Rapporter 42/2011

*Torbjørn Hægeland, Lars J. Kirkebøen,*
*Bernt Bratsberg og Oddbjørn Raaum*

# Value added-indikatorer

## Et nyttig verktøy i kvalitetsvurdering av skoler?

**Statistisk sentralbyrå • Statistics Norway**
Oslo–Kongsvinger

*Rapporter* I denne serien publiseres analyser og kommenterte statistiske resultater fra ulike undersøkelser. Undersøkelser inkluderer både utvalgsundersøkelser, tellinger og registerbaserte undersøkelser.



ISBN 978-82-537-8249-2 (trykt)
ISBN 978-82-537-8250-8 (elektronisk)
ISSN 0806-2056
Emne: 04.02

Trykk: Statistisk sentralbyrå

| **Standardtegn i tabeller** | **Symbol** |
|---|---|
| Tall kan ikke forekomme | . |
| Oppgave mangler | .. |
| Oppgave mangler foreløpig | … |
| Tall kan ikke offentliggjøres | : |
| Null | - |
| Mindre enn 0,5 av den brukte enheten | 0 |
| Mindre enn 0,05 av den brukte enheten | 0,0 |
| Foreløpig tall | * |
| Brudd i den loddrette serien | — |
| Brudd i den vannrette serien | \| |
| Desimaltegn | , |

# Forord

I de fleste OECD-land har det i de senere år blitt lagt større vekt på resultatkvalitet i skolen, og dokumentasjon av dette, gjennom såkalte kvalitetsvurderingssystemer. Gode indikatorer for skolens bidrag til elevenes resultater (i tillegg til kunnskap om på hvilken måte skolene eventuelt kommer til kort) er fundamentalt for nytten av og tilliten til et kvalitetsvurderingssystem og de vurderinger som gjøres og beslutninger som fattes med dette som en del av faktagrunnlaget. Formålet med denne rapporten er å utrede hvordan value added-indikatorer kan implementeres innenfor Nasjonalt kvalitetsvurderingssystem (NKVS). Prosjektet bygger videre på tidligere arbeid knyttet til skolebidragsindikatorer basert på avgangskarakterer fra ungdomsskolen (Hægeland, Kirkebøen, Raaum og Salvanes, 2005a, 2005b, 2007).

Arbeidet med rapporten er finansiert av Utdanningsdirektoratet.

# Sammendrag

I de fleste OECD-land har det i de senere år blitt lagt mer vekt på resultatkvalitet i skolen og tilhørende dokumentasjon. Verdien av dette avhenger imidlertid kritisk av at de måleinstrumentene man benytter gir pålitelige anslag på forskjeller i kvalitet mellom skoler. Det er etter hvert allment anerkjent at ukorrigerte resultatgjennomsnitt på skolenivå kan være sterkt påvirket av faktorer som er utenfor skolens egen kontroll. Selv om slike resultatmål gir verdifull informasjon om elevenes kunnskapsnivå og prestasjoner, kan de gi et ufullstendig og misvisende bilde av skolekvalitet og hva som er skolens bidrag til resultatene. Mange studier har etter hvert vist at elevsammensetning og tilfeldig variasjon er viktige bidragsytere til resultatforskjeller mellom skoler. Resultatmål som ikke tar hensyn til disse faktorene, er med stor sikkerhet misvisende som indikatorer på skolekvalitet. Spørsmålet er om man, ved hjelp av ulike datakilder, kan konstruere resultatmål som bedre reflekterer skolens bidrag til elevenes læring enn ukorrigerte skoleprestasjoner.

Value added-indikatorer er i prinsippet mer nøyaktige enn andre resultatmål som uttrykk for skolens kvalitet eller bidrag til elevenes læring. Value added-indikatorer skiller seg fra andre indikatorer ved at de også benytter informasjon om elevenes resultater på et tidligere tidspunkt. De korrigerer dermed for viktige forskjeller mellom skoler med hensyn til elevsammensetning som ikke bør fanges opp av indikatorer for skolens bidrag. I norsk sammenheng har det tidligere blitt beregnet lignende indikatorer, men her har man korrigert for forskjeller i elevsammensetning mellom skolens gjennom å kontrollere for elevenes sosioøkonomiske bakgrunn. Ved å ta hensyn til elevenes kunnskaper på et tidligere tidspunkt får indikatoren også en tydeligere tolkning som skolenes bidrag til endring i kunnskaper *i tidsrommet mellom de to målepunktene*.

Denne rapporten drøfter og gjengir beregninger av value added-indikatorer for ulike deler av grunnopplæringen. En viktig forutsetning for arbeidet er at innføringen av nasjonale prøver gir tilgjengelige testresultater for de samme enkeltelevene på ulike trinn på en slik måte at resultatene kan ses i sammenheng. Formålet med rapporten har vært å se nærmere på hvordan value added-indikatorer kan beregnes med de data som er tilgjengelig i Norge i dag, og drøfte hvordan de eventuelt kan implementeres innenfor Nasjonalt kvalitetsvurderingssystem (NKVS).

Resultatene viser gjennomgående at skoler som skårer høyt med hensyn til ujusterte resultater, også tenderer til å skåre høyt når vi ser på value added-indikatorer og skolebidragsindikatorer. Sammenhengen er imidlertid langt fra perfekt. Resultatmål som tar hensyn til at skoler har ulik elevsammensetning gir betydelig tilleggsinformasjon sammenlignet med ujusterte resultater. Indikatorer der elevutfall er justert for kjønn og tidligere resultater synes å være en robust beregningsmåte og har den åpenbare fordel at den kan implementeres av utdanningsmyndighetene selv uten tilleggsinformasjon fra eksterne kilder. Usikkerhet bør rapporteres sammen med indikatorene som grunnlag for å vurdere om resultatforskjeller mellom skoler kan avvises som tilfeldige (statistisk signifikante). I tillegg bør indikatorene beregnes med bakgrunn i data for flere årskull.

Value added- og skolebidragsindikatorer er et hjelpemiddel til å sammenligne resultatene til skoler med forskjellig elevsammensetning, og kan tolkes som det resultatgjennomsnittet vi forventer at en skole ville hatt, om dens elevmasse var gjennomsnittlig i forhold til alle de elevkjennetegn som vi inkluderer i analysen. Indikatorene er et supplement til eksisterende informasjon om skoler og skolekvalitet. De kan ikke *erstatte* eksisterende informasjon, men kan bidra til å gi et mer utfyllende bilde av virksomheten som foregår på skolene. Erkjennelsen av at skolekvalitet ikke kan oppsummeres i et enkelt tall, gjør at value added-indikatorer bør presenteres sammen med annen relevant informasjon om skoler, slik at det er mulig å danne seg et mer helhetlig bilde av virksomheten ved den enkelte skole. Ukorrigerte resultatmål og value added-indikatorer vil uansett bare være verktøy for å *identifisere* god praksis i skolen, dvs. finne de skoler som bidrar mye til elevenes læring. For å *karakterisere* god praksis, dvs. finne hva som kjennetegner skoler med høyt bidrag eller enda mer ambisiøst *hvorfor* noen skoler bidrar mer enn andre, kreves andre data og andre analyseverktøy, men value added-indikatorene vil være et viktig grunnlag for slike mer overordnede analyser

# Abstract

In recent years, most OECD countries have added emphasis on assessment and documentation of performance and quality of schools. The value of such focus depends critically, however, on measures that provide reliable estimates of quality differences between schools. It is widely recognized that unadjusted results averaged at the school level can be strongly influenced by factors that are outside the school's own control. Although unadjusted measures provide valuable information about students' knowledge and performance, they will typically give an incomplete and misleading picture of school quality and the school's contribution to student outcomes. A broad research literature shows that the composition of pupils and random variation affect performance differences across schools. Targets that fail to take such factors into account are likely to be biased indicators of school quality. The question is whether, based on different data sources, one can construct measures that better reflect the school's contribution to student learning than that captured by the unadjusted school average.

In principle, value-added indicators are more accurate than other existing measures in terms of expressing the school's quality and contribution to pupil learning. Value-added indicators differ from alternative indicators in that they use information about student performance at an earlier stage of the education process. The indicator thus controls for differences between schools in terms of student composition, which should not influence assessment measures of the school's contribution. Prior Norwegian studies have estimated similar indicators, but with adjustments for differences in pupil composition between schools based on students' socioeconomic background. When taking into account students' outcomes at an earlier date, the value-added indicator provides a more direct measure of the school's contribution to the change in the stock of knowledge between the two measurement dates.

This report discusses the methodology and presents estimates of value-added indicators for different stages of primary and secondary education. An important prerequisite for the study is the introduction of comprehensive testing that permits analysis of the evolvement of students' test scores over time. The purpose of the report is to examine how value-added indicators can be computed with the data available in Norway as of today, and to discuss how value-added indicators can be implemented within the national quality assessment system ("Nasjonalt kvalitetsvurderingssystem", NKVS).

Our results consistently show that high-performing schools in terms of unadjusted results also tend to score highly when we look at value-added indicators of the school's contribution. This relationship is, however, far from complete. Performance indicators that consider schools differences in student composition provide important additional information when compared to unadjusted results. Indicators that adjust student outcomes for gender and previous test results appear to yield robust estimates of school contributions, and the method has the obvious advantage that it can be implemented by education authorities without collecting additional information from external sources. Measures of uncertainty should be reported along with indicator values as a basis for assessing whether performance differences between schools can be dismissed as random variation (statistically significant). Furthermore, indicator values for schools should be based on data covering multiple cohorts.

Value-added and other indicators of schools' contributions are useful tools in order to compare the performance of schools that differ in student composition, and can be interpreted as the expected outcome of a school had its student body been average along all student characteristics included in the analysis. The indicator values provide supplements to other information about schools and school quality. They cannot replace existing information but can help provide a more complete picture of the activities that take place in schools. Recognizing that school quality cannot be summarized in a single number, value-added indicators should be presented together with other relevant information about schools, so that it is possible to form a more complete picture of the activities at each school. Unadjusted performance measures and value-added indicators will still only be tools to identify good practice in schools, i.e., to identify schools that excel in their contribution to pupil learning. To describe good practice and characteristics of schools with high contributions, or even more ambitiously, to explain why some schools contribute more than others, requires other data and other analytical approaches. Value-added indicators may, however, be an important input in broader analyses of school performance and quality.

# Innhold

# 1. Innledning og bakgrunn

I de fleste OECD-land har det i de senere år blitt lagt mer vekt på resultatkvalitet i skolen, og dokumentasjon av dette. Verdien av et slikt økt fokus avhenger imidlertid kritisk av at de måleinstrumentene man faktisk benytter gir pålitelig informasjon om variasjon i skolekvalitet. Det er etter hvert vel kjent at rene resultatgjennomsnitt på skolenivå kan være et svært misvisende mål skolenes bidrag til elevenes læring, fordi det er influert av andre faktorer som i stor grad er utenfor skolens kontroll, som for eksempel sammensetningen av elevmassen. Et skolebidrag er knyttet til en spesiell periode og skal representere effekten skolemiljøet har hatt på endringen i elevenes kunnskaper i den aktuelle tidsperioden. Gode indikatorer for skolens bidrag til elevenes resultater (i tillegg til kunnskap om på hvilken måte skolene eventuelt kommer til kort) er fundamentalt for nytten av og tilliten til et kvalitetsvurderingssystem og de vurderinger som gjøres og beslutninger som fattes med dette som en del av faktagrunnlaget.

Med utgangspunkt i en interesse for å kunne vite noe om forskjeller i skolenes bidrag til elevenes læring, og en erkjennelse av at ukorrigerte resultatgjennomsnitt på skolenivå kan gi et ufullstendig og misvisende bilde av dette, er skolebidragsindikatorer et potensielt viktig hjelpemiddel i arbeidet med å identifisere god praksis i skolen. Det er allerede nå viktig å understreke at skolebidragsindikatorer aldri kan bli det eneste verktøyet i dette arbeidet. Selv om ukorrigerte resultatmål som for eksempel skolens gjennomsnittsresultat ved skriftlig eksamen, eller andelen elever under et visst nivå på nasjonale prøver ikke nødvendigvis reflekterer skolens bidrag til elevenes læring på en god måte, har slike mål selvsagt betydelig informasjonsverdi. Uavhengig av hvor godt skolens bidrag er, er det bekymringsfullt dersom mange elever ved en skole har resultater som vitner om et kunnskapsnivå som er for lavt i forhold til å skulle klare seg i videre utdanning og i arbeidslivet. Ukorrigerte resultatmål og skolebidragsindikatorer vil uansett bare være verktøy for å *identifisere* god praksis i skolen, dvs. finne de skoler som bidrar mye til elevenes læring. For å *karakterisere* god praksis, dvs. finne hva som kjennetegner skoler med høyt bidrag eller enda mer ambisiøst *hvorfor* noen skoler bidrar mer enn andre, kreves andre data og andre analyseverktøy. Denne erkjennelsen av at skolekvalitet ikke kan oppsummeres i et enkelt tall, gjør at skolebidragsindikatorer bør presenteres sammen med annen relevant informasjon om skoler.

