i>	ES (<i>d</i>)
300 (n = 159)	28,7	6,6					
400 (n = 150)	32,4	7,4					
400-300			3,7	307	4,6	,000	0,53
600 (n = 157)	28,3	6,5					
700 (n = 118)	31,4	6,9					
700-600			3,1	273	3,8	,000	0,46

Kommentar. Skillnaden avser MP1 och är skattad genom oberoende t-tester.

Som vi kan se i tabell 28, är skillnaden mellan trinnen signifikant och *d*-värdena ligger något över värden som brukar rapporteras när man talar om ett års extra undervisning (jfr ovan). Mellan 300- och 400-elever är skillnaden uttryckt i effektstorlek $d = 0,53$ och mellan 600- och 700-elever $d = 0,46$.

Vi ser också att det finns en skillnad mellan småskole- och mellomtrinnen här. Detta återspeglar den skillnad vi noterade ovan, om utveckling *inom* respektive grupp. Generellt verkar alltså mellomtrinns elever ha en mindre brant resultatutvecklingskurva (jfr Hattie, 2009).

I tabell 29 återfinns vi medelvärdeskillnaden och *d*-värdet för skillnaden *mellan* projekt- och kontrollskolor på 300- respektive 600-nivån. Resultaten rapporteras både för det större och det mindre urvalet.

Tabell 29. Skillnader mellan respektive grupp mellan på MP7

	<i>M diff.</i>	<i>Df</i>	<i>t</i>	<i>p</i>	ES (<i>d</i>)
300 _{i_ak} ^a (n = 75, 54)	5,4	127	3,9	,000	0,67
300 _{i_uk} (n = 71, 54)	6,2	123	4,5	,000	0,81
600 _{i_ak} (n = 78, 58)	2,4	134	1,6	,108	n/s
600 _{i_uk} (n = 71, 58)	2,5	127	1,6	,104	n/s

Kommentar. Skillnaden är skattad oberoende t-tester och urvalet är elever som genomfört både MP1 och MP7.

^a "i_ak" = samtliga projektskoleelever och kontrollskolegruppen, "i_uk" = urval av projektskoleelever och kontrollskolegruppen.

I tabellen redovisas skillnaden mellan projektgrupperna (både det större och mindre urvalet) och kontrollgrupperna. För båda grupper 300-elever (dvs. båda urvalen) är effektstorleken att betrakta som stor; att ingå i projektet verkar spela positiv roll. Detta är i linje med vad vi sett när vi genomfört ANCOVA-analyserna. På 300-nivå överstiger effektstorleken den som är förbunden med ett års extra skolgång. För det större urvalet på 300-nivå motsvarar effekten, tillspetsat, drygt ett års extra undervisning ($d = 0,67$ mot $d = 0,53$ för 300–400). För det mindre urvalet motsvarar effekten, återigen tillspetsat, cirka ett och ett och ett halvt års extra undervisning ($d = 0,81$ mot $d = 0,53$ för 300–400).

På 600-nivån, liksom på 300-nivån, förstärks den bild som vi fått från tidigare analyser. För det första är skillnaden på MP7 icke-signifikant när vi inte kontrollerar för MP1 och för det andra är effekten inte lika hög. Vi skall dock komma ihåg vår initiala analys som visade att effekten bland kontrollskolorna på 600-nivå var lägre än på 300-nivå och vi ser att effekten mellan 600- och 700 är lägre än den mellan 300- och 400.

När vi, avslutningsvis, återvänder till de två grupper som i analysen utmärkt sig – elever på 300-nivå som presterar runt normen, och det särskilda urvalet på 600-nivå – får vi värden på $d = 0,82$ i båda fallen. För båda urval motsvarar effekten, tillspetsat, i närheten av ett och ett halvt respektive två års undervisning. I tabell 30 återfinns en sammanställning av resultaten.

Tabell 30. Starka, moderata och svaga effektstorlekar

Stark ES		Moderat ES	
Grupp	<i>d</i>	Grupp	<i>d</i>
300-elever (urval)	0,81 (0,53)	300-elever	0,67 (0,53)
300-elever på normen	0,82 (0,53)		
600-elever särskilt urval	0,82 (0,46)		

Kommentar. Inom parentes syns referensvärdet för det aktuella trinnet.

Generellt indikerar dessa och tidigare rapporterade resultat att de flesta 300-eleverna har haft ett stort utbyte av projektet. Elever på småskoletrinnet verkar alltså ha mycket att vinna på systematisk skrivundervisning och skrivbedömning.

Utvecklingen för 600-elever ser annorlunda ut och vi ser, som sagt, att skillnaden mellan MP1 och MP7 mellan 600- och 700-elever är mindre. Emellertid visar analysen av det särskilda urvalet högpresterande 600-skolor att det inte finns någon anledning att hävda att 600-eleverna har haft mindre utvecklingspotential.

3.4 Variasjon og stabilitet i elevenes skriveferdigheter

För att undersöka i vilken utsträckning elevernas resultat på olika målepunkter kan förutspås med hjälp av resultatet på MP1 har vi genomfört bivariata regressionsanalyser.¹⁵

Syftet med en regressionsanalys är att undersöka hur stor del av variansen i en beroende variabel (y) kan förklaras med en oberoende variabel (x). Resultatet av analysen är en regressionskvation, som vi kan använda när vi vill förutspå y -värdet vid ett givet värde på x .

I tabell 31 framgår hur stor del av variansen på respektive målepunkt som kan förklaras med resultaten på målepunkt 1.

Tabell 31. Förklarad varians

	MP2, r ²	MP3, r ²	MP5, r ²	MP6, r ²	MP7, r ² (kontr)
300	,362	,283	,305	,127	,370 (,325)
600	,153	,157	,208	,330	,228 (,244)

Kommentar. I analysen har MP1 använts som oberoende variabel

I tabellen framkommer ett intressant mönster. För 300-elever gäller i huvudsak att mellan 30 % och 37 % av variansen förklaras med utgångspunkt i MP1. Förklarad varians för

¹⁵ Analysen bygger på det tidigare nämnda mindre urvalet (dvs. utan skolorna Dalen, Heia, Kilden og Quadra).

kontrollskoleeleverna på MP7 bidrar till att bekräfta att resultatet inte bara gäller projektskoleeleverna. Vi ser dock att för MP6 kan bara 12 % av variansen förklaras, vilket antyder att prediktionen inte kan förklaras med personliga, fasta egenskaper utan att de andra målepunkterna delar något som MP6 inte har. Det kan exempelvis vara så att uppgiften eller undervisningen som omgärdat uppgiften vid MP6 på ett markant sätt har skiljt sig från uppgifter och undervisning förbundna med övriga målepunkter. Kommande undersökningar, exempelvis av uppgiftstyper, kan generera resultat för nya regressionsanalyser.