Value added-indikatorer er i prinsippet mer nøyaktige enn andre resultatmål som uttrykk for skolens kvalitet eller bidrag til elevenes læring. Value added-indikatorer skiller seg fra andre skolebidragsindikatorer ved at de også benytter informasjon om elevenes resultater på et tidligere tidspunkt. De korrigerer dermed for viktige forskjeller mellom skoler med hensyn til elevsammensetning som ikke bør fanges opp av indikatorer for skolens bidrag. I norsk sammenheng har det tidligere blitt beregnet lignende skolebidragsindikatorer, men her har man korrigert for forskjeller i elevsammensetning mellom skoler gjennom å kontrollere for elevenes sosioøkonomiske bakgrunn. Ved å ta hensyn til elevenes kunnskaper på et tidligere tidspunkt får indikatoren også en tydeligere tolkning som skolenes bidrag til endring i kunnskaper *i tidsrommet mellom de to målepunktene*.

I denne rapporten bruker vi begrepet *skolebidragsindikatorer* som en fellesbetegnelse på indikatorer som søker å gi et uttrykk for skolenes bidrag til elevenes læring gjennom å kontrollere for forskjeller i elevsammensetning på tvers av skoler og på den måten korrigere ujusterte resultatforskjeller mellom skoler. Vi vil skille mellom to typer skolebidragsindikatorer: Indikatorer som kontrollerer for elevene tidligere resultater betegnes value added-indikator*er,* mens de som bare baserer seg på informasjon om elevenes sosioøkonomiske bakgrunn betegnes *tverrsnittsindikatorer.*

Formålet med denne rapporten er å utrede hvorvidt og hvordan value added-indikatorer kan beregnes med de data som er tilgjengelig for norske elever og

skoler. Prosjektet bygger videre på tidligere arbeid knyttet til skolebidrags-indikatorer basert på avgangskarakterer fra ungdomsskolen (Hægeland, Kirkebøen, Raaum og Salvanes, 2005a, 2005b, 2007). I tillegg til dette arbeidet, ble det i 2006 publisert en rapport (Hægeland, Kirkebøen, Raaum og Salvanes, 2006), som så nærmere på mulighetene for benytte tilsvarende type rammeverk for å se på resultatforskjeller mellom videregående skoler. Studien pekte på ytterligere begrensninger, problemer og muligheter man står overfor når man studerer videregående skole, sammenlignet med grunnskolen. Senere ble dette arbeidet fulgt opp med et prosjekt som omfattet videregående skoler i Oslo kommune (Hægeland, Kirkebøen og Raaum, 2010). I tillegg vil vi trekke veksler på internasjonale erfaringer, blant annet er dokumentert i OECD (2008).

Tidligere har det ikke vært mulig å beregne value added-indikatorer for norske skoler fordi testresultater på ulike alderstrinn ikke har vært tilgjengelig for enkelt-elever på en slik måte at resultatene kan ses i sammenheng. Datasituasjonen på dette feltet har imidlertid blitt vesentlig forbedret. Nasjonale prøver på 5. og 8. trinn ble innført på årlig basis fra høsten 2007. Dette innebærer at det nå er mulig å beregne value added-indikatorer for:

1. Mellomtrinnet: Basert på nasjonale prøver for 8. og 5. trinn
2. Ungdomstrinnet: Basert på avgangskarakterer for 10. trinn og nasjonale prøver for 8. trinn
3. Videregående skole: Basert på karakterer/fullføring/frafall fra VG1/VG2/VG3 og avgangskarakterer for 10. trinn.
4. I tillegg er det mulig å beregne tverrsnittsindikatorer for barnetrinnet, basert på nasjonale prøver for 5. trinn

Rapporten er disponert på følgende måte. I kapittel 2 gis det en kort prinsipiell drøfting av hva som kan skape resultatforskjeller mellom skoler. Deretter defineres value added-indikatorer i kapittel 3. I kapittel 4 beskriver og drøfter vi metoden for å estimere value added-indikatorer. Kapittel 5 beskriver datagrunnlaget. I kapitlene 6-9 presenteres beregninger av indikatorer for ulike resultatmål og trinn. Det avsluttende kapitlet gir en sammenfattende drøfting.

# 2. Hva skaper resultatforskjeller mellom skoler?

De fleste studier av variasjoner i skoleprestasjoner, vår egen inkludert, bygger mer eller mindre eksplisitt på en teoretisk modell der en elevs skoleprestasjoner avhenger av elevens forutsetninger og miljø, skolens bidrag til læring og tilfeldig variasjon. Skoleprestasjoner kan i denne sammenheng være mål på kunnskapsnivå på et gitt tidspunkt, eller endringer i kunnskapsnivå over et tidsrom. Fra dette tankeskjemaet følger det at gjennomsnittsresultat på skolenivå grovt sett kan tilskrives tre hovedfaktorer:

1. Skolens bidrag til læring, inkludert bidraget på tidligere klassetrinn
2. Elevens kunnskapsnivå fra tidligere og forutsetninger for å tilegne seg ny kunnskap
3. Tilfeldig variasjon og målefeil

De to siste faktorene ligger utenfor skolens kontroll. Den enkelte skole kan styrke sitt eget bidrag, selv om handlingsrommet for rektorer og lærere begrenses av rammer og ressurser som skoleeiere og sentrale myndigheter fastsetter. Skolene som skårer høyest målt i rene elevresultater gir ikke nødvendigvis det største bidraget til læring. Det kan skyldes fordelaktig elevsammensetning eller tilfeldigheter. På samme måte er det langt fra opplagt at skoler med svake resultater gir elevene et dårlig læringsutbytte.

### Skolens bidrag til læring

Skolens bidrag til læringsutbyttet kan tilskrives flere forhold. Hvorvidt resultatforskjeller mellom skoler faktisk reflekterer forskjeller i skolens bidrag, avhenger kritisk av hvor viktig elevenes forutsetninger – og tilfeldig variasjon – er for karakterer og testresultater. Dette er et empirisk spørsmål, og kan bare fastslås ved nøyaktige undersøkelser basert på faktiske resultater. Formålet med å korrigere skolens resultater for faktorer utenfor dens kontroll er nettopp å komme nærmere en kvantifisering av forskjeller mellom skoler i deres bidrag til læring. Slike korrigerte resultatforskjeller kan *ikke* kaste lys over *hvilke* "skolefaktorer" (f.eks. forskjeller i ressursbruk, lærerkompetanse) som eventuelt betyr mest for forskjellene. Identifikasjon av slike faktorer, og effekter av politiske virkemidler spesielt, er en svært krevende oppgave, og utfordringen henger blant annet sammen med at ressursbruk i skolen ikke er uavhengig av andre faktorer, både observerte og uobserverte, som påvirker elevresultater. Hanushek (2003), Krueger (2003) og Todd og Wolpin (2003) gir til sammen en bred oversikt over denne tematikken. Hægeland, Kirkebøen, Raaum og Salvanes (2005c) gir en ikke-teknisk diskusjon av hvilke problemer man møter i slike studier.

### Elevenes forutsetninger og bakgrunn

Utallige undersøkelser fra ulike land og tidsperioder viser at utdanningsutfall henger nært sammen med sosioøkonomiske kjennetegn ved familien en vokser opp i. Når det gjelder grunnskoleresultater i Norge viser for eksempel Hægeland, Kirkebøen, Raaum og Salvanes (2004) at familiebakgrunn, målt ved et svært rikt sett av registerbaserte variabler som reflekterer foreldrenes utdanning, inntekt, formue, arbeidsmarkedstilknytning, trygdeforhold, sivilstand, familiestørrelse osv., kan forklare omtrent 30 prosent av variasjonen i karakterer mellom enkeltelever. Det er også dokumentert klare sammenhenger mellom elevers resultater på ulike trinn i utdanningen, se for eksempel Hægeland, Kirkebøen og Raaum (2006) for en kartlegging av sammenhengen mellom resultater fra grunnskolen og videregående skole for norske elever. Som vi også viser senere i denne rapporten, kan tidligere resultater gjennomgående forklare en større andel av resultatvariasjonen mellom elever enn hva et rikt sett med familiebakgrunnsvariable er i stand til.

Den positive samvariasjonen mellom enkeltelevers resultater på ulike trinn i utdanningen er ikke overraskende. I den grad det kreves samme *type* ferdigheter på forskjellige nivåer, og de elevkjennetegnene som påvirker ferdighetene (bortsett fra

alder) er noenlunde konstante over tid, vil vi forvente at en elev som gjør det godt på et nivå også vil gjøre det godt på et høyere nivå. Vi vil ikke kunne si noe om hvorvidt samvariasjonen mellom resultater på ulike nivåer en slik sammenheng skyldes elevens (medfødte) evner, motivasjon eller oppfølging fra foreldrene, eller at kunnskap ervervet på et tidligere tidspunkt er en viktig innsatsfaktor i innlæring av ny kunnskap. Når man i empiriske analyser av skoleresultater bruker tidligere resultater som kontrollvariabel (som er det man gjør ved beregning av value added-indikatorer), kan tidligere resultater tolkes både som et direkte mål på ervervet kunnskap og som et signal om den totale effekten av uobserverte faktorer som evner, motivasjon m.v.

Selv om empiriske studier viser sterk samvariasjon mellom familiebakgrunn og skoleprestasjoner og mellom elevers skoleprestasjoner på ulike nivåer, er det viktig å presisere at det ikke dreier seg om et en-til-en-forhold. Familiebakgrunnen er en svært viktig faktor for å forklare skoleprestasjoner, men det er samtidig et stort rom for andre faktorer. Tross alt kan en stor del av variasjonen i karakterer tilskrives andre forhold enn hva vi kan kartlegge om familiene. Selv om barn av foreldre med høy utdanning og god økonomi gjennomsnittlig oppnår bedre resultater enn klassekamerater som har foreldre med kort skolegang og lav inntekt, finnes det mange *enkelttilfeller* der forholdet er motsatt. Det vil heller ikke være en lovmessig sammenheng mellom tidligere og framtidige resultater. Det finnes åpenbart umotiverte elever med svake resultater på ungdomskolen, som blir mer motiverte, jobber hardere og dermed får bedre resultater på videregående.

Elevers skoleprestasjoner varierer altså systematisk med ulike kjennetegn, men hvordan påvirkes resultatene på skolenivå? Dersom det ikke var systematiske forskjeller mellom skoler med hensyn til elevsammensetning, ville ikke gjennomsnittsresultater på skolenivå påvirkes av for eksempel at barn av foreldre med høyere utdanning i gjennomsnitt gjør det bedre på skolen enn barn av lavt utdannede foreldre. Slik er det imidlertid ikke, elever med ulik bakgrunn fordeler seg ikke jevnt utover skolene. Det er tvert imot en klar tendens til at de som har relativt lik bakgrunn "klumper seg sammen" på samme skole. Dette kan skje delvis ved at like familier i stor grad velger tilsvarende boligområder, kombinert med at barn og ungdom typisk går på skoler nær hjemmet. For videregående skjer det også gjennom elevens aktive valg av studieretning og skole de søker på, og i den grad opptakssystemet sorterer elevene etter karakterer – direkte eller indirekte – vil også dette bidra til at elever med tilsvarende resultater fra grunnskolen går på samme skole. Dermed vil mye av karakterforskjellene mellom skoler være påvirket av elevsammensetningen.

### Tilfeldig variasjon

Resultatforskjeller mellom skoler kan også skyldes tilfeldig variasjon. Et skolegjennomsnitt er beheftet med statistisk usikkerhet, som skyldes både tilfeldigheter bak enkeltelevers prestasjoner og særskilte og "uvanlige" hendelser på skolen eller klassetrinnet. Det kan synes merkelig å snakke om usikkerhet knyttet til en indikator som i prinsippet omfatter alle elevene på et klassetrinn ved en skole. Gjennomsnittskarakteren ved eksamen i norsk for skole A i 2010 er jo et eksakt mål på gjennomsnittskarakteren ved eksamen i norsk *for skole A i 2010*. Så lenge vi aksepterer at karakterer faktisk måler det vi er interessert i, er det kun innslag av tekniske registreringsfeil som skaper usikkerhet.

Når det likevel er viktig å fokusere på usikkerhet eller tilfeldig variasjon, skyldes det at vår interesse strekker seg ut over hva elevene på skole A og B oppnådde ett bestemt år. Vi ønsker en pekepinn på læringsutbyttet som elevene får på den enkelte skole, det vil si et mer permanent kjennetegn ved skolen. En gjennomsnittskarakter for ett enkelt år er bare ett enkelt resultat, for et bestemt elevkull. Med dette perspektivet blir det nødvendig å ta hensyn til tilfeldig variasjon og statistisk usikkerhet.

En viktig kilde til usikkerhet er knyttet til antall elever ved skolen. Jo færre elever som danner grunnlaget for å regne ut et gjennomsnittsresultat, jo større vil variasjonen i resultatet typisk være. Når karakteren for hver elev i noen grad styres av tilfeldigheter (god eller dårlig dag), reduseres usikkerheten for gjennomsnittet jo flere elever det representerer. Det er også andre enkeltstående faktorer, eller skolespesifikke hendelser, som gjør at skolenes resultater svinger fra år til år. Slike tilfeldigheter som alle elever kan bli eksponert for, kan være av betydning for gjennomsnittsresultatene. Det viktig å ta hensyn til også denne formen for usikkerhet når man sammenlikner resultater mellom skoler.

Til slutt er det et spørsmål hvor godt prøveresultater eller karakterer fanger opp det som egentlig interesserer oss, elevenes kunnskap eller ferdigheter. En forutsetning for at det skal være meningsfullt å beregne kvalitetsindikatorer er at vi har et mål på elevenes prestasjoner. Elevenes prestasjoner vil avhenge av deres ferdigheter, men trenger imidlertid ikke å fullt gjenspeile disse. Særlig ved enkeltprøver og eksamener er det vanskelig å teste alle relevante ferdigheter, slik at resultatet bare måler deler av disse. Standpunktkarakterer baserer seg på grundigere observasjon, og fanger sannsynligvis opp et bredere spekter av ferdigheter, men til gjengjeld vil ikke forskjellige elevers prestasjoner nødvendigvis vurderes ut fra samme skala (se Galloway, Kirkebøen og Rønning, 2011).

# 3. Hva er value added?

OECD (2008) gir følgende definisjon av *the value added contribution of a school* (skolens bidrag til elevens læring):

> *the contribution of a school to students' progress towards stated or prescribed education objectives (e.g. cognitive achievement). The contribution is net of other factors that contribute to students' educational progress.*

Ut fra denne definisjonen gis så følgende definisjon av value added-modeller:

> *a class of statistical models that estimate the contributions of schools to student progress in stated or prescribed education objectives (e.g. cognitive achievement) measured at at least two points in time.*

Det er viktig å merke seg presiseringen av at value added-modeller omfatter de som benytter seg av resultatmål fra minst to ulike tidspunkter. Dette innebærer at skolebidragsindikatorer basert på avgangskarakterer fra grunnskolen, slik de så langt har vært beregnet og publisert i Norge (se for eksempel Hægeland et al., 2005a), ikke faller inn under denne definisjonen. Modeller som estimerer skolens bidrag til elevers læring ved hjelp av tverrsnittsdata for elevprestasjoner, skoletilhørighet og informasjon om elevenes sosioøkonomiske bakgrunn, kalles i OECD-rapporten for *contextualized attainment models*, og vi velger her å kalle dem tverrsnittsindikatorer. Slike modeller har mange av de samme egenskapene og bruksområdene som value added-modeller. Rent teknisk/statistisk er de i prinsippet tilsvarende, siden elevens tidligere resultater kan ses på som et elevkjennetegn på lik linje med andre familiebakgrunnsvariable. Tidligere elevprestasjoner kan imidlertid bidra til å fange opp uobserverbare faktorer (for eksempel motivasjon og evner) som ikke nødvendigvis reflekteres fullt ut i mål på sosioøkonomisk bakgrunn (Raudenbush, 2004). Når vi for eksempel kontrollerer for foreldrenes utdanning i estimering av tverrsnittindikatorer, kontrollerer vi for at barn av høyt utdannede foreldre *i gjennomsnitt* gjør det bedre enn barn av lavt utdannede foreldre, og vi tilordner denne gjennomsnittsforskjellen til alle elevene. Tidligere resultater fanger dette opp på individnivå.

Det som imidlertid skiller value added-modeller fra tverrsnittsmodeller er at de estimerte skoleeffektene som value added-modellene produserer kan gis en mye mer presis tolkning som skolens bidrag til elevenes læring i den perioden som ligger *mellom* de ulike målene på elevprestasjoner, siden man betinger på kunnskaps- og ferdighetsnivået ved inngangen til perioden. Dette gjør indikatorene mer hensiktsmessige til bruk for eksempel skoleutvikling. Dette er ikke tilfelle med den typen av skolebidragsindikatorer som hittil har vært beregnet for Norge, hvor det er mindre klart hva man faktisk betinger på når man kontrollerer for familie-bakgrunn og der resultatforskjeller vil avspeile mulige kvalitetsforskjeller mellom skoler på alle tidligere trinn av opplæringen.

Det bør nevnes at valg av type modell ofte styres av hva slags data som er tilgjengelig. For Norges del har det hittil stort sett bare vært mulig å beregne indikatorer på bakgrunn av tverrsnittsinformasjon, og mangelen på informasjon om elevprestasjoner over tid har i stor grad blitt kompensert med å benytte et rikt sett av variable for elevenes sosioøkonomiske bakgrunn.

# 4. Formelt rammeverk

I dette kapitlet drøfter vi tolkningen av value added-indikatorer innenfor at analytisk rammeverk for kunnskapstilegnelse. Rammeverket bygger på Todd og Wolpin (2003) og Rothstein (2010), men er tilpasset vårt fokus på å identifisere skolenes bidrag til elevenes resultater (skoleeffekter). Formålet er å få klart fram hvilke antakelser ulike skolebidragsindikatorer bygger på.[1] Rammeverket beskriver en rekke ulike typer effekter som det i praksis ikke er mulig å identifisere empirisk. Likevel er de nyttige for drøftingen av mulig kilder til skjevhet der skolebidrags-indikatorene fanger opp andre forhold enn skolenes rolle i elevenes kunnskaps-tilegnelse.

Formelt ser vi på utfallet av testen i klassetrinn $g$ ($A_g$) for elev $i$ som et resultat av skolehistorien til eleven selv (dvs. alle skolebidragene fra tidligere klassetrinn), personlige egenskaper, familie- og miljøforhold under oppveksten og tilfeldigheter;

$$(1) \qquad A_{ig} = \alpha_g + \sum_{h=1}^{g} \beta_{hgs(i,h)} + \mu_i \tau_g + \sum_{h=1}^{g} \varepsilon_{ih} \phi_{hg} + \nu_{ig}$$

La oss se litt nærmere på de ulike elementene i (1). Skoleeffektene representert ved beta-koeffisientene er konstruert slik at skolene man har gått på gjennom hele skolekarrieren er potensielt viktig for senere resultater. Skole s på klassetrinn h, dvs. *s(i,h)* har effekt på testresultatet påfølgende trinn (g minst lik stor som h). Et sentralt spørsmål i alle studier av kunnskapsakkumulasjon er varigheten av eventuelle effekter fra tidligere miljøer. For vår problemstilling er spørsmålet: Hvor lenge vedvarer skoleeffekten fra klassetrinn *h*? På dette punktet gir teorien et tynt grunnlag for hvordan dette skal modelleres. Vi vil konsentrere oss alternativer innenfor en ramme av såkalt uniform geometrisk "forvitring" (decay) der

$$(2) \qquad \beta_{hg's} = \beta_{hgs} \lambda^{g-g'}, 0 \le \lambda \le 1 \qquad h \le g \le g'$$

I ligning (2) avtar skoleeffekten fra et bestemt klassetrinn med en fast rate pr tidsenhet. Forskjellen mellom effekten av *skole s (for en enkelt elev) i trinn h* på to ulike senere trinn vil da avhenge av avstand mellom trinnene. Innenfor dette rammeverket er det to ekstremtilfeller: Hvis λ=0, forsvinner skoleeffektene umiddelbart, slik at skolen eleven går på bare har effekt på testresultatet på samme trinn, og ikke senere. Motsatsen er en permanent effekt (λ=1), der alt skolen gir eleven av kunnskaper på et gitt klassetrinn varer hele skoleløpet. Det er god grunn til å tro at virkeligheten ligger et sted mellom de to ytterpunktene.

Det neste leddet i (1) *($\mu_i \tau_g$)* fanger opp permanente personlige egenskaper ($\mu_i$) og disse tillates i vårt teoretiske rammeverk å ha ulike effekter ($\tau_g$) på forskjellige klassetrinn.

Totaleffekten av alle de andre systematiske faktorene ($\varepsilon_{ih}$) , herunder familie- og nærmiljø, på trinn g er gitt ved

$$\omega_{ig} = \sum_{h=1}^{g} \varepsilon_{ih} \phi_{hg}$$

og tilfeldig variasjon inkludert målefeil er i (1) representert ved $v_{ig}$.

Når formålet er å tallfeste skoleeffekter og konkret avklare eventuelle forskjeller mellom skoler i deres bidrag til elevenes resultater, er det sentrale spørsmålet

[1] Det finnes en rekke alternative rammeverk for kunnskapsakkumulasjon som kunne vært benyttet og illustrert antakelsene bak value added-modeller på en tilsvarende måte.

*hvilke forutsetninger* som må være oppfylt for at våre anslag på skoleeffekter faktisk skal representere bidrag fra skolen - og ikke andre faktorer med innflytelse på testresultater – slike som samtidig varierer systematisk på tvers av skoler. Med andre ord: Under hvilke forutsetninger gir "value added" pålitelige resultater for skolekvalitet? Kan vi stole på at disse forutsetningene er oppfylt?

## 4.1. Eksempel - karakterer på 10. trinn

For å konkretisere drøftingen ser vi på et eksempel hvor vi skal identifisere skoleeffektene med utgangspunkt i avgangsresultater fra ungdomsskolen. Vi tenker oss for enkelhets skyld at ferdighetsnivået ved utgangen av henholdsvis 10. trinn og 7. trinn måles ved et kontinuerlig utfallsmål på den samme skalaen. Ut fra likning (1) kan vi legge og trekke fra $A_{i7}$ (merk at denne er generert av den samme modellen (1)) som $A_{i10}$ multiplisert med en faktor slik at

$$(3) \qquad A_{i10} = (\alpha_{10} - \lambda^* \alpha_7) + \lambda^* A_{i7} + \sum_{h=8}^{10} \beta_{h10s(i,h)} + \theta_{i10}$$

Skolebidraget vi er på jakt etter er summen av *effekten fra skole s i 8., 9. og 10. trinn på ferdighetsnivået ved utgangen av 10.trinn.* I ligning (3) uttrykkes dette ved $\sum_{h=8}^{10} \beta_{h10s(i,h)}$ . Anta nå at vi kjenner hvilken betydning kunnskapsnivået på 7. trinn har for kunnskapene tre år senere ($\lambda^*$). Så lenge vi observerer $A_{i7}$ er det kritisk for konsistente anslag på skolebidragene at den uobserverte variabelen $\theta_{i10}$ er tilfeldig fordelt mellom skoler. Det må altså være slik at når vi kontrollerer for tidligere resultater, er det ingen systematisk sortering på tvers av skoler når det gjelder uobserverte elevkjennetegn som påvirker resultatene på 10. trinn.

Den uobserverte variabelen $\theta_{i10}$ består av ulike komponenter:

$$(4) \qquad \theta_{i10} = (\sum_{h=1}^{7} \beta_{h10s(i,h)} - \lambda^* \sum_{h=1}^{7} \beta_{h7s(i,h)}) + \mu_i(\tau_{10} - \lambda^* \tau_7) + (\omega_{i10} - \lambda^* \omega_{i7}) + (\nu_{i10} - \lambda^* \nu_{i7})$$

La oss drøfte dem etter tur:

*(i) Skolebidragene fra tidligere (barne- og mellomtrinnet)*

Under uniform forvitring av kunnskap er hvert av leddene i summen $\sum_{h=1}^{7} (\beta_{h10s(i,h)} - \lambda^* \beta_{h7s(i,h)})$ lik null ettersom $\beta_{h10s(i,h)} = \lambda^3 \beta_{h7s(i,h)}$. All framtidig effekt av skolebidragene fra barne- og mellomtrinnet er med andre ord fanget opp av ferdighetsnivået målt ved utgangen av mellomtrinnet. Det er lettest å tenke på dette i grensetilfellet når λ går mot 1 ettersom skoleeffekter da varer evig. Hva som hendte på for eksempel 5. trinn har da den samme effekten på utfall både på 7. og 10. trinn.

*(ii) Individuelle evner (μ)*

For individuelle evner gjelder ikke argumentet om forvitring. Dessuten kan effekten av individuelle evner for den enkelte elev kan være ulik på 10. og 7. trinn. Elever utvikler seg ulikt med hensyn til modenhet, evne til konsentrasjon og beherske prøvesituasjoner. Samtidig stiller skolen varierende krav til disse evnene over tid og variasjon i evner kan dermed få ulike konsekvenser. I så fall blir ikke konsekvenser av variasjon i individuelle evner fullt ut fanget opp av i resultatene

på 7. trinn, og det andre leddet i (4) vil generelt være forskjellig fra null for den enkelte elev.

*(iii) miljøfaktorer (ω)*

Miljøfaktorer kan endre seg over tid (eksempelvis familiehendelser) samtidig som et gitt miljø kan ha ulik effekt over tid. Også her vil forvitringsmønsteret ha betydning for eventuelle skjevheter. Anta som for skoleeffekter at "framtidseffekten" av miljøet på klassetrinn *h* er mindre jo lengre inn i framtida vi ser:

$$\phi_{hg'} = \varphi_{hg}\xi^{g-g'}, 0 \le \xi \le 1 \qquad h \le g \le g'$$

Dersom $\varepsilon_{ih} = \varepsilon_i$ (stabilt miljø for hver enkelt elev over tid) kan miljøeffekten i likning (4) skrives som

$$(5) \qquad (\omega_{i10} - \lambda^* \omega_{i7}) = \varepsilon_i \sum_{h=8}^{10} \phi_{h10}$$

Intuisjonen bak (5) er den samme som for skoleeffektene omtalt over. Når vi betinger på resultatet i slutten av 7. trinn fanger vi samtidig opp effekten av historien (miljøet i 1. til 7. klasse) dersom vi har geometrisk avtakende framtidseffekt (eller full persistens; λ =λ*=1). Det som gjenstår er ekstrabidragene fra miljøfaktorene i ungdomsskoleårene, jf. (5). En viktig komponent i det vi kaller miljøfaktorer er familien. Det at miljøfaktorene kan ha spesifikke effekter på 8. - 10. trinn kan være viktig er et teoretisk argument for å kontrollere for familiebakgrunnskjennetegn i estimeringen av skoleeffekter, selv om vi korrigerer for ferdighetsnivå ved inngangen til perioden. Hvor mye dette faktisk har å si for de estimerte skoleeffektene, er et empirisk spørsmål. Dermed gjenstår problemet med uobserverte miljøfaktorer som ikke er reflektert i resultatet fra 7. trinn *eller* fanges opp av observert familiebakgrunn.