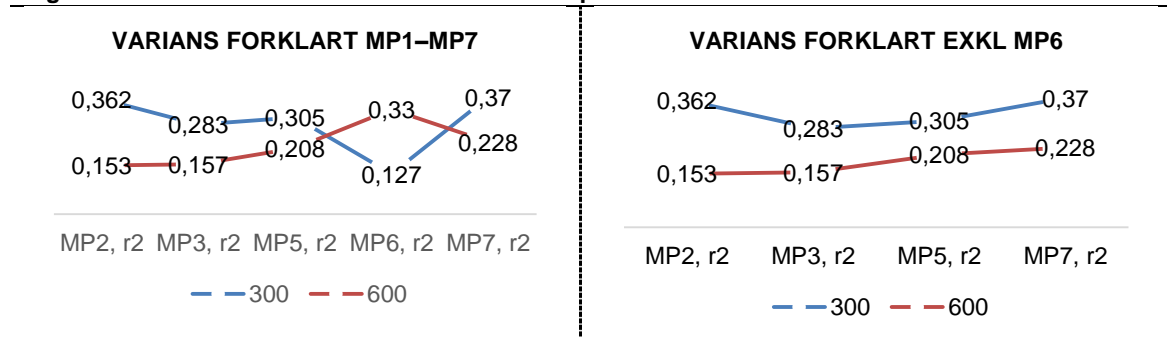
För 600-elever gäller att mellan 15 och 33 % av variansen kan förklaras. Även på 600-nivå uppvisar kontrollskoleeleverna ett liknande resultat som projektskolorna. En intressant skillnad mot 300-eleverna är att 600-eleverna är mindre stabila – en lägre andel av variansen går att förklara med MP1 som oberoende variabel. Samtidigt har projektet sammantaget mindre verkan på 600-eleverna.

Paradoxalt nog har alltså den elevgrupp vars resultat bäst kan förutspås också störst nytta av projektet. En möjlig förklaring kan förstås vara att effekten är stor på gruppnivå på 300-eleverna, vilket gör att gruppen som sådan rör sig uppåt. Detsamma sker inte bland 600-eleverna.

En annan intressant skillnad mellan 300- och 600-eleverna är MP6. För 300-eleverna innebär den att en lägre andel varians kan förklaras, och för 600-eleverna att en för gruppen onormalt hög andel kan förklaras.

I figur 5 ser vi att när vi tar bort MP6 är de båda grupperna någorlunda parallella. Samtidigt indikerar figur 5 liten grad av stabilitet. Det är knappast tal om fullständiga förutsägelser; en stor del av variansen är oförklarad.

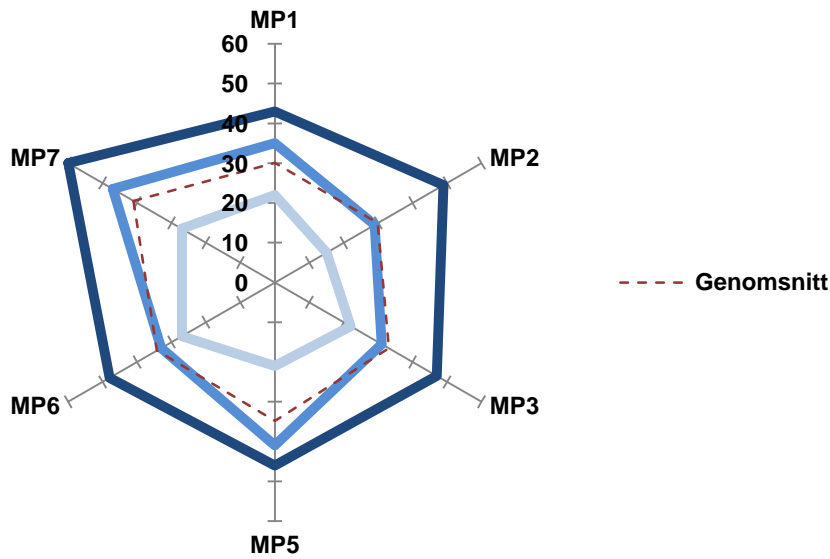
Figur 5. Varians för 300- och 600-elever alla målepunkter förklarad.



Siffrorna indikerar att det, som ovan nämnt, finns egenskaper vid respektive målepunkt som är av stor betydelse (se också utvecklingsgrafer, figur 3 och figur 4).

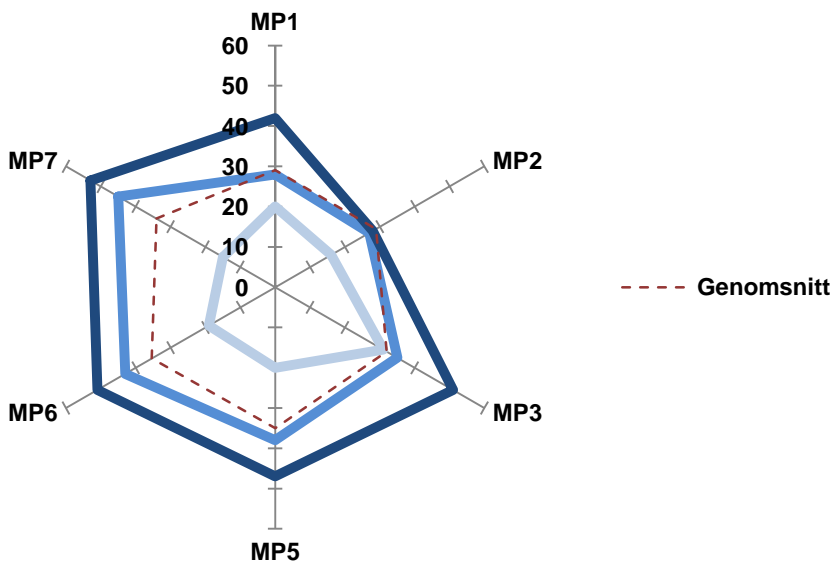
I figurerna nedan har vi illustrerat individuell variation från MP till MP. De olika blå strecken representerar enskilda elever som ligger på (ljusblå färg) eller kraftigt under (svagt blå färg) eller över förväntningsnormen (mörkblå färg). Den röda streckade linjen representerar projektskolegruppens medelvärde.

Figur 6. Individuell variation bland 300-elever



Figur 6. De blå linjerna avser 300-elever, som presterar under, på eller över normen. Den röda streckade linjen avser genomsnittet.

Figur 7. Individuell variation bland 600-elever



Figur 7. De blå linjerna avser 600-elever, som presterar under, på eller över normen. Den röda streckade linjen avser genomsnittet.

Figurerna 6 och 7 hjälper oss att studera en del av den *individuella* variationen från målepunkt till målepunkt. Vi kan exempelvis se att för 300-eleverna ligger den elev som presterar bäst nästan alltid bättre till än sina medelever, men på MP5 är det ingen skillnad i resultat i jämförelse med den elev som i genomsnitt ligger på normen. På 600-nivå tecknar sig en bild av större instabilitet, vilket är förväntat. För t.ex. MP3 – för vilken MP1 kan förklara 15 % av variansen – ser vi att både den lägst presterande och medeleven får ungefär samma resultat. För MP6 är avstånden mellan eleverna mer lika de för MP1, vilket förklaras av den stora andelen förklarad varians.