*(iv) Tilfeldig variasjon (v)*

Med tilfeldig variasjon mener vi hendelser av individ- eller skolespesifikk art som påvirker resultatene. For enkeltelever kan det være at man hadde en god eller dårlig dag da prøven ble holdt. Det i tillegg mange forhold av tilfeldig karakter som påvirker større grupper av elever. Emnet for eksamen kan slå heldig eller uheldig ut for enkeltskoler, avhengig av hva som har vært vektlagt i undervisningen. Dersom skolen er rammet av en influensaepidemi rundt tiden for prøven, kan det slå uheldig ut, med mange halvsyke elever ved eksamenspulten. Støyende byggearbeider i nærheten kan virke forstyrrende og påvirke resultatene. Andre former for tilfeldig variasjon kan påvirke selve læringen gjennom skoleåret. En lærer kan ha spesielt god kjemi med en klasse, slik at forholdene for læring blir uvanlig gode. Langvarig sykefravær hos én eller flere lærere og varierende stabilitet og kvalitet på vikarlærerne kan hemme tilegnelsen av nye kunnskaper. Én eller flere problemelever kan virke forstyrrende på undervisningen og ødelegge læringen for hele klassen. Fra indikatorperspektivet er tilfeldigheter på elevnivå uproblematisk når de ikke samvarierer mellom elever på samme skole. Tilfeldigheter på skolenivå vil påvirke anslag på skolebidraget og er en viktig grunn til at de bør basere seg på observasjoner av flere elevkull.

## 4.2. Hva slags antakelser ligger bak ulike spesifikasjoner?

I lys av dette rammeverket, hvilke antakelser ligger bak en standard value added-modell (VAM)? Når gir VAM korrekte svar på hva de enkelte skoler bidrar med i elevenes kunnskapstilegnelse? Modellen basert på endringen i resultater fra ett tidspunkt til et annet innbærer en antakelse om fullstendig persistens: λ=1. Alt

skolen og miljøet for øvrig gir eleven av kunnskaper på et gitt klassetrinn antas å vare hele skoleløpet ut. I dette tilfellet får vi riktig svar, dvs. at anslag på skoleeffektene er konsistente med gjennomsnittet av $\theta_{i10}$ i forventning det samme for alle skoler dersom følgende er oppfylt: (i) effekten av individuelle evner er det samme på alle klassetrinn ($\tau_7=\tau_{10}$), og (ii) familie-/miljøvariabler under ungdomsskoletiden påvirker ikke utfallet i 10. trinn, dvs. at summen av familie-/miljøvariablene er de samme på 7. som på 10. trinn ($\omega_7=\omega_{10}$). Når individuelle evner har en konstant effekt på tvers av klassetrinn er all innflytelse fra uobserverte forskjeller kontrollert for gjennom kunnskapsnivået ved inngangen til perioden vi ser på. Tilsvarende fanger resultatet fra det laveste klassetrinnet opp alle miljøvariabler som har lik påvirkning på ulike alderstrinn.

Merk at dette er tilstrekkelige betingelser, skoleeffektanslagene er konsistente dersom de er oppfylt. Men antakelsene er ikke nødvendige i den forstand at alt blir galt dersom de i virkeligheten ikke er oppfylt. Det er kun hvis de samlete effektene av individuelle evner og miljø varierer systematisk mellom skoler at anslagene for skoleeffektene blir gale. Dessverre finnes det ingen test i vårt tilfelle som kan gi oss svar på omfanget av en slik mulig skjevhet.

I den empiriske analysen vil vi også benytte andre modellspesifikasjoner enn den mest restriktive spesifikasjonen. Vi vil operere med modeller der vi estimerer $\lambda$-parameteren. Det innebærer at resultatmålet vi studerer ikke er endringen i prøveresultat fra ett tidspunkt til et annet, men vi søker å forklare resultater på et tidspunkt med tidligere resultater, skoletilhørighet og eventuelle andre variable. Vi legger dermed ingen antakelse om graden av persistens til grunn, men denne estimeres sammen med skoleeffektene. Vi vil også undersøke betydningen av å inkludere ulike familiebakgrunnsvariable (jf. "miljøfaktorer" i drøftingen ovenfor). Dette gjør at vi tillater at effekten av familiebakgrunn ikke fullt ut fanges av resultatmålet fra inngangen til perioden. Som nevnt innledningsvis er det et empirisk spørsmål hvor mye dette betyr for de estimerte skoleeffektene.

Når det gjelder hvilke tidligere prøveresultater man skal kontrollere for, er det viktig å merke seg at bortsett fra i tilfellet med den mest restriktive value added-modellen, hvor man ser på endringen i resultater/kunnskapsnivå fra den ene perioden til den andre, så er det ikke nødvendigvis noe krav om at resultatene på de to tidspunktene skal stamme fra samme type prøve, eller å ha samme vurderingsskala. Som definisjonen av en value added-modell i kapittel 3 poengterer, er kjennetegnet ved en slik modell at den tar hensyn til eller kontrollerer for kunnskaps- og ferdighetsnivået ved inngangen til den perioden man ser på. Det kan gjøres ved å inkludere tidligere resultater i det eller de fagene man ser på, men ettersom hensikten er å gi en best mulig beskrivelse av kunnskaps- og ferdighetsnivået til elevene ved inngangen til den perioden man ser på, vil det generelt være mulig å forbedre indikatoren ved å ta hensyn til et rikere sett av kjennetegn. Hvorvidt dette har noen betydningsfull effekt på de estimerte indikatorene er et empirisk spørsmål.

## 4.3. Relasjon mellom value added og andre skolebidragsindikatorer

Ideen bak value added-indikatorer, og dermed det modellmessige rammeverket, sammenfaller med tilsvarende skolebidragsindikatorer for grunnskolen basert på data der elvenes utfall kun observeres en gang (tverrsnittsindikatorer). Utgangspunktet er ønske om å kunne si noe om forskjeller i skolenes bidrag til elevenes læring, sammen med en erkjennelse av at ukorrigerte resultatgjennomsnitt på skolenivå kan gi et misvisende bilde av dette skolebidraget.

Hovedårsaken til dette er at forskjeller i elevgrunnlag mellom skoler og tilfeldig variasjon i resultater i stor grad kan påvirke anslagene på bidragene til læring på de enkelte skolene. Disse faktorene kan i liten grad sies å være innenfor skolens kontroll. Ved hjelp av tilgjengelige data søker man derfor å korrigere resultatene på

skolenivå for forskjeller i elevgrunnlag. Med regresjonsanalyse trekker man ut den delen av resultatet som skyldes at elevenes bakgrunn ved en skole avviker fra gjennomsnittet blant elevene på alle skolene som er med i analysen. For hver skole sitter vi da igjen med et "skolebidrag", som kan tolkes som det gjennomsnittet vi forventer at skolen ville hatt, om elevsammensetningen ved skolen var lik gjennomsnittet blant de elevene/skolene som er inkludert i analysen.

Tolkningen av de justerte skoleresultatene er selvsagt avhengig av hvilke elevkjennetegn som er inkludert i modellen. I beregningsopplegget for tverrsnittsindikatorer (Hægeland, Kirkebøen, Raaum og Salvanes, 2005a), baserte vi oss på en lang rekke variable – hentet fra administrative registre - som beskrev elevenes sosioøkonomiske bakgrunn. Formålet med dette var ikke å måle betydningen av familiebakgrunn i seg selv, men å kontrollere for forskjeller i resultater som kan tilskrives andre faktorer enn selve skolen eleven går på. I prinsippet er beregningsopplegget for value added-indikatorene tilsvarende som for tverrsnittsindikatorer, bortsett fra at man har et rikere sett av elevkjennetegn. Mens tverrsnittsindikatorer kan tolkes som det gjennomsnittsresultatet vi ville forvente at skolen ville hatt, dersom alle elevene hadde en gjennomsnittlig familiebakgrunn, kan value added-indikatorene tolkes som det gjennomsnittsresultatet vi ville forvente dersom elevene ved skolen hadde gjennomsnittlige resultater fra tidligere (og gjennomsnittlig familiebakgrunn i den grad man også kontrollerer for dette). Med kjennskap til tidligere resultater har vi kartlagt kunnskapsnivået den enkelte elev hadde ved inngangen til den perioden vi ser på langt bedre enn ved hjelp av familiebakgrunnskjennetegn alene. Når man bare har familiebakgrunnskjennetegn, kontrollerer man for elevenes forutsetninger på en indirekte måte, ettersom man tilordner hver elev et kunnskapsnivå lik gjennomsnittet av de elevene med samme sosioøkonomiske bakgrunn.

Som nevnt tidligere i rapporten, har den estimerte skoleeffekten i value added-modeller en tolkning som gjør den mer egnet til bruk i skoleutvikling enn hva som er tilfellet for tverrsnittsindikatorer. Siden vi kontrollerer for tidligere resultater, har value added-indikatoren en presis tolkning som skolens bidrag til elevenes læring i løpet av tiden som ligger mellom de to måletidspunktene, relativt til andre skoler.[2] Tolkningen av skolebidragsindikatorer fra tverrsnittsdata er ikke så klar siden bidraget målt på ett tidspunkt lett fanger opp skoleeffekter fra mange år tilbake, så lenge disse er korrelert over klassetrinn.

Et interessant spørsmål er hvilken betydning det har å kontrollere for familiebakgrunnsvariable gitt at man kontrollerer for grunnskoleresultater, med andre ord hvorvidt familiebakgrunn gir noen vesentlig tilleggsinformasjon når vi allerede kontrollerer for grunnskoleresultater. Dette er drøftet i OECD (2008), og Hægeland og Kirkebøen (2008). Hovedinnsikten herfra er at dette i stor grad er et empirisk spørsmål. I vår anvendelse er det viktig å understreke at selv om det å inkludere et sett av variabler ikke bidrar særlig til å øke forklaringskraften til modellen totalt sett, kan det ha betydning for indikatorene til enkeltskoler. Hvis det er slik at indikatorer uten familiebakgrunnsvariable stort sett gir samme resultater, er det mulig for skoleeiere å beregne disse indikatorene selv basert på de data de selv rår over.

Alt i alt gir dette følgende kriterier for hva slags bakgrunnsvariabler som skal tas med i modellen:

1. Det må være en sammenheng mellom variabelen og skoleresultater,
2. skolenes elevsammensetning varierer med hensyn til det aktuelle kjennetegnet, og
3. variabelen, målt på en konsistent måte, må være tilgjengelig for (tilnærmet) hele elevmassen siden vi ønsker å lage indikatorer for alle skoler, basert på samtlige elever på trinnet.

[2] Det er verdt å minne om at alle indikatorer er relative innen kull av norske elever. Hvorvidt norske skoler i gjennomsnitt er gode eller dårlige til å øke elevenes kunnskaper krever sammenlikning over tid eller mellom land.

Det er verd å merke seg at disse kriteriene gjelder betinget på hvilke andre variable som er med i modellen. Hvis det er slik at vi uansett kontrollerer for elevenes tidligere resultater, må kriteriene (1) og (2) være oppfylt betinget på disse, dvs. at det å inkludere flere variable i modellen bidrar med relevant tilleggsinformasjon om elevmassen.

Når det i det aktuelle datamaterialet ikke er noen sammenhenger mellom en variabel og skoleresultatene, påvirkes ikke resultatene for justerte skolegjennomsnitt. Fravær av systematiske forskjeller mellom skoler i elevsammensetning langs en bestemt dimensjon representerer for så vidt ikke noe problem, den eneste konsekvensen blir at modellen blir mer omfattende. Samtidig er det ingen grunn til å gjøre modellen mer komplisert og omfattende enn nødvendig.

## 4.4. Estimering av value added-indikatorene

Value added-indikatorene fremkommer ved å estimere følgende regresjonsmodell, som er en forenklet versjon av rammeverket i kapittel 4.1:

$$A_{isg} = \alpha_g + \sum_{s=1}^{N} b_s S_{isg} + \gamma F_{ig} + \lambda A_{ig'} + u_{isg},\ g > g' \quad (6)$$

Her er $A_{isg}$ et resultatmål for elev $i$ på trinn $g$ ved skole $s$, $F_{ig}$ er en vektor av familiebakgrunnsvariable og $A_{ig'}$ en vektor (eller en enkelt skalar) av tidligere skoleresultater på trinn g' for elev $i$. $S_{isg}$ er en såkalt dummyvariabel, som er lik 1 dersom elev $i$ er elev ved skole $s$ på trinn $g$ og null ellers, mens $u_{isg}$ er et restledd som fanger opp utelatte kjennetegn og tilfeldigheter.