Kommande undersøkingar bör, som sagt, inriktas på att undersöka vad som kvalitativt kännetecknar de olika målepunkterna. Skulle det exempelvis visa sig att uppgifterna skiljer sig åt, kan det bidra till att förklara det vi ser och också indikera vilken typ av uppgifter som minskar respektive ökar avståndet mellan elever.

Avslutningsvis bör understrykas att resultatprofilerna också påminner oss om att skrivförmåga knappast är en stabil förmåga. För att studera enskilda individers skrivande bör kommande undersøkingar baseras på ett rikt urval texter från samma elev (jfr Bouwer, Béguin, Sanders, & van den Bergh, 2015).

3.4 Forholdet mellom skriveferdighetsnivå og eksterne variabler (kjønn, morsmål, målform, fødselsmåned)

Vi har i forrige kapittel avdekket en rekke mønstre som peker på hva som kan forklare kvaliteten på og variasjonen i elevenes skriveferdigheter. Vi har sett at 300-elevene er mer stabile enn 600-elevene. Mer av variansen i elevenes skriveprestasjoner kan forklares med 300-elevenes ferdighetsnivå før intervensjon, enn med 600-elevenes skriveferdighetsnivå på samme tidspunkt. Dessuten har vi påvist at noen klasser og skoler har hatt betydelig bedre utbytte av intervensjonen enn andre klasser og skoler. Noen få skoler/klasser/elever har hatt lite eller ikke noe utbytte av intervensjonen.

Vi har også påvist at kvaliteten i elevenes tekster kan variere fra målepunkt til målepunkt. Dette funnet peker i retning av at kvalitetsvariasjonen på elevenes tekster kan forklares med oppgave, tematikk, skrivehandling, interesse, motivasjon og andre faktorer. Vi skal komme tilbake til slike mulige forklaringer på variasjonen i skriveferdighet noe lenger framme i rapporten.

I Normprosjektet er det samlet inn informasjon om egenskaper ved elevene som teoretisk kan tenkes å forklare hvordan elever skriver, men som ikke er relevant for hvordan intervensjonen er presentert og gjennomført i undervisningssituasjonen. Imidlertid er dette egenskaper som ifølge tidligere forskning ofte samvarierer med skriveferdighet, eller kan antas å kunne forklare skriveferdighetsvariasjon på en eller annen måte. Følgende slike «eksterne» egenskaper er samlet inn for alle intensivelever: kjønn, det første språket eleven lærte hjemme (morsmål), målform, skole og fødselsmåned.

Av disse egenskapene er det forskningsbelegg for å hevde at kjønn og elevenes morsmål til en viss grad kan forklare skriveprestasjoner. I tidligere forskning (Berge, Evensen, Hertzberg, & Vagle, 2005; Ministry of Education, 2012; Skolverket/Siris, 2015; Thygesen et al., 2007; Utdanningsdirektoratet/Skoleporten, 2015) er sammenhengene mellom skriveferdigheter og kjønn tilsynelatende entydige: Jenter skriver bedre enn gutter når kulturkonteksten er skolens skriveopplæring.

En tilsvarende tendens er ikke avdekket når det gjelder forholdet mellom elever med norsk som førstespråk og L2-elever (Bakken, 2010). Men studiene når det gjelder forholdet mellom morsmål og skriveferdigheter på norsk er få.

Når det gjelder elevenes målform, finnes det få eller ingen vitenskapelige studier av sammenhengen mellom elevens målform og skriveferdigheter. Nynorskelever har få språklige forbilder i storsamfunnets skriftkultur. Siden elever med nynorsk som hovedmål er en minoritet i det norske språksamfunnet, har vi undersøkt om det lar seg påvise en systematisk sammenheng mellom intensivelevens målform og elevenes skriveferdigheter.

Sammenhengen mellom skriveferdighet og fødselsmåned er regnet ut for å se om alder på trinnet samvarierer systematisk med skriveferdighet. Som vi har avdekket tidligere (se tabell 12

og 28), er det klare kvalitetsforskjeller i skriveferdighet mellom ulike trinn i intensivutvalget ved MP1. En systematisk undersøkelse av forholdet mellom fødselsmåned og trinn kan avdekke om det er trinnet og klassen eleven tilhører, på den ene sida, eller elevenes alder på den andre som kan forklare skriveferdighetsnivået. Viss det avdekkes at elevenes fødselsmåned samvarierer signifikant med skriveferdighet, samtidig med at det er en systematisk forskjell mellom skriveprestasjoner mellom trinn, finnes det belegg for å forklare skriveutviklingen med en mulig kognitiv og/eller biologisk modenhetsutvikling. Er det ikke tilfellet, må andre alternative forklaringer regnes som mer rimelige og relevante.

Siden vurderingene i Normprosjektet er basert på en ordinalskala der teksters kvaliteter rangeres etter en femdelt skala, har vi brukt ikke-parametriske analyser for å studere forholdet mellom de omtalte eksterne variablene og elevens skriveferdigheter. For relasjonen mellom skriveprestasjon og målform, morsmål og kjønn er det brukt Mann-Whitney U-test. For relasjonen mellom skriveprestasjon og fødselsmåned er det brukt Kruskal-Wallis-test (Pallant, 2013). Vi rapporterer antallet elever som er vurdert (n), rangeringssnitt og signifikans der $p < .10$ i tabellene som følger.

3.4.1 Forholdet mellom kjønn og skriveferdighet

I tabell 32 er resultatene for forholdet mellom skriveprestasjon og kjønn presentert når det gjelder prosjektelevene (300-elevene). Vi ser at det for 300-elevene er en signifikant forskjell mellom gutter og jenters skriveprestasjoner på pre- og posttestene (MP1 og MP7). For de andre målepunktene er det ingen signifikante forskjeller. Vi ser også at den innbyrdes rangeringen mellom gutter og jenter (snitt rank) på de fleste målepunkter er til jentenes fordel. På MP6 rangeres guttenes skriveprestasjoner noe over jentenes ferdigheter.

Tabell 32. Mann-Whitney U-test: kjønnsforskjeller i skriveprestasjon for 300-prosjektelever

	N		Snitt rank		Signifikans
	Gutt	Jente	Gutt	Jente	
MP1	38	55	34,5	55,7	.00
MP2	38	54	45,2	47,5	n/s
MP3	41	53	42,4	51,5	n/s
MP5	40	53	43,8	49,4	n/s
MP6	38	50	46,7	42,8	n/s
MP7	33	46	33,6	44,6	.04

I tabell 33 er resultatene for forholdet mellom skriveprestasjon og kjønn presentert når det gjelder kontrollskoleelevene. Også for denne gruppa skriver jenter bedre enn guttene på pre- og posttestene (MP1 og MP7).

Tabell 33. Mann-Whitney U-test: kjønnsforskjeller i skriveprestasjon for 300-kontrollskoleelever

	N		Snitt rank		Signifikans
	Gutt	Jente	Gutt	Jente	
MP1	36	30	29,8	38,1	.03
MP7	31	23	22,7	34,0	.01

Disse funnene peker i retning av at jenter skriver bedre på prøver. Det vil si i testsituasjoner der elevene ikke følges systematisk opp gjennom formativ vurdering. Når situasjonen er undervisning der elevene skriver utkast- og presentasjonstekster, utjevnes forskjellene mellom

gutter og jenters skriveprestasjoner. Det ser altså ut til at gutteskriverne underpresterer i testsituasjoner. Funnene gir ikke støtte for å hevde at jenter generelt skriver bedre enn gutter uansett skrivesituasjon.

Finner vi det samme mønsteret for 600-elevene? Tabell 34 antyder at for 600-elevene på prosjektskolene, er det et annet mønster enn vi har sett for 300-elevene. Målt i rangeringssnitt skriver jentene på alle målepunkt bedre enn guttene, men det er først på slutten av prosjektintervensjonen at kjønnsforskjellene er signifikant. Bare på MP6 og MP7 skriver jentene signifikant bedre enn guttene.

Tabell 34. Mann-Whitney U-test: kjønnsforskjeller i skriveprestasjon for 600-prosjektelever

	N		Snitt rank		Signifikans
	Gutt	Jente	Gutt	Jente	
MP1	43	47	41,3	49,3	n/s
MP2	46	44	43,7	47,4	n/s
MP3	46	44	41,6	49,6	n/s
MP5	45	42	40,4	47,9	n/s
MP6	44	44	36,0	53,0	.00
MP7	43	41	36,2	49,1	.01

Når det gjelder 600-elevene på kontrollskolene, bekreftes heller ikke det mønsteret som vi avdekket for 300-elevene. Det er ingen signifikant forskjell i skriveprestasjon som kan forklares med elevens kjønn.

Tabell 35. Mann-Whitney U-test: kjønnsforskjeller i skriveprestasjon for 600-kontrollskoleelever

	N		Snitt rank		Signifikans
	Gutt	Jente	Gutt	Jente	
MP1	31	36	32,0	35,7	n/s
MP7	27	31	26,6	32,0	n/s

Konklusjonen for analysene av forholdet mellom skriveprestasjon og kjønn er at det i løpet av intervensjonen ikke kan påvises noen systematisk sammenheng mellom skriveprestasjon og elevens kjønn. Riktignok skriver jenter som gruppe betraktet oftest noe bedre enn guttene, men de skriver ikke forskjellig nok til at vi kan hevde med statistisk sikkerhet at elevens kjønn forklarer de tilsynelatende prestasjonsforskjellene. Andre typer tester, slik som ikke-parametriske means-test og parametriske t-test, gir samme resultat.¹⁶

Bare i to tilfeller er det forekomster av kjønnsforskjeller der systematiske forklaringer er rimelige og relevante. Både når det gjelder 300- og 600-elevene, underpresterer guttene i skrivesituasjoner som er preget av prøving/testing og ikke av undervisning. Med unntak av ett tilfelle (MP6 for 600-elevene) er det i testsituasjoner at jenter skriver signifikant bedre enn guttene. En mulig forklaring på dette mønsteret vil være at egenskaper ved undervisningen bidrar til å gjøre kjønn irrelevant som forklaring på skriveferdigheten. Forklaringen ville da kunne være en slags kulturell kontekstsensitivitet som varierer med kjønn. Gutter presenterer bedre i undervisningssituasjoner enn i testsituasjoner.

For 600-elevene ser det likevel ut til at forskjellen i gutters og jenters skriveferdigheter kan øke med alder. Derfor kan vi ikke se bort fra at kjønnsforskjellene i skriveprestasjonen mellom

¹⁶ Siden resultatene av means-testen og t-testen ikke avviker fra Mann-Whitney U-test, rapporterer vi ikke resultatene av disse testene.

600-elevene som er avdekket på MP6 og MP7 kan forklares med forskjeller i modenhetsutvikling. I så fall ville forklaringen være naturlig utvikling blant elevene, og ikke trekk ved vurderings- og undervisningssituasjonen.

3.4.2 Forholdet mellom morsmål og skriveferdighet

Tabell 36 gir en oversikt over skriveferdighetene for prosjektskoleelever med norsk som morsmål på den ene sida og L2-elever på den andre. Vi ser at for 300-elevene er det en systematisk forskjell mellom de to elevgruppene på pretesten og i løpet av det første intervensjonsåret (MP1-MP3). Men i det andre intervensjonsåret er det ingen forskjell i skriveprestasjon mellom elevgruppene. Målt i rangeringssnitt skriver L2-elever til og med noe bedre enn elever med norsk som morsmål våren 2014 (MP5 og MP6).

Tabell 36. Mann-Whitney U-test: språkforskjeller i skriveprestasjon for 300-prosjektelever

	N		Snitt rank		Signifikans
	norsk	L2	norsk	L2	
MP1	76	11	47,3	31,5	.04
MP2	75	13	46,6	32,1	.06
MP3	77	13	48,2	29,4	.02
MP5	78	11	44,9	45,7	n/s
MP6	74	10	41,8	47,9	n/s
MP7	70	5	38,3	33,1	n/s

Studerer vi resultatene fra kontrollelevne som er 300-elever, viser resultatene i tabell 34 ingen systematisk forskjell. På pretesten (MP1) var kontrollelevne med norsk som morsmål noe bedre til å skrive enn L2-elevne. På posttesten var forholdet det motsatte.

Tabell 37. Mann-Whitney U-test: språkforskjeller i skriveprestasjon for 300-kontrollskoleelever

	N		Snitt rank		Signifikans
	norsk	L2	norsk	L2	
MP1	59	7	34,0	29,4	n/s
MP7	48	6	22,7	34,0	n/s

En mulig forklaring på disse noe merkelige mønstrene kan være at intervensjonen bidrar til å utjevne forskjellene mellom elever med norsk som morsmål og L2-elever. Men en slik antakelse ser ikke videre ut til å kunne bekreftes om vi ser på forholdet mellom skriveferdighet og elevenes morsmål for 600-elevne. Som tabell 38 viser, varierer prestasjonene til L2-elevne ganske mye fra målepunkt til målepunkt.

Tabell 38. Mann-Whitney U-test: språkforskjeller i skriveprestasjon for 600-prosjektelever

	N		Snitt rank		Signifikans
	norsk	L2	norsk	L2	
MP1	79	8	45,1	33,4	n/s
MP2	78	9	44,1	42,7	n/s
MP3	77	10	46,2	25,1	.01
MP5	75	10	43,2	41,8	n/s
MP6	77	9	45,3	28,4	.05
MP7	72	10	41,5	41,7	n/s

Når det gjelder kontrollskoleelevene blant 600-elevene, er L2-elevene signifikant dårligere til å skrive enn elevene med norsk som morsmål kun på pretesten (MP1). Det framgår av tabell 39.