Modellen er ekvivalent med en modell med såkalte "faste effekter" på skolenivå. Modellen, dvs. $\gamma$- og $\lambda$-vektoren (eller parameteren i tilfelle $A_{ig'}$ er en skalar) samt settet av $b_s$ estimeres ved hjelp av minste kvadraters metode. Når familiebakgrunnsvariablene og tidligere resultater måles som avvik fra sine respektive gjennomsnitt i datamaterialet, har de estimerte skoleparameterne $\hat{b}_s$ tolkning som gjennomsnittsresultater på skolenivå justert for resultatforskjeller som skyldes ulik observert elevsammensetning. De estimerte skoleparameterne $\hat{b}_s$ er anslag på skoleeffektene i rammeverket i kapittel 4.1;

$$b_s = \sum_{h=g}^{g'} \beta_{hgs(i,h)},$$

og utgjør således skolebidragsindikatorene basert på value added. Ved å utelate variablene i $F_i$ eller $A_{ig'}$ får vi indikatorer basert på bare grunnskoleresultater og indikatorer basert på bare familiebakgrunnsvariable, henholdsvis. De estimerte skoleparameterne $\hat{b}_s$ representerer konsistente anslag på skoleeffektene hvis og bare hvis forventningen til det gjennomsnittelige restledd for hver enkelt skole er lik null.

Hvis vi i (6) lar $A_{ig}$ og $A_{ig'}$ begge være skalarer, med samme skala og måltall, vil vi få en mer restriktiv modellspesifikasjon ved å sette $\lambda = 1$. Dette gir oss følgende empiriske modell:

$$A_{isg} - A_{ig'} = \alpha_g + \sum_{s=1}^{N} b_s S_{isg} + \gamma F_{ig} + u_{isg},\ g > g' \quad (7)$$

Modellen i (7) er mer intuitiv enn modell (6), siden den avhengige variabelen i større grad har tolkningen av "kunnskapsøkning". Imidlertid (7) mer restriktiv i den forstand at den pålegger en en-til-en sammenheng mellom tidligere og nåværende

resultater. Modellen av typen (6) er den vanligste i den internasjonale litteraturen, jf. OECD (2008).

Skolebidragsindikatorene vi tidligere har beregnet baserer seg på et statistisk modellrammeverk som går under betegnelsen "faste effekter" ("fixed effect")-modeller. Dette rammeverket er benyttet også i mange value added-modeller. Det finnes imidlertid andre rammeverk som er mye benyttet, bl.a. beskrevet i OECD (2008). De empiriske studiene som finnes på feltet, konkluderer stort sett med at valg av rammeverk blant de mest benyttede har moderate konsekvenser for indikatorene. Som en del av rapporten beregner vi også value added-indikatorer basert på noen andre modelltyper. De statistiske modelltypene vi ser på, skiller seg fra hverandre i forhold til hvordan skoleeffekten spesifiseres. I "faste effekter"-modellen betraktes skoleeffektene som faste parametere som skal estimeres, mens de i den såkalte "tilfeldige effekter" ("random effects")-modellen betraktes som stokastiske variable. En tredje mye brukt spesifikasjon er å estimere modellen ved hjelp av vanlig minste kvadraters metode, dvs. uten å utnytte informasjon om hvilken skole elevene tilhører, for så å beregne skoleeffektene som gjennomsnittlige residualer pr. skole.

Den siste metoden, å se på gjennomsnittlige residualer, utnytter ikke all informasjon som ligger i data, noe som gjør den mer upresis enn de andre tilnærmingene. Dette vil bety mer jo sterkere de faktiske skoleeffektene er. "Tilfeldige effekter"-modellen bygger på en forutsetning om at alle de forklaringsvariablene som er inkludert i modellen er ukorrelert med skoleeffekten, mens "faste effekter"-modellen tillater fri korrelasjon mellom skoleeffekten og de andre variablene. Forutsetningen om at skoleeffektene er ukorrelerte med de andre variablene i modellen, som ligger under random effects-modellen, er restriktiv. Dersom den er oppfylt, gir den imidlertid estimater som har lavere statistisk usikkerhet. I modellen med faste effekter pålegger vi ingen restriksjoner på skole-effektene. Dette er dermed en metode som i utgangspunket er mer robust i forhold til å gi forventningsrette estimater av skole-effektene, men for en gitt størrelse på datamaterialet gir de noe større standardfeil. I den empiriske delen av rapporten undersøker vi hvilken praktisk betydning det har for valg mellom alternative modellspesifikasjoner.

## 4.5. Mulige feilkilder knyttet til frafall

Utgangspunktet for analysene er et ønske om å si noe om skolenes kvalitet. Imidlertid vil alle estimeringer basere seg på det datamaterialet som finnes, de elevene som har registrert resultater. Som vi viser i kapittel 5 utgjør elevene med registrerte resultater på nasjonale prøver og avgangskarakterer en svært stor andel av antall elever på tilsvarende trinn i følge Grunnskolens informasjonssystem (GSI) på landsbasis. Det kan imidlertid være variasjoner, skoler kan ha forskjellig praksis for fritak, og enkeltskoler kan ha en stor andel elever som ikke tar prøvene.

Dersom elever som fritas ikke er representative for elevgruppen som helhet – dette vil de sannsynligvis ikke være, dersom fritaksreglementet følges – vil heller ikke de beregnede indikatorene være representative for opplæringen til hele elevgruppen. Dette trenger imidlertid ikke å være et problem som skaper avvik i forhold til hva indikatoren skal måle. Value added-indikatorene måler i utgangspunktet skolens bidrag til elevenes læring generelt, men trenger ikke være særlig godt egnet til å vurdere kvaliteten på tilbudet til et lavt antall elever med spesielle utfordringer og tilbud. Dermed vil value added-indikatorene fange opp kvaliteten først og fremst på skolens ordinære tilbud, til flertallet av elevene, uavhengig av fritaksandelen – dvs. hvor stort dette flertallet er.

Det kan imidlertid oppstå skjevheter som forstyrrer sammenligning av forskjellige skoler, dersom fritakspraksisen varierer mellom skoler. Dersom lavt presterende elever ved en skole i mindre grad avlegger nasjonale prøver vil dette bidra til at skolens snittresultat blir kunstig høyt. Motivasjonen for dette vil reduseres ved bruk

av value added-indikatorer, ettersom vi da tar hensyn til elevenes forutsetninger. En svakt presterende elev kan bidra til en god value added-indikator for sin skole, dersom skolen har lyktes med å få eleven til å prestere godt *relativt til elevens (observerte) forutsetninger*. Dersom en skole skulle ønske å manipulere value added-indikatoren må den utelate elever som presterer lavt, relativt til deres (observerte) forutsetninger. Dette er en mindre tydelig gruppe enn elever som generelt presterer lavt. Det er likevel mulig at praksis mht. hvilke elever som deltar varierer mellom skoler, og at dette påvirker resultatene. For informasjonsverdien til value added-indikatorer, eller mer generelt, all sammenligning av elevers prestasjoner, er det åpenbart viktig at gjennomføring av nasjonale prøver foregår på mest mulig samme måte på forskjellige skoler.

I prinsippet er det mulig å studere deltagelse i for eksempel nasjonale prøver, og estimere en egen indikator. Dette svarer omtrent til hva Hægeland, Kirkebøen og Raaum (2010) gjør når de ser på både karakterer fra Vg1 og andelen som fullfører Vg1 innen en gitt tid. Overdrevent fritak vil da kunne gi urealistisk høye indikatorer basert på poeng, men vil samtidig fanges opp som en lav andel som avlegger prøven, hensyn tatt til elevsammensetning. For at dette skal kunne gjennomføres må det være mulig å ha en oversikt over hvilke elever som ikke avlegger prøven. Ettersom vi finner et (beskjedent) avvik mellom totalt antall elever i resultatene og GSI, kan dette være krevende. Videre bør en fortrinnsvis ha mest mulig presise kjennetegn som kan forklare prøvedeltagelse (informasjon om spesialundervisning, særskilt språkopplæring, norskferdigheter), både for de som deltar og de som ikke deltar. Skoletilknytning kan også være uklar for noen av elevene som ikke deltar, for eksempel dersom de går på en spesialskole, men er registrert ved sin lokale grunnskole.

Til slutt, ettersom value added-indikatorene tar hensyn til tidligere resultater, vil manglende resultater også påvirke estimerte resultater på senere trinn. Dersom en elev med resultater på 8. trinn mangler resultater fra 5. trinn, tas dette hensyn til ved at vi forventer at denne presterer som gjennomsnittet av alle elever som mangler resultater.

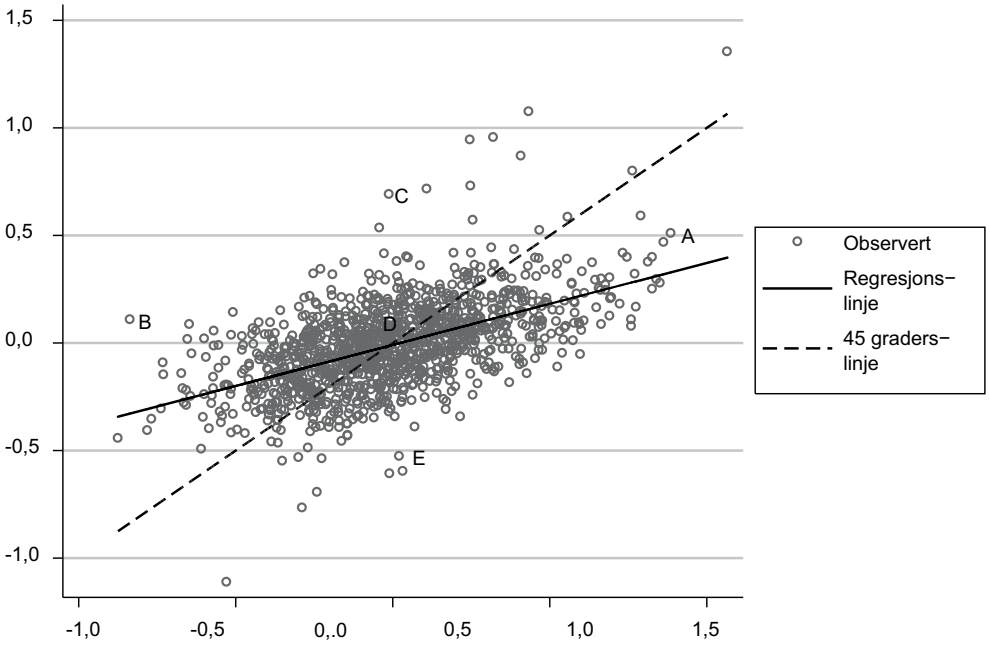
## 4.6. Presentasjon av indikatorene i denne rapporten

En viktig motivasjon for rapporten er å undersøke i hvilken grad value added-indikatorer gir et annet bilde av hvilke skoler som gir gode bidrag til elevenes læring enn ujusterte resultater, og om hvordan justeringen for elevsammensetning foretas (spesifikasjon av modellen) har stor betydning for resultatene. For å illustrere dette benytter vi både tabeller og grafiske framstillinger. Figuren nedenfor er en prototype vi benytter ofte i denne rapporten, både for å sammenligne indikatorer med ujusterte resultater, og for å sammenligne to ulike indikatorer. Her gir vi en veiledning i hvordan slike figurer kan leses.

I figuren nedenfor, som er identisk med Figur 6.6, sammenligner vi en value added-indikator for mellomtrinnet basert på nasjonale prøver 8. trinn med tilsvarende ujusterte resultater for nasjonale prøver 8. trinn. Hver sirkel i figuren representerer en skole, og det ujusterte resultatet måles langs den horisontale aksen, mens value added-indikatoren måles langs den vertikale. Dersom value added-indikatoren ikke innebar noen justering av resultatene for en skole, ville sirkelen for denne skolen ligge på den stiplede linjen (45-graderslinjen) i figuren. Avstanden fra 45-graderslinjen sier noe om hvor stor justering value added-indikatoren innebærer. Hvis vi for eksempel ser på skole A i figuren, har den et ujustert resultat på ca. 0,9, men et justert resultat på 0,5. Alle skoler som ligger under 45-graderslinjen får sine resultater nedjustert ved SBI. Skole B får derimot oppjustert sine resultater betydelig ved beregningen av value added-indikatoren. Hvis vi ser på skolene C, D og E, så har de temmelig like ujusterte resultater, mens deres value added-indikatorer er nokså forskjellige.

Videre inneholder figuren en regresjonslinje. Helningen på denne forteller oss styrken på sammenhengen mellom ujusterte resultater og value added-indikatoren. Det samme gjør de rapporterte korrelasjonskoeffisientene. I dette tilfellet er det en klar positiv samvariasjon, det er altså en tendens til at skolebidraget er høyere jo sterkere ujusterte resultater skolen oppnår, selv om det er mange enkelteksempler på det motsatte.

**Figur 4.1. Eksempelfigur**




1211 skoler med 50809 elever. Elevvektet reg.koeff: 0.381; korrelasjon=0.567; rangkorr.=0.563

# 5. Datagrunnlag og avgrensninger

Analysene gjøres mulig av og baserer seg på rike data fra koblede administrative registre. To typer datakilder benyttes, skoleresultater og data som beskriver elevenes bakgrunn, herunder bl.a. foreldres utdanning, inntekt og innvandrings-bakgrunn.