Tabell 39. Mann-Whitney U-test: språkforskjeller i skriveprestasjon for 600-kontrollskoleelever

	N		Snitt rank		Signifikans
	norsk	L2	norsk	L2	
MP1	56	11	36,8	20,0	.01
MP7	48	10	31,2	21,5	.10

Generelt gir ikke resultatene vi har presentert om forholdet mellom morsmål og skriveferdighet, støtte for noen antakelser om at L2-elevene skriver signifikant dårligere enn elever med norsk som morsmål. Likevel er det påfallende at skriveferdighetene til L2-elevene varierer ganske mye fra målepunkt til målepunkt. Mulige forklaringer på denne kvalitetsvariasjonen kan være undervisningen og tilretteleggingen av skriving i de klassene der det er mange L2-elever. Disse spørsmålene krever mer kvalitative studier. Det er også grunn til å anta at gruppa som består av elever med annet morsmål enn norsk, ikke er en kulturelt og språklig homogen gruppe. Mer inngående analyser av L2-elevenes språklige og kulturelle bakgrunn kan bidra til mer innsikt om den påfallende variasjonen i skriveferdighet vi har påvist i analysene i dette avsnittet.

3.4.4 Forholdet mellom målform og skriveferdighet

Av tabellene 40 til 43 ser vi at det med noen svært få unntak ikke er signifikante forskjeller i skriveferdighet mellom bokmåselever og nynorskelever. For 300-elevene finner vi ingen forskjeller verken blant kontrollskoleelevene eller prosjektskoleelevene. Ser vi på rangeringssnittet for 300-elevene, rangeres kvaliteten på nynorskelevens tekster noe høyere enn bokmåselevens.

Tabell 40. Mann-Whitney U-test: målformforskjeller i skriveprestasjon for 300-prosjektelever

	N		Snitt rank		Signifikans
	bm	nn	bm	nn	
MP1	77	16	45,2	55,4	n/s
MP2	77	15	44,8	55,3	n/s
MP3	77	17	45,9	54,8	n/s
MP5	76	17	46,5	49,1	n/s
MP6	71	17	44,5	44,6	n/s
MP7	63	16	31,1	43,3	n/s

Tabell 41. Mann-Whitney U-test: målformforskjeller i skriveprestasjon for 300-kontrollskoleelever

	N		Snitt rank		Signifikans
	bm	nn	bm	nn	
MP1	56	10	32,1	41,4	n/s
MP7	47	8	28,2	26,9	n/s

For 600-elevene er det både blant kontrollskoleelevene og prosjektelevene en signifikant forskjell i skriveferdighet mellom bokmåselever og nynorskelever på pretesten (MP1). I begge

tilfeller er bokmåselevne bedre til å skrive enn nynorskelevne. For de andre målepunktene er det ingen forskjell mellom nynorskelevne og bokmåselevne skriveferdigheter

Tabell 42. Mann-Whitney U-test: målformforskjeller i skriveprestasjon for 600-prosjektelever

	N		Snitt rank		Signifikans
	bm	nn	bm	nn	
MP1	74	16	48,2	33,0	.03
MP2	73	17	45,0	47,7	n/s
MP3	74	16	45,2	49,6	n/s
MP5	71	16	40,4	46,7	n/s
MP6	71	17	44,1	43,7	n/s
MP7	68	16	42,8	41,2	n/s

Tabell 43. Mann-Whitney U-test: målformforskjeller i skriveprestasjon for 600-kontrollskoleelever

	N		Snitt rank		Signifikans
	bm	nn	bm	nn	
MP1	57	10	36,4	20,5	,02
MP7	48	10	29,5	29,3	n/s

På grunnlag av disse analysene kan vi hevde at det ikke er statistisk støtte for å hevde at bokmåselever og nynorskelever har utviklet skriveferdigheter som skiller seg kvalitativt fra hverandre. Også her vil videre kvalitative studier kunne bidra til å utdype og nyansere resultatene.

3.4.5 Forholdet mellom fødselsmåned på trinnet og skriveferdighet

En interessant undersøkelse er det å forsøke å avdekke forskjeller i skriveferdighet som kan forklares med elevers alder innenfor de aktuelle trinnene. Som tidligere påvist (tabell 12, side 20) er det signifikante forskjeller i skriveferdighet mellom 300- og 400-elever på den ene sida og mellom 600- og 700- elever på den andre. Som Kruskal-Wallis-testene i tabellene fra 44 til 47 dokumenterer, er det ingen systematisk samvariasjon mellom skriveferdighet og fødselsmåned på trinnet for noen subgrupper verken på prosjekt- eller kontrollskoler, med ett unntak for pretesten (MP1) på kontrollskolene.

Tabell 44. Kruskal-Wallis-test: fødselsmånedforskjeller i skriveprestasjon for 300-prosjektelever

	N	Signifikans
MP1	93	n/s
MP2	92	n/s
MP3	94	n/s
MP5	93	n/s
MP6	88	n/s
MP7	79	,10

Tabell 45. Kruskal-Wallis-test: fødselsmånedforskjeller i skriveprestasjon for 300-kontrollskoleelever

	N	Signifikans
MP1	66	,01
MP7	55	n/s

Tabell 46. Kruskal-Wallis-test: fødselsmånedforskjeller i skriveprestasjon for 600-prosjektelever

	N	Signifikans
MP1	89	n/s
MP2	89	n/s
MP3	89	n/s
MP5	87	n/s
MP6	88	n/s
MP7	84	n/s

Tabell 47. Kruskal-Wallis-test: fødselsmånedforskjeller i skriveprestasjon for 600-kontrollskoleelever

	N	Signifikans
MP1	67	n/s
MP7	58	n/s

En mulig forklaring på disse analyseresultatene er at undervisningen i klassen og på trinnet visker ut aldersforskjeller mellom elevene. Skriveprestasjonene ser ut til å preges mer av undervisningen elevene er utsatt for, enn av når på året de er født.

4 Drøfting av og videre forskning på funnene i Normprosjektet

4.1 Intervensjonens læringseffekt: en oppsummering

En generell konklusjon vi kan trekke på grunnlag av analysene som er dokumentert i denne rapporten, er at intervensjonen som har blitt gjennomført på de 20 prosjektskolene i regi av Normprosjektet, har hatt læringseffekt. Normprosjektets intervensjon har utviklet elevenes skriveferdighet med en effekt som tilsvarer fra et halvt til to ekstra undervisningsår sammenliknet med kontrollskolene, avhengig av hvilken elevgruppe som undersøkes. Imidlertid varierer effektstørrelsen avhengig av subgruppe. Det er også forekomster av ingen eller svært liten læringseffekt blant subgrupper (klasser, skoler, elevgrupper, elever) i prosjektet.

På de fleste skolene i prosjektet har intervensjonen bidratt til å utvikle elevenes skriveferdigheter til et kvalitetsnivå som de ikke ville vært på om de ikke hadde deltatt i prosjektet. Elevene som har utviklet skriveferdighetene ved å delta i prosjektet, utgjør om lag 90 % av intensivelevne. På en rekke skoler og/eller klasser er effekten av prosjektet sterk (J. Cohen, 1988). Studier av hvordan intervensjonen er gjennomført på disse skolene og/eller klassene, og av elevenes tekster og lærernes summative og formative vurderinger, vil kunne gi mer informasjon om hva det er som gjør at intervensjonen virker spesielt sterkt der. På noen få skoler har deltakelsen i Normprosjektet hatt liten eller ingen effekt. Elevene som ikke har utviklet skriveferdighetene påviselig i løpet av prosjektet, utgjør om lag 10 % av intensivelevne. På disse skolene er det grunn til å undersøke kvalitativt hvordan intervensjonen har foregått, og om den reelt sett kan sies å ha vært gjennomført.