## 5.1. Karakterer og resultater fra nasjonale prøver

Skoleresultatene måles ved elevprestasjonene slik det fremkommer gjennom karakterer og resultater på nasjonale prøver. Skolekvalitet estimeres som skolenes bidrag til disse. For alle resultatene med unntak av de nasjonale prøvene for 5. trinn beregner vi value added-indikatorer, dvs. indikatorer som utnytter informasjon om tidligere prestasjoner på lavere trinn og dermed måler skolebidraget for en spesifikk tidsperiode. Resultatene vi bruker er

1. Nasjonale prøver 5. trinn
2. Nasjonale prøver 8. trinn
3. Avsluttende standpunkt- og eksamenskarakterer ved fullført grunnskole

I tillegg trekker vi på tilsvarende analyser for videregående skoler i Oslo (Hægeland, Kirkebøen og Raaum, 2010). Her benyttes karakterer og fullføring på videregående skole som resultatmål, mens avgangsresultater fra grunnskolen representerer tidligere resultater.

### Nasjonale prøver 5. trinn

Formålet med nasjonale prøver er å vurdere i hvilken grad skolen lykkes med å utvikle elevenes ferdigheter i lesing og regning, og i deler av faget engelsk. Resultatene skal brukes av skoler og skoleeiere som grunnlag for kvalitetsutvikling i opplæringen. Nasjonale prøver er ikke prøver i enkeltfag, men i grunnleggende ferdigheter. Prøvene i lesing og regning tar derfor ikke bare utgangspunkt i kompetansemålene i norsk og matematikk, men også i andre fag der mål for lesing og regning er integrert. Prøvene i engelsk skiller seg fra de to andre prøvene ved at de tar utgangspunkt i kompetansemål i ett fag. Nasjonale prøver gjennomføres på høsten, kort tid etter at elevene har startet på 5., 8. og 9. trinn.[3]

Hver prøve skåres med et antall poeng, der det maksimale antallet poeng varierer mellom fag. Maksimalt antall poeng er imidlertid høyt nok til at poengsummen med rimelighet kan behandles som en kontinuerlig variabel. Ettersom prøvene ikke har noen naturlig tolkbar skala, og antall poeng også varierer mellom fag, standardiserer vi resultatene.,Vi regner resultatene i enheter av standardavvik og beregner indikatorer for hver enkeltprøve samt gjennomsnitt for alle prøver. Standardiseringen gir en felles skala, som gjør det meningsfullt å regne ut gjennomsnitt. Tabell 5.1 og Tabell 5.2 gir beskrivende statistikk for både opprinnelige og standardiserte prøvepoeng.. I tråd med tidligere arbeider med skolebidragsindikatorer bruker vi data fra to årganger(2009 og 2010).

[3] Formålet og beskrivelsen av nasjonale prøver er hentet fra http://www.udir.no/Vurdering/Nasjonale-prover/Om-nasjonale-prover/, der det også finnes mer informasjon om prøvene.

**Tabell 5.1. Beskrivende statistikk, poeng på nasjonale prøver 5. trinn. 2009 og 2010**

| | Antall elever | Snitt | Standard-avvik | Min | 10. per-sentil | 25. per-sentil | 50. per-sentil | 75.per-sentil | 90.per-sentil | Maks. |
|---|---|---|---|---|---|---|---|---|---|---|
| **2009** | | | | | | | | | | |
| Engelsk ..... | 57 145 | 22,67 | 8,304 | 0 | 12 | 16 | 22 | 29 | 34 | 40 |
| Regning .... | 57 121 | 25,89 | 9,263 | 0 | 14 | 19 | 25 | 33 | 39 | 48 |
| Lesing ..... | 55 818 | 19,84 | 6,732 | 0 | 10 | 15 | 21 | 25 | 28 | 32 |
| **2010** | | | | | | | | | | |
| Engelsk ..... | 57 368 | 25,61 | 7,388 | 0 | 16 | 20 | 26 | 31 | 36 | 42 |
| Regning .... | 57 785 | 25,70 | 8,943 | 0 | 14 | 19 | 26 | 33 | 38 | 44 |
| Lesing ..... | 54 842 | 18,89 | 6,485 | 0 | 10 | 14 | 19 | 24 | 27 | 33 |

**Tabell 5.2. Beskrivende statistikk, standardiserte poeng på nasjonale prøver 5. trinn. 2009 og 2010**

| | Antall elever | Snitt | Standard-avvik | Min | 10. per-sentil | 25. per-sentil | 50. per-sentil | 75.per-sentil | 90.per-sentil | Maks. |
|---|---|---|---|---|---|---|---|---|---|---|
| **2009** | | | | | | | | | | |
| Engelsk .. | 57 145 | 3,024 | 1,108 | 0 | 2 | 2 | 3 | 4 | 5 | 5 |
| Regning . | 57 121 | 2,614 | 0,935 | 0 | 1 | 2 | 3 | 3 | 4 | 5 |
| Lesing .... | 55 818 | 3,095 | 1,050 | 0 | 2 | 2 | 3 | 4 | 4 | 5 |
| Snitt ..... | 58 994 | 2,894 | 0,894 | 0 | 2 | 2 | 3 | 4 | 4 | 5 |
| **2010** | | | | | | | | | | |
| Engelsk .. | 57 368 | 3,417 | 0,985 | 0 | 2 | 3 | 3 | 4 | 5 | 6 |
| Regning . | 57 785 | 2,595 | 0,903 | 0 | 1 | 2 | 3 | 3 | 4 | 4 |
| Lesing .... | 54 842 | 2,946 | 1,011 | 0 | 2 | 2 | 3 | 4 | 4 | 5 |
| Snitt ..... | 58 926 | 2,971 | 0,842 | 0 | 2 | 2 | 3 | 4 | 4 | 5 |

Vi ser at de opprinnelige fordelingene er tilsvarende på tvers av fag og år, men ikke helt like. De standardiserte fordelingene er likere, både mellom fag innen år og over tid, men heller ikke disse er helt identiske. Til slutt ser vi at spredningen i snittet av de tre prøvene er lavere enn spredningen i de enkelte testene.[4] Dette kan være et uttrykk for at prøvene i noen grad fanger det samme, en form for underliggende ferdighet, men ikke måler dette helt presist. Snittet av flere prøver kan imidlertid gi et mer presist mål enn enkeltprøvene. Tabell 5.3 viser samvariasjonen mellom prøvene. Vi ser at resultatene i enkeltprøvene er forholdsvis høyt korrelerte, med korrelasjonskoeffisienter på 0,5-0,6. Alle enkeltprøvene er høyt korrelert med snittet, med korrelasjonskoeffisienter på 0,8-0,9.

**Tabell 5.3. Korrelasjon mellom forskjellige nasjonale prøver, 5. trinn. 2009 og 2010**

| | Engelsk | Regning | Lesing | Snitt |
|---|---|---|---|---|
| Engelsk ..... | 1 | | | |
| Regning ..... | 0,514 | 1 | | |
| Lesing ..... | 0,620 | 0,621 | 1 | |
| Snitt ..... | 0,853 | 0,825 | 0,883 | 1 |

Alle elever som har registrert prøveresultater inngår i datamaterialet, unntatt totalt 144 elever (jevnt fordelt på 2009 og 2010) som vi ikke er i stand til å knytte til noen skole. Elevenes prestasjoner knyttes til skolen de er registrert ved på prøvetidspunktet. For snittpoeng utgjør datamaterialet alle elever med resultater i minst ett fag, vi tar i analysene hensyn til hvorvidt elevene mangler resultater i enkelte fag. Totalt har vi registrert minst ett resultat for 117 920 elever, mens vi for 121 664 elever har registrert enten ett eller flere resultater, eller en deltattkode (deltatt, fritatt eller ikke deltatt). Dette utgjør hhv. 95 og 98 prosent av de elevene som i følge Grunnskolens informasjonssystem (GSI) var registrert på 5. trinn i 2009 eller 2010. Fra Tabell 5.1 ser vi at det er flere som mangler resultat i lesing enn de andre prøvene.

Vi har ikke tilgang til noe tidligere prestasjonsmål, kan dermed ikke beregne noen value added-indikatorer for disse resultatene. Ettersom prøvene avlegges tidlig på høsten på 5. trinn blir det estimerte skolebidraget et mål på bidraget over trinnene 1-4.

[4] Standardavviket til de enkelte testene avviker litt fra 1 fordi de er standardisert med standardavviket for alle årene 2007-2010.

## Nasjonale prøver 8. trinn

Elevene testes i lesing, regning og engelsk, som på 5. trinn. Totalt har vi minst ett prøveresultat for 60 696 elever, samt 636 som vi ikke er i stand til å knytte til noen skole, og deltattstatus for 61 731 elever. Dette utgjør hhv. 96 og 97 prosent av elevene registrert på 8. trinn i 2010 i GSI. Skolen vi knytter elevene til er avgiverskolen, dvs. skolen de gikk i ved utgangen av 7. trinn. I de tilfellene der elever bytter skoler ved overgangen til ungdomstrinnet tolker vi elevenes resultat tidlig på høsten i 8. trinn som et resultat av forrige skoles kvalitet. Vi antar altså at ungdomsskolen der de faktisk gjennomførte prøvene ikke har noen vesentlig påvirkning på resultatene etter så kort tid. Ettersom vi i beregningen av value added-indikatorer tar hensyn til ferdighetsnivået ved starten av 5. trinn gir indikatorene et mål på kvaliteten over trinnene 5-7.

Vi beregner også her indikatorer for hvert fag, samt for snittet av alle prøvene. Som for prøvene på 5. trinn standardiserer vi poengsummen for å få en felles skala. Elevenes prestasjoner knyttes til registrert avgiverskole, dvs. skolen elevene avsluttet mellomtrinnet ved. Tabell 5.4 og Tabell 5.5 gir beskrivende statistikk for prøveresultatene, mens Tabell 5.6 viser samvariasjonen mellom elevenes resultater på forskjellige prøver. Hovedmønstrene er som for prøvene på 5. trinn.

**Tabell 5.4. Beskrivende statistikk, poeng på nasjonale prøver 8. trinn. 2010**

| | Antall elever | Snitt | Standard-avvik | Min | 10. per-sentil | 25. per-sentil | 50. per-sentil | 75.per-sentil | 90.per-sentil | Maks. |
|---|---|---|---|---|---|---|---|---|---|---|
| Engelsk ... | 58 378 | 28,57 | 9,965 | 0 | 14 | 21 | 30 | 37 | 41 | 48 |
| Lesing ..... | 56 982 | 22,82 | 7,516 | 0 | 13 | 17 | 23 | 28 | 33 | 42 |
| Regning .. | 59 047 | 30,97 | 11,70 | 0 | 15 | 22 | 31 | 40 | 47 | 58 |

**Tabell 5.5. Beskrivende statistikk, standardiserte poeng på nasjonale prøver 8. trinn. 2010**

| | Antall elever | Snitt | Standard-avvik | Min | 10. per-sentil | 25. per-sentil | 50. per-sentil | 75.per-sentil | 90.per-sentil | Maks. |
|---|---|---|---|---|---|---|---|---|---|---|
| Engelsk .. | 58 378 | 2,867 | 1,000 | 0 | 1 | 2 | 3 | 4 | 4 | 5 |
| Lesing .... | 56 982 | 3,036 | 1,000 | 0 | 2 | 2 | 3 | 4 | 4 | 6 |
| Regning . | 59 047 | 2,648 | 1,000 | 0 | 1 | 2 | 3 | 3 | 4 | 5 |
| Snitt ..... | 60 696 | 2,830 | 0,889 | 0 | 2 | 2 | 3 | 4 | 4 | 5 |

**Tabell 5.6. Korrelasjon mellom forskjellige nasjonale prøver, 8. trinn. 2010.**

| | Engelsk | Lesing | Regning | Snitt |
|---|---|---|---|---|
| Engelsk ..... | 1 | | | |
| Lesing ..... | 0,711 | 1 | | |
| Regning ..... | 0,578 | 0,652 | 1 | |
| Snitt ..... | 0,876 | 0,903 | 0,856 | 1 |

Vi beregner value added-indikatorer ved å ta hensyn til resultater på 5. trinn. Alle elever med resultater fra 8. trinn inngår i analysene, som for analysene av 5. trinn, men vi tar hensyn til hvorvidt elevene mangler enkeltprøver fra 5. trinn, samt hvorvidt de i det hele tatt finnes i datamaterialet for 5. trinn. Ettersom vi ønsker å korrigere for variasjon i tidligere resultater er vi henvist til å bare bruke én årgang med elever, de som var på 5. trinn høsten 2007 (og dermed hadde de første nasjonale prøvene på dette trinnet) og på 8. trinn høsten 2010. For 57969 elever har vi minst ett resultat fra nasjonale prøver på 5. trinn, dvs. for omtrent 6 prosent har vi ingen resultater. For de elevene som faktisk har resultater fra 8. trinn er det omtrent 5 prosent som ikke har noen registrert på 5. trinn. Tabell 5.7 gir beskrivende statistikk for resultatene på 5. trinn til de elevene vi observerer på 8. trinn.