Når vi i analysene skiller mellom 300- og 600-elever, finner vi interessante forskjeller i hvordan intervensjonen har virket. Generelt har 300-elevne fra småskoletrinnet utviklet skriveferdighetene i betydelig grad gjennom intervensjonen. For 600-elevne er det en stor variasjon mellom skoler, klasser og elever. 300-elevne er, sett i forhold til 600-elevne, en mer homogen og mer utviklingsorientert/læringsvillig elevgruppe. 600-elevne er mer sammensatt. Et stort antall 600-elever har utviklet skriveferdighetene i betydelig grad ved å delta i prosjektet, mens andre 600-elever har hatt lite eller ikke noe utbytte av å inngå i Normprosjektet.

Prosjekt- og kontrollelever som begynte på 3. trinn skoleåret 2012/2013 (300-elever) og 6. trinn skoleåret 2012/2013 (600-elever), tok utvalgsprøven i skrijving som posttest (MP7) høsten 2014. Prosjektelevne tok skriveprøven som posttest et halvt år etter at to års intervensjon var avsluttet. At prosjektelevne som gruppe betraktet presterer signifikant bedre på posttesten enn kontrollelevne, innebærer at vi vil hevde at skriveferdighetene elevene har utviklet gjennom intervensjonsperioden, er vedvarende.

De viktigste analysefunnene når det gjelder effekten av Normprosjektet på utviklingen av elevenes skriveferdigheter, delt opp i 300- og 600-elever, kan oppsummeres slik.

Effekten på 300-elevnes skriveferdigheter:

- På de fleste prosjektskoler har intervensjonen signifikant læringseffekt for 300-elevne.
- Effekten ved å delta i prosjektet innebærer for 300-elevne minst ett års ekstra undervisning i forhold til kontrollskolelevne ($p < .05$).

- Det er en viss variasjon i hvilke undergrupper av 300-elever som i størst grad har utviklet skriveferdighetene gjennom å delta i prosjektet. Elevene som ved pretest (MP1) presterer omkring normen, har hatt den sterkeste ferdighetsutviklingen. For denne undergruppa av 300-elever, tilsvarer effekten av å delta i prosjektet opp mot to års ekstra undervisning i forhold til kontrollskolelevnene.

Effekten på 600-elevenes skriveferdigheter:

- Elever som er fulgt gjennom 6. og 7. skoleår, har bedret egne skriveferdigheter signifikant ved å delta i Normprosjektet. For 600-elevnene som gruppe betraktet, tilsvarer effekten minst ett halvt års ekstra undervisning i forhold til kontrollskolelevnene ($p > .10$).
- Det er en stor variasjon i hvilke undergrupper av 600-elever som har utviklet skriveferdighetene gjennom å delta i Normprosjektet. Det er særlig ved seks skoler (Aura, Casa, Engen, Marka, Lia og Raset) at elevene har utviklet ferdighetene i betydelig grad. For elevene på disse skolene tilsvarer effekten opp mot to års ekstra undervisning i forhold til kontrollskolelevnene.
- Ved fire skoler (Dalen, Heia, Kilden, Quadra) har ingen av intensivelevnene fra 600-gruppen påviselig effekt av å delta i prosjektet.

4.2 Mulige forklaringer på intervensjonens læringseffekt: eksterne variabler

Vi har i analysene i denne rapporten undersøkt om elevenes ferdighetsutvikling kan forklares med elevenes skriveferdigheter før intervensjonen. Når det gjelder 300-elevnene, kan mellom 13 og 37 % av variansen i skriveprestasjonene gjennom intervensjonens to år forklares på grunnlag av hvordan de skrev før intervensjonen ble gjennomført. Funnet er basert på en undersøkelse av forholdet mellom skriveferdigheten vurdert ved pretest (MP1) og de øvrige skrivehendelsene (MP2–MP7). Tilsvarende verdier for 600-elevnene er mellom 15 og 23 %. Derfor er det en ganske stor del av variasjonen i elevenes skriveferdigheter som må forklares med andre forhold enn elevenes skriveferdigheter før intervensjonen.

Basert på etablert forskning og utbredte erfaringsbaserte (doxiske) hverdagsantakelser er forholdet mellom skriveferdigheter før, gjennom og etter hele intervensjonsperioden på den ene sida og eksterne variabler som kjønn, målform, morsmål og fødselsmåned på den andre, undersøkt. Slik har vi avdekket om de eksterne variablene samlet eller enkeltvis kan bidra til å forklare elevenes skriveferdighetsnivå. Vi kan oppsummere analysene der forholdet mellom de eksterne variablene og skriveferdighet er undersøkt, slik:

- Elevenes skriveferdigheter samvarierer ikke systematisk med elevens *kjønn* bortsett fra i de fleste testsituasjoner (pretest og posttest). For 600-elevnenes del er det også en tendens til at jentene skriver bedre mot slutten av og etter intervensjonen)
- Elevenes skriveferdigheter samvarierer ikke systematisk med *når på året eleven er født*.

- Elevenes skriveferdigheter samvarierer ikke systematisk med elevenes *målform*.
- Elevenes skriveferdigheter samvarierer ikke systematisk med elevenes *morsmål*, *unntatt ved enkelte målepunkter/ vurderingshendelser*.

Disse undersøkelsene av forholdet mellom elevenes skriveferdigheter før, gjennom og etter hele intervensjonsperioden og eksterne variabler, peker i retning av at eksterne variabler som målform, morsmål, kjønn og fødselsmåned samlet eller enkeltvis ikke kan forklare effekten av å delta i Normprosjektet. Andre mulige forklaringer på elevenes ferdighetsutvikling gjennom intervensjonen må regnes med, undersøkes og eventuelt avdekkes. Av disse analysene er det særlig interessant at verken elevens kjønn, morsmål, målform eller fødselsmåned på noen entydig måte ser ut til å samvariere systematisk med skriveferdighet.

Når det gjelder forholdet mellom kjønn og skriveferdighet, avviker funnene i Normprosjektet fra etablerte hypoteser i tidligere forskning (Berge et al., 2005). Guttene som har deltatt i Normprosjektet, skriver like godt som jentene i prosjektet. Et unntak fra denne tendensen er skriveferdighetsutviklingen blant 600-elevne, der jentene skriver signifikant bedre enn guttene mot slutten av prosjektperioden.

Funnene i Normprosjektet viser L2-elever etter intervensjonen stort sett mestrer å skrive tekster på samme ferdighetsnivå som elever med norsk som morsmål. Funnet bekrefter resultater fra liknende studier (Bakken & Elstad, 2012; Bakken, 2010).