**Tabell 5.7. Resultater fra standardiserte prøver 5. trinn, for elever testet på 8. trinn i 2010**

| | Antall elever | Snitt | Standardavvik |
|---|---|---|---|
| Engelsk | 56 774 | 20,769 | 6,464 |
| Regning | 56 372 | 31,368 | 10,658 |
| Lesing | 56 219 | 16,971 | 6,488 |
| Snitt (standardisert) | 57 969 | 0,000 | 0,859 |

## Grunnskoleresultater

Ved fullført grunnskole registreres alle standpunkt- og eksamenskarakterer. Vi benytter oss bare av skriftlig eksamenskarakter og standpunktkarakterene i basisfagene norsk hovedmål, matematikk og engelsk skriftlig. Når vi utelater 234 personer uten gyldig skoletilknytning, har vi 61030 elever med minst en av disse karakterene. Dette utgjør ca 96 prosent av elevene registrert på 10. trinn skoleåret 2009/2010 i GSI.[5] De aller fleste av disse har én skriftlig eksamenskarakter og alle standpunktkarakterene. Elevene knyttes til skolen der de avslutter 10. trinn. Ettersom vi tar hensyn til ferdigheter ved starten av 8. trinn blir estimerte indikatorer et mål på kvalitet på ungdomstrinnet.

Vi beregner to snittkarakterer, en som er et gjennomsnitt av standpunktkarakterene i fagene, og en som er den skriftlige eksamenskarakteren i ett av disse fagene. Den første beregnes bare for de elevene som har standpunktkarakter i alle tre fagene, mens den andre beregnes bare for de elevene som har eksamenskarakter i eksakt ett av fagene. Ettersom karakterpoeng er en enhet som har en rimelig klar tolkning og en relativt stabil fordeling standardiserer vi ikke karakterene. Fra Tabell 5.8, som viser beskrivende statistikk for enkeltkarakterene, og Tabell 5.9, som viser beskrivende statistikk for snittkarakterene, ser vi imidlertid at standardavvikene er nær en. Hvorvidt vi standardiserer eller ikke gjør dermed liten forskjell i akkurat her. Skriftlig eksamenskarakter og standpunktkarakter er forholdsvis høyt korrelert på elevnivå, med en korrelasjonskoeffisient på 0,713.

**Tabell 5.8. Beskrivende statistikk, avgangskarakterer grunnskolen. 2010**

| | Antall elever | Snitt | Standardavvik |
|---|---|---|---|
| Skriftlig eksamen | | | |
| Engelsk | 20 671 | 3,756 | 1,079 |
| Matematikk | 20 712 | 3,239 | 1,198 |
| Norsk hovedmål | 17 684 | 3,518 | 0,987 |
| Standpunkt | | | |
| Engelsk | 59 936 | 3,840 | 1,094 |
| Matematikk | 60 419 | 3,588 | 1,203 |
| Norsk hovedmål | 60 045 | 3,838 | 1,006 |

**Tabell 5.9. Beskrivende statistikk, snittkarakterer fra grunnskolen. 2010**

| | Antall elever | Snitt | Standard-avvik | Min | 10. per-sentil | 25. per-sentil | 50. per-sentil | 75.per-sentil | 90.per-sentil | Maks. |
|---|---|---|---|---|---|---|---|---|---|---|
| Standpunkt | 58 956 | 3,777 | 0,957 | 1 | 3 | 3 | 4 | 4 | 5 | 6 |
| Skriftlig eksamen | 58 870 | 3,505 | 1,117 | 1 | 2 | 3 | 3 | 4 | 5 | 6 |

Vi beregner value added-indikatorer ved å ta hensyn til resultater fra nasjonale prøver i 8. trinn. Datamaterialet omfatter én årgang med elever, de som avla nasjonale prøver på 8. trinn høsten 2007 og fullførte grunnskolen våren 2010. Tabell 5.10 gir beskrivende statistikk for Resultater fra nasjonale prøver 8. trinn, for elever som gikk ut av grunnskolen i 2010. For hver av prøvene er det omtrent 7 prosent av elevene som har minst én grunnskolekarakter som mangler resultater, og omtrent 4 prosent har ingen resultater fra nasjonale prøver på 8. trinn.

[5] Hvis vi også inkluderer elever med for eksempel fraværskoder i minst ett fag, men ikke nødvendigvis noen karakterer, øker antallet elever til 61500.

**Tabell 5.10. Resultater fra nasjonale prøver 8. trinn, for elever som gikk ut av grunnskolen i 2010**

| | Antall elever | Snitt | Standardavvik |
|---|---|---|---|
| Engelsk | 56 487 | 24,496 | 9,433 |
| Regning | 56 472 | 40,762 | 15,483 |
| Lesing | 56 503 | 27,272 | 7,964 |

## 5.2. Elevbakgrunn

Informasjon fra en lang rekke administrative datakilder gir oss en relativt detaljert beskrivelse av personkjennetegn og familiebakgrunn for alle elever som inngår i datamaterialet. Ut fra denne informasjonen har vi beregnet en lang rekke variable, som til sammen gir et bredt grunnlag for å karakterisere elevgrunnlaget ved den enkelte skole. I tillegg til grunnleggende demografisk informasjon, har vi opplysninger om foreldres utdanning, innvandrerbakgrunn, foreldres inntekt, formue og arbeidsledighets- og trygdeforhold.

I de fleste analysene kontrollerer vi for elevbakgrunn ved å inkludere variable for kjønn, mors og fars utdanning (grunnskole, videregående, kort og lang høyere utdanning), familieinntekt siste ti år (i kvintiler) samt innvandringsbakgrunn. Tabell 5.11 viser hvordan elevene fordeler seg på disse kjennetegnene. Ettersom alle elevene i hvert årskull i prinsippet er fanget opp, er det små forskjeller i sammensetning mellom de forskjellige elevgruppene. Utdanningsnivået til foreldrene er høyere for de yngste kullene, mens en økende andel med innvandrerbakgrunn er født i Norge. I analysene av resultater for nasjonale prøver bruker vi også et større sett med familiebakgrunnsvariable. Dette er beskrevet bl.a. i Hægeland, Kirkebøen, Raaum og Salvanes (2005a).

**Tabell 5.11. Andel elever med forskjellige kjennetegn**

| | Nasjonale prøver, 5. trinn | | Nasjonale prøver, 8. trinn | Grunnskole-karakterer |
|---|---|---|---|---|
| | 2009 | 2010 | | |
| Jenter | 0,483 | 0,485 | 0,485 | 0,488 |
| Mors utdanning | | | | |
| Grunnskole | 0,165 | 0,166 | 0,184 | 0,203 |
| Mellomnivået | 0,023 | 0,023 | 0,025 | 0,025 |
| Høyere (inntil 4 år) | 0,325 | 0,342 | 0,316 | 0,293 |
| Høyere (minst 5 år) | 0,072 | 0,074 | 0,062 | 0,053 |
| Mangler informasjon | 0,056 | 0,058 | 0,054 | 0,056 |
| Fars utdanning | | | | |
| Grunnskole | 0,172 | 0,169 | 0,184 | 0,188 |
| Mellomnivået | 0,050 | 0,049 | 0,051 | 0,050 |
| Høyere (inntil 4 år) | 0,205 | 0,212 | 0,197 | 0,185 |
| Høyere (minst 5 år) | 0,099 | 0,101 | 0,095 | 0,090 |
| Mangler informasjon | 0,071 | 0,072 | 0,072 | 0,078 |
| Innvandrer | 0,039 | 0,042 | 0,053 | 0,060 |
| Fra utenfor Vest-Europa mm.[6] | 0,031 | 0,033 | 0,043 | 0,050 |
| Barn av innvandrere | 0,051 | 0,057 | 0,042 | 0,038 |
| Fra utenfor Vest-Europa mm. | 0,045 | 0,048 | 0,038 | 0,035 |
| Antall elever | 60818 | 60846 | 61731 | 61500 |

[6] Utenfor Vest-Europa (omfatter ikke nye EU-land i Øst- og Sentral-Europa), USA, Canada, Australia, New Zealand.

# 6. Indikatorer for mellomtrinnet

Med bakgrunn i den metodiske drøftingen og redegjørelsen for datagrunnlaget tidligere i rapporten, drøfter vi alternative beregninger av value added-indikatorer for mellomtrinnet i dette kapitlet. Vi tar her utgangspunkt i nasjonale prøver for 8. trinn som resultatmål, og bruker resultatene fra nasjonale prøver på 5. trinn for å kontrollere for kunnskapsnivået ved starten av mellomtrinnet. Nasjonale prøver tas på høsten, og prøvene på 8. trinn fungerer da som en test på kunnskapsnivået ved slutten av mellomtrinnet. For mange elever er skolen der testen tas på 8. trinn en annen enn den de var elever ved på mellomtrinnet, men som nevnt tidligere tilordnes disse elevene mellomtrinns-skolen. Vi kontrollerer ikke for hvilken skole testen ble tatt på. Det innebærer en forutsetning om at resultatet på nasjonale prøver 8. trinn ikke er systematisk påvirket av testskolen. Vi ser både på indikatorer for alle fag under ett og indikatorer for enkeltfag.

Hensikten med drøftingen er – med bakgrunn i den teoretiske drøftingen tidligere i rapporten, å undersøke hvor stor betydning valg av modellspesifikasjon har for resultatene. Det er særlig tre ting man må ta stilling til; (i) hvordan man skal kontrollere for tidligere resultater, (ii) hvor stor betydning det har å kontrollere for familiebakgrunn og (iii) hvor stor betydning forutsetningen om skole-effekten (faste eller tilfeldige effekter) har for resultatene. Ut fra denne drøftingen finner vi en foretrukket modell. Målet er en modell som kontrollerer for så mye som mulig av resultatvariasjon som er utenfor skolens kontroll, men uten at modellen blir unødig komplisert.

## 6.1. Indikatorer basert gjennomsnitt for alle prøver

I dette avsnittet gir vi en nokså detaljert gjennomgang av mulige indikatorer basert på resultater fra nasjonale prøver 8. trinn, med utgangspunkt i gjennomsnittlig skår på tvers av prøveemner (lesing, regning, engelsk). Ettersom bruk av poengskalaen på de nasjonale prøvene varierer mellom emner og trinn er resultatene målt i standardavvik, beregnet separat for hvert emne-trinn.

Vår vurdering av egnet modell for hvordan resultater påvirker testutfall tre år senere og betydningen av kontroll for familiebakgrunn tar utgangspunkt i Tabell 6.1, som viser resultatene fra sju ulike spesifikasjoner. I kolonne (1) kontrollerer vi kun for familiebakgrunn. I kolonne (2) og (3) er resultatmålet vi ser på, differansen mellom gjennomsnittsresultatene på 8. og 5. trinn hvilket innebærer at kontrollen for tidligere resultater er implisitt gjennom spesifikasjonen av resultatmålet. Dette er ekvivalent til å sette $\lambda$ lik 1, jf. drøftingen i kapittel 4. I de andre kolonnene er resultatmålet gjennomsnittsresultatet på 8. trinn, målt i standardavvik. Kolonne (4) – (7) kontrollerer for tidligere resultater, enten ved å inkludere gjennomsnittsresultatet fra NP 5. trinn eller resultatet fra hvert enkelt emne for seg. For hver av spesifikasjonene med tidligere resultater, rapporterer vi resultater med og uten kontroll for familiebakgrunn.

Siden deltakelse på prøvene ikke er tilfeldig kontrollerer vi gjennomgående for om eleven mangler resultater på noen av de nasjonale prøvene for 8. trinn. Dette korrigerer for eventuelle poengnivå forskjeller mellom fag, men også for selektivt frafall på enkeltprøver. En implisitt antakelse er her at skolene selv ikke strategisk påvirker frafallet. I tillegg har vi en dummyvariabel for om eleven ikke finnes i datamaterialet for tidligere resultater (i dette tilfellet NP 5. trinn). Vi kontrollerer også for kjønn. For tidligere resultater kontrollerer vi for nivå, samt en dummy-variabel som indikerer om resultatet mangler.

Raden $R^2$ uttrykker hvor stor andel av variansen i det aktuelle resultatmålet som blir forklart av modellen. Kolonne (2) og (3) er dermed ikke sammenlignbare med de andre ettersom all variasjon mellom elever er fjernet. Et hovedinntrykk er at det å inkludere resultater fra nasjonale prøver 5.trinn bidrar til å forklare mye av variasjonen i individuelle resultater, men at forklaringskraften ikke varierer mye

med *hvordan* dette gjøres. Familiebakgrunn forklarer relativt mindre av variasjonen mellom elever. Å inkludere familiebakgrunnsvariable når man først har kontrollert for tidligere resultater har liten betydning. Ser vi nærmere på effekten av tidligere resultater (λ) ligger anslaget rundt 0,8, enten vi ser på felleskoeffisienten i kolonne (4) eller summen av de tre fagene i (6). Dersom denne koeffisienten var lik eller nær 1, ville den rendyrkede value added-modellen basert på differansen i resultater ha vært å foretrekke. I dette tilfellet er imidlertid λ-koeffisienten relativt langt unna en, og forskjellen er klart statistisk signifikant. Å benytte modellen basert på differansen i resultater, vil dermed innebære å pålegge en restriksjon i modellen (λ=1) som ikke har støtte i data. Det finnes imidlertid ingen klart definert grense for hvor nær 1 λ må være før vi ville foretrukket den modellen. Valg av modell bestemmes mer av i hvilken grad de estimerte skolebidragsindikatorene påvirkes av modellvalget. Som vi vil se nedenfor (Tabell 6.3) er korrelasjonen mellom indikatorer basert på modeller med lambda lik 1 og modeller der λ tillates å variere fritt alle lavere enn 0,9. Selv om dette er en høy korrelasjon, innebærer dette at det estimerte skolebidraget for en del skoler kan påvirkes betydelig av at denne restriksjonen pålegges.