Et interessant funn i Normprosjektet er at fødselsmåned på trinnet ikke samvarierer systematisk med skriveferdighetsnivået i klassen. Dette funnet peker i retning av at skriveundervisningen på skolen og i klassen utjevner mulige ferdighetsforskjeller elevene imellom som kan skyldes elevenes kognitive og biologiske modning.

4.3 Mulige forklaringer på intervensjonens læringseffekt: videre forskning

Når vi i Normprosjektet skal utvikle en dypere innsikt og mer relevant forståelse av hva som kan forklare de utviklingsmønstrene som de kvantitative analysene i denne rapporten har avdekket, mener vi at tre typer undersøkelser peker seg ut som særlig interessante:

1 Hvordan er Normprosjektet blitt implementert på skolene og i klassene av skoleledelse og de deltakende lærerne?

Følgende underspørsmål kan presisere spørsmål 1:

- Kan skriveferdighetene på prosjektskoler der effekten av intervensjonen fra Normprosjektet har vært *sterk eller særlig sterk*, forklares med hvordan skolene har implementert forventningsnormene og vurderingsperspektivene fra Normprosjektet? I så fall: Hvordan har intervensjonen vært implementert på de skolene der effekten av å delta i Normprosjektet har vært særlig sterk?
- Kan skriveferdighetene på prosjektskoler der det *ikke har vært noen effekt av intervensjonen* fra Normprosjektet, forklares med hvordan skolene har eller ikke har implementert

forventningsnormene og vurderingsperspektivene fra Normprosjektet? I så fall: Hvordan har intervensjonen vært implementert?

2 Hvordan har skriveopplæringen og vurderingsarbeidet på skolene i blitt gjennomført og utviklet gjennom intervensjonen fra Normprosjektet?

Eliminering av de eksterne variablene som mulige forklaringer på elevenes skriveferdighetsnivå før, gjennom og etter prosjektet, peker i retning av at effekten på elevenes skriveferdigheter kan forklares med *kvaliteten på lærernes skriveundervisning*. I følge etablert forskning (Berge & Thygesen, 2015; Graham, MacArthur, & Fitzgerald, 2013; Graham, McKeown, Kiuahara, & Harris, 2012; Hattie & Timperley, 2007; Huot, 2002; Jesson, McNaughton, & Parr, 2011; Matre & Solheim, 2014; Otnes, 2015; Parr & Jesson, 2011; Smidt, 2010; Solheim & Matre, 2014) handler kvaliteten i skriveundervisningen blant annet om:

- Hvordan *oppgavene* elevene skriver, er utviklet og forankret i skolens læringsaktiviteter.
- Hvordan lærerne *modellerer skriveoppgavene* i undervisningssituasjoner i klassen.
- Bruk av *formativ vurdering* gjennom skriveprosessene.
- Elevenes *motivasjon* for skriving.

Mot en slik bakgrunn kan følgende underspørsmål presisere spørsmål 2:

- Kan skriveferdighetene på prosjektskoler der effekten av intervensjonen fra Normprosjektet har vært sterk eller særlig sterk, *forklares med kvalitetene på skriveundervisningen på disse skolene?* I så fall: Hvilke egenskaper ved undervisningen er det som mest sannsynlig kan forklare effekten av skriveundervisning på den enkelte skolen og/eller klassen?
- Kan skriveferdighetene på prosjektskoler der det *ikke* har vært noen effekt av intervensjonen fra Normprosjektet, *forklares med kvalitetene på skriveundervisningen på disse skolene?* I så fall: Hvilke egenskaper ved undervisningen på den enkelte skolen og/eller klassen er det som mest sannsynlig kan forklare de manglende effektene av å delta i Normprosjektet?

3 Hva kjennetegner skriveutviklingen i intensivelevenenes tekster?

- Hvilke språklige og tekstlige kvaliteter preger tekstene til elevene som befinner seg på forventningsnormen på de ulike vurderingsområdene ved ulike målepunkt/vurderingshendelser?
- Hvilke språklige og tekstlige kvaliteter preger tekstene ved ulike målepunkt/vurderingshendelser til elevene som utvikler egne skriveferdigheter i stor eller betydelig grad gjennom å delta i Normprosjektet?

- Hvilke språklige og tekstlige kvaliteter preger tekstene ved ulike målepunkt/vurderingshendelser til elevene som overhodet ikke eller i liten grad utvikler egne skriveferdigheter gjennom å delta i Normprosjektet?
- Hvilke språklige og tekstlige kvaliteter preger tekstene der skriveferdighetene til skriveren varierer med målepunkt/vurderingshendelse?

Litteratur

- Bakken, A. (2010). Prestasjonsforskjeller i Kunnskapsløftets første år - kjønn, minoritetsstatus og foreldres utdanning. NOVA Rapport 9/2010. Oslo: Norsk institutt for forskning om oppvekst, velferd og aldring.
- Bakken, A., & Elstad, J. I. (2012). Sosial ulikhet og eksamensresultater i Oslo-skolen. Trender i perioden 2002–2011. *Tidsskrift for Ungdomsforskning*, 12(2), 67–87.
- Berge, K. L. (1993a). Hvorfor er det så vanskelig å få til en pålitelig bedømming av elevtekster? I P. A. Michelsen (red.), *Sjangeroppbrudd* (s. 65–94). Oslo: Cappelen.
- Berge, K. L. (1993b). På søking etter tolkningsfellesskap. Grunnleggende problemer med å etablere evalueringskriterier for skriveopplæringen i den vg skolen med et sideblikk til tilsvarende problemer i andre land og kulturer. I L. S. Evensen & I. Moslet (red.), *Skrivepedagogisk fornying* (s. 151–190). Oslo: Ad Notam.
- Berge, K. L. (1996). *Norskensensorenes tekstnormer og doxa: en kultursemiotisk og sosiotekstologisk analyse*. (Doktoravhandling, Norges teknisk-naturvitenskapelige universitet).
- Berge, K. L. (2009). Er tolkningsfellesskap mulig å oppnå i skriveprøver? I O. K. Haugaløkken, L. S. Evensen, F. Hertzberg, & H. Otnes (red.), *Tekstvurdering som didaktisk utfordring* (s. 44–54). Oslo: Universitetsforlaget.
- Berge, K. L., Evensen, L. S., Hertzberg, F., & Vagle, W. (red.). (2005). *Ungdommers skrivekompetanse* (Bind I & II). Oslo: Universitetsforlaget.
- Berge, K. L., Evensen, L. S., & Thygesen, R. (under publisering). The Wheel of Writing. A Model of the Writing Domain for the Teaching and Assessing of Writing as a Key Competency. *The Curriculum Journal*.
- Berge, K. L., & Thygesen, R. (2015). Skrivande som grundläggande färdighet: förståelse, bedömning och undervisning. I G. Skar & M. Tengberg (red.), *Bedömning i svenskämnet* (s. 63–86). Stockholm: Natur och kultur.
- Borgström, E., & Ledin, P. (2014). Bedömarvariation. Balansen mellan teknisk och hermeneutisk rationalitet vid bedömning av skrivprov. *Språk och Stil*, 24, 133–165.
- Bouwer, R., Béguin, A., Sanders, T., & van den Bergh, H. (2015). Effect of genre on the generalizability of writing scores. *Language Testing*, 32(1), 83–100.
<http://doi.org/10.1177/0265532214542994>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7. utg.). Milton Park, Abingdon, Oxon: Routledge.
- Eckes, T. (2011). *Introduction to Many-Facet Rasch Measurement: analyzing and evaluating rater-mediated assessments*. Frankfurt am Main: Peter Lang.
- Eriksen, A., & Roe, A. (2011). Den nasjonale prøven i lesing på 8. og 9. trinn, 2011. Rapport basert på populasjonsdata. Institutt for lærerutdanning og skoleforskning, Universitetet i Oslo. Tilgjengelig på:
http://www.udir.no/Upload/Nasjonale_prover/2011/Vedlegg1_Rapport_NPLES_8og9_2011.pdf?epslanguage=no
- Evensen, L. S. (2010). En gyldig vurdering av elevers skrivekompetanse? I J. Smidt (red.), *Rammer for skriving: om skriveutvikling i skole og yrkesliv* (s. 13–31). Trondheim: Tapir akademisk forlag.

- Glass, G. V., & Hopkins, K. D. (1996). *Statistical methods in Education and Psychology* (3. utg.). Boston: Allyn and Bacon.
- Graham, S., MacArthur, C. A., & Fitzgerald, J. (red.). (2013). *Best practices in writing instruction* (2. utg.). New York: The Guilford Press.
- Graham, S., McKeown, D., Kiuahara, S., & Harris, K. R. (2012). A Meta-Analysis of Writing Instruction for Students in the Elementary Grades. *Journal of Educational Psychology*, 104(4), 879–896. <http://doi.org/10.1037/a0029185>
- Gwet, K. L. (2014). *Handbook of inter-rater reliability*. Gaithersburg, MD: Advanced Analytics, LLC.
- Hattie, J. (2009). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81–112. <http://doi.org/10.3102/003465430298487>
- Huot, B. (2002). *(Re)Articulating Writing Assessment for Teaching and Learning*. Logan, Utah: Utah State University Press. Tilgjengelig på: http://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=1136&context=usupress_pubs
- Jesson, R., McNaughton, S., & Parr, J. M. (2011). Drawing on intertextuality in culturally diverse classrooms: Implications for transfer of literacy knowledge. *English Teaching: Practice and Critique*, 10(2), 65–77. Tilgjengelig på: <http://education.waikato.ac.nz/research/files/etpc/files/2011v10n2art5.pdf>
- Johnson, B., & Christensen, L. B. (2000). *Educational Research: Quantitative and Qualitative Approaches*. Boston: Allyn and Bacon.
- Linacre, J. M. (2014). Facets® (version 3.71.4) [Computer Software]. Beaverton, Oregon: Winsteps.com.
- Matre, S., Berge, K. L., Evensen, L. S., Fasting, R. B., Solheim, R., & Thygesen, R. (2011). Developing National Standards for the Teaching and Assessment of Writing. Rapport frå forprosjekt Utdanning 2020. Trondheim: Nasjonalt senter for skriveoppl ring og skriveforskning.
- Matre, S., & Solheim, R. (2014). L rarsamtalar om elevtekstar – mot eit felles fagspr k om skrivning og vurdering. I R. Hvistendahl & A. Roe (red.), *Alle tiders norskdidaktiker* (s. 219–244). Oslo: Novus forlag.
- McNamara, T. F. (1996). *Measuring second language performance*. New York: Longman.
- Ministry of Education. (2012). e-asTTle writing: The technical report. Wellington: Ministry of Education. Tilgjengelig p : [http://e-asTTle.tki.org.nz/content/download/1651/6660/file/e-asTTle Writing Technical Report v.3.pdf](http://e-asTTle.tki.org.nz/content/download/1651/6660/file/e-asTTle%20Writing%20Technical%20Report%20v.3.pdf)
- Otnes, H. (Ed.). (2015). *  invitere elever til skrivning. Ulike perspektiver p  skriveoppgaver*. Bergen: Fagbokforlaget.
- Pallant, J. (2013). *SPSS survival manual*. Maidenhead: McGraw-Hill.
- Parr, J. M., & Jesson, R. (2011). If students are not succeeding as writers, teach them to self-assess using a rubric. I D. Lapp & B. Moss (red.), *Exemplary Instruction in the Middle Grades* (s. 174–188). New York: The Guilford Press.
- PROBA Samfunnsanalyse. (2014). Unders kelse av nynorsk som hovedm l. Nedlastet den 4. juni 2015 fra: <http://www.udir.no/Upload/Forskning/2014/Nynorsk.pdf>
- Purves, A. C. (1992). A Comparative Perspective on the Performance of Students in Written Composition. I A. C. Purves (red.), *The IEA study of written composition* (vol. 2, s. 129–152). Oxford: Pergamon.
- Skar, G., Evensen, L. S., & Iversen, J. M. (2015). L ringsst ttende pr ver i skrivning 2014. Teknisk rapport. Trondheim: Nasjonalt senter for skriveoppl ring og skriveforskning.

- Skolverket/Siris. (2015). Grunnskolan - Resultat nationella prov årskurs 9, per delprov och kön - Engelska, Matematik och Svenska/svenska som andraspråk 2013/14. Hentet 4. juni 2015, fra <http://siris.skolverket.se/>
- Smidt, J. (2010). Skrivekulturer og skrivesituasjoner i bevegelse - fra beskrivelser til utvikling. I J. Smidt (red.), *Skriving i alle fag – innsyn og utspill* (s. 11–35). Trondheim: Tapir akademisk forlag.
- Solheim, R., & Matre, S. (2014). Forventninger om skrivekompetanse. Perspektiver på skrijving, skriveopplæring og vurdering i «Normprosjektet». *Viden om læsning*, (15), 76–88.
- Svennevig, J. (2001). Abduction as a methodological approach to the study of spoken interaction. *Norskrikt*, 103, 1–22.
- Thygesen, R., Evensen, L. S., Berge, K. L., Fasting, R. B., Vagle, W., & Haanæs, I. R. (2007). Nasjonale prøver i skrijving som grunnleggende ferdighet. Sluttrapport. Nasjonalt senter for leseopplæring og leseforskning, Universitetet i Stavanger.
- Utdanningsdirektoratet. (2015). Tall fra Grunnskolens informasjonssystem (GSI) 2014/15. Hentet 4. juni 2015, fra [http://www.udir.no/Upload/Statistikk/GSI/Grunnskole GSI notat 2014-15.pdf?epslanguage=no](http://www.udir.no/Upload/Statistikk/GSI/Grunnskole%20GSI%20notat%202014-15.pdf?epslanguage=no)
- Utdanningsdirektoratet/Skoleporten. (2015). Eksamenskarakterer (grunnskole, 2013-2014). Hentet 4. juni 2015, fra <https://skoleporten.udir.no/rapportvisning.aspx?enhetsid=00&vurderingsomrade=11&skoletype=0&skoletypemenuid=0&underomrade=21&fordeling=4#rapport>