**Tabell 6.1. Regresjonsresultater, basert på gjennomsnittsresultat nasjonale prøver 8. trinn**

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| | SnittNP8 | Snitt NP8-snitt NP5 | Snitt NP8-snitt NP5 | SnittNP8 | SnittNP8 | SnittNP8 | SnittNP8 |
| Mangler res NP8 engelsk | -0,357*** | -0,0786*** | -0,0709*** | -0,202*** | -0,194*** | -0,194*** | -0,188*** |
| | (0,0177) | (0,0109) | (0,0109) | (0,0116) | (0,0113) | (0,0114) | (0,0112) |
| Mangler res NP8 regning | -0,141*** | 0,0762*** | 0,0874*** | -0,0190 | -0,00319 | -0,00614 | 0,00581 |
| | (0,0203) | (0,0126) | (0,0125) | (0,0133) | (0,0131) | (0,0131) | (0,0129) |
| Mangler res NP8 lesing … | -0,232*** | -0,108*** | -0,104*** | -0,187*** | -0,176*** | -0,178*** | -0,170*** |
| | (0,0167) | (0,0103) | (0,0103) | (0,0110) | (0,0107) | (0,0108) | (0,0106) |
| Mangler observasjon fra tidligere år …… | -0,459*** | 0,0583 | 0,0771 | 0,916 | 0,856 | -0,0993*** | -0,108*** |
| | (0,0181) | (0,481) | (0,479) | (0,509) | (0,498) | (0,0247) | (0,0246) |
| Jente …… | 0,0587*** | 0,0171*** | 0,0173*** | 0,0269*** | 0,0303*** | 0,0241*** | 0,0276*** |
| | (0,00641) | (0,00397) | (0,00396) | (0,00421) | (0,00412) | (0,00421) | (0,00413) |
| Mangler resultat NP5 | | 0,0234 | -0,000155 | -1,631** | -1,516** | | |
| | | (0,481) | (0,479) | (0,509) | (0,498) | | |
| Gjennomsnitt NP5 | | | | 0,829*** | 0,786*** | | |
| | | | | (0,00265) | (0,00272) | | |
| Poeng engelsk NP5 | | | | | | 0,201*** | 0,200*** |
| | | | | | | (0,00281) | (0,00276) |
| Poeng regning NP5 | | | | | | 0,338*** | 0,318*** |
| | | | | | | (0,00291) | (0,00289) |
| Poeng lesing NP5 | | | | | | 0,309*** | 0,291*** |
| | | | | | | (0,00315) | (0,00312) |
| Mangler res NP8 engelsk | | | | | | -0,174*** | -0,161*** |
| | | | | | | (0,0166) | (0,0163) |
| Mangler res NP8 regning | | | | | | -0,207*** | -0,187*** |
| | | | | | | (0,0150) | (0,0147) |
| Mangler resultat NP8 lesing | | | | | | -0,258*** | -0,241*** |
| | | | | | | (0,0145) | (0,0143) |
| Konstantledd …… | 2,853*** | -0,0625*** | -0,0635*** | 2,867*** | 2,861*** | 2,885*** | 2,878*** |
| | (0,00470) | (0,00289) | (0,00290) | (0,00306) | (0,00302) | (0,00309) | (0,00306) |
| Familiebakgrunnsvar inkl. …… | Ja | Nei | Ja | Nei | Ja | Nei | Ja |
| Forklaringskraft ($R^2$) …… | 0,175 | 0,00452 | 0,0133 | 0,645 | 0,661 | 0,656 | 0,669 |
| Antall elever …… | 60 696 | 60 696 | 60 696 | 60 696 | 60 696 | 60 696 | 60 696 |

Estimerte standardfeil i parentes. Statistisk signifikans: * p<0,05, ** p<0,01, *** p<0,001

Tabell 6.2 viser beskrivende statistikk for de ulike skolebidragsindikatorene tilhørende kolonnene i Tabell 6.1, i tillegg til ujustert gjennomsnitt. Verdiene i tabellen er veid med antall elever ved skolene. Bare indikatorer for skoler som har minst 20 elever med i datamaterialet rapporteres, hvilket medfører at gjennomsnittet ikke er presist lik null[7]. Som vi ser er spredningen, målt ved standavviket, for alle

[7] Årsaken til at vi bare rapporterer indikatorer for skoler med mer enn 20 elevobservasjoner, er at indikatorer for skoler med færre observasjoner er mer upresist estimert. Hægeland et al. (2004) gir en drøfting av dette. På bakgrunn av denne drøftingen ble det besluttet at skolebidragsindikatorene som ble publisert på www.skoleporten.no i 2005 (se Hægeland et al, 2005a) baserte seg på to elevkull, og begrenset seg til skoler med minst 20 elevobservasjoner hvert av årene. For arbeidet med denne rapporten har vi bare tilgang til ett elevkull, men vi beholder likevel kravet om minst 20 elevobservasjoner.

indikatorene betydelig lavere enn for de ujusterte resultatene. Unntaket er de mest ekstreme observasjonene som blir enda mer ekstreme, uten at det nødvendigvis er de samme skolene. Mens 90-10 forskjellen for ujusterte skolegjennomsnitt ligger rundt 0.75 standardavvik faller den til rundt 0.45 når vi kontrollerer for resultater fra 5.trinn.

**Tabell 6.2. Beskrivende statistikk for indikatorer basert på gjennomsnittsresultat nasjonale prøver 8. trinn (elevvektet)**

| | Gj.snitt | Std. Avvik | Min. | 10. per-sentil | 25. per-sentil | 50. per-sentil | 75. per-sentil | 90. per-sentil | Maks. |
|---|---|---|---|---|---|---|---|---|---|
| Ujustert gjennomsnitt ..... | 0,014 | 0,293 | -0,875 | -0,343 | -0,184 | -0,015 | 0,197 | 0,413 | 1,066 |
| Kontroll for familiebakgrunn (1) ..... | -0,001 | 0,210 | -0,704 | -0,273 | -0,149 | 0,006 | 0,135 | 0,263 | 1,062 |
| Diff. NP8-NP5 (2) ..... | -0,003 | 0,179 | -1,183 | -0,224 | -0,118 | 0,004 | 0,117 | 0,215 | 0,583 |
| Diff. NP8-NP5, fam.bak.(3) .. | -0,005 | 0,177 | -1,147 | -0,224 | -0,122 | -0,001 | 0,111 | 0,218 | 0,593 |
| Kontroll NP5 (4) ..... | -0,002 | 0,196 | -1,089 | -0,244 | -0,126 | -0,001 | 0,110 | 0,224 | 1,301 |
| Kontroll NP5, fam.bak. (5) .. | -0,006 | 0,180 | -0,980 | -0,232 | -0,121 | -0,003 | 0,098 | 0,200 | 1,144 |
| Kontroll alle NP5 (6) ..... | -0,004 | 0,197 | -1,109 | -0,241 | -0,124 | -0,004 | 0,108 | 0,224 | 1,358 |
| Kontroll alle NP5, fam.bak. (7) ..... | -0,007 | 0,183 | -1,005 | -0,236 | -0,120 | -0,005 | 0,097 | 0,202 | 1,205 |
| Antall skoler ..... | 1 211 | | | | | | | | |

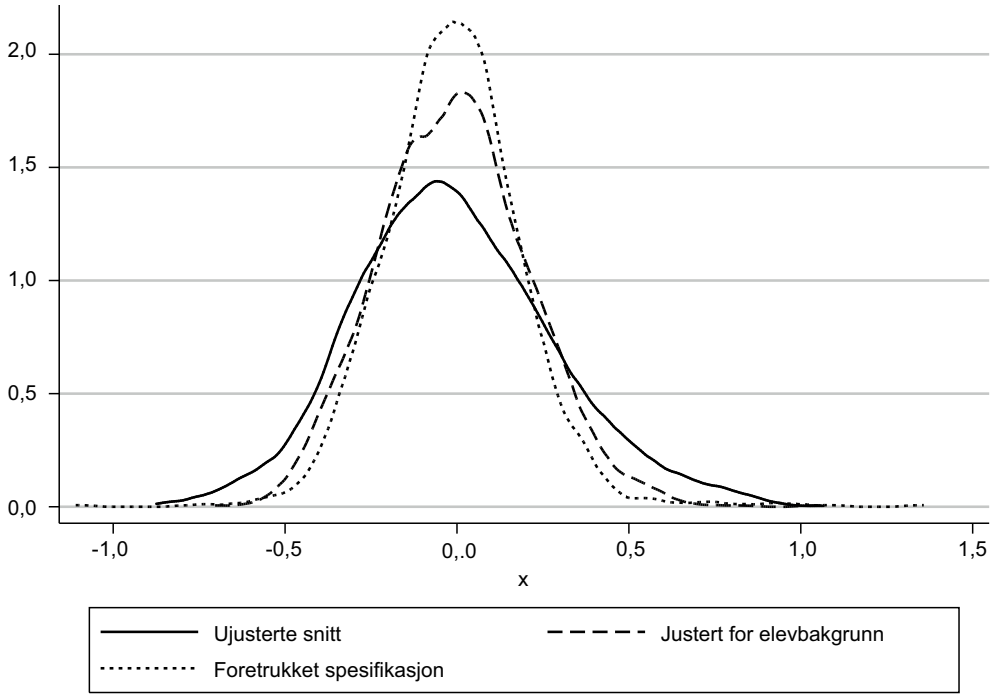
Et nærliggende spørsmål er om 0,45 av et standardavvik representerer store forskjeller mellom de beste og dårligste skolene? En mulig sammenlikning er elever fra ulike familier der foreldrene har ulikt utdanningsnivå. I de modellene hvor vi kontrollerer for familiebakgrunn, har vi også med variable for foreldrenes utdanning. Ved å sammenligne koeffisienter for ulike utdanningsnivåer, finner vi forskjeller i resultater mellom elever som har foreldre med ulikt utdanningsnivå, betinget på de andre variablene som er inkludert i modellen. Går vi nærmere inn på estimeringene som ligger bak kolonne (7) i Tabell 6.1 (disse regresjonskoeffisientene er ikke rapportert i denne rapporten), finner vi en forskjell på 0,40 mellom en elev med begge foreldrene med lang universitetsutdanning, sammenliknet med en annen der begge har grunnskole. Det innebærer at hensyn tatt til forskjell i resultater fra 5.trinn gjør elevene med høyt utdannede foreldre bedre ved de nasjonale prøvene på 8. trinn. Dersom man tolker skolebidragsindikatorene som uttrykk for skolens bidrag, og forskjellen knyttet til foreldres utdanning, så sier tallene at det å gå på en av de beste skolene kontra en av de dårligste, kan være minst like viktig for læringen på mellomtrinnet som det å ha høyt kontra lavt utdannede foreldre.

Alle indikatorene som kontrollerer for tidligere resultater har lavere spredning enn der vi kun kontrollerer for forskjeller i familiebakgrunn. Likevel fanger familiebakgrunn opp en betydelig andel av variasjonen mellom skoler. Mens Tabell 6.1 viser at elev-kjennetegn som tidligere prestasjoner er en viktig forklaringsfaktor for elevresultatene, viser Tabell 6.2 at bildet av forskjeller mellom skoler blir noe annerledes når man korrigerer resultater for forskjeller i elevsammensetning. Figur 6.1 viser fordelingen til ulike resultatmål. Den heltrukne grafen viser fordelingen til ujusterte resultater, den stiplede til indikatoren som følger av å kun kontrollere for familiebakgrunn (1), mens den prikkede viser fordelingen til vår foretrukne indikator som kontrollerer for alle resultater fra NP 5. trinn. Som vi ser er spredningen mindre enn for ujusterte resultater når man kontrollerer for familiebakgrunn, og enda mindre når man kontrollerer for tidligere resultater. Dersom elevene var tilfeldig fordelt på skoler, ville spredningen i resultatene/ indikatorene ha vært uavhengig av hva man kontrollere for. Selv om spredningen blir lavere, er det likevel er det fremdeles betydelige forskjeller mellom skolene som skårer blant de høyeste og de som skårer blant de laveste.

Selv om Tabell 6.2 viser at det er relativt beskjedne forskjeller i spredning mellom de ulike mulige indikatorene, gir den ikke noe svar på hvordan *samvariasjonen* er mellom de ulike indikatorene, dvs. i hvilken grad de gir forskjellige svar på hvilke skoler som bidrar mye eller lite til elevenes læring. Tabell 6.3 viser korrelasjons-matrisen for de ulike indikatorene. (Korrelasjonskoeffiesienten uttrykker graden av lineær samvariasjon mellom to variable. Den kan variere mellom -1 og 1, der -1 uttrykker at det er en perfekt negativ lineær sammenheng, og 1 at det er en tilsvarende positiv sammenheng). Som vi ser av tabellen, er det en meget høy

korrelasjon (over 0,85) mellom alle indikatorer hvor man kontrollerer for tidligere resultater (markert med grått). Vi ser også at gitt hvordan man kontrollerer for tidligere resultater, så har det liten betydning for indikatorene om man i tillegg kontrollerer for familiebakgrunn: Korrelasjonene mellom (2) og (3), mellom (4) og (5) og mellom (6) og (7) er alle rundt 0,97-0,98. Det innebærer at det gir lite ny informasjon med hensyn til skolens bidrag å kontrollere for familiebakgrunn i tillegg til tidligere resultater.

**Figur 6.1. Fordeling av ulike resultatmål basert på gjennomsnittskår nasjonale prøver 8. trinn**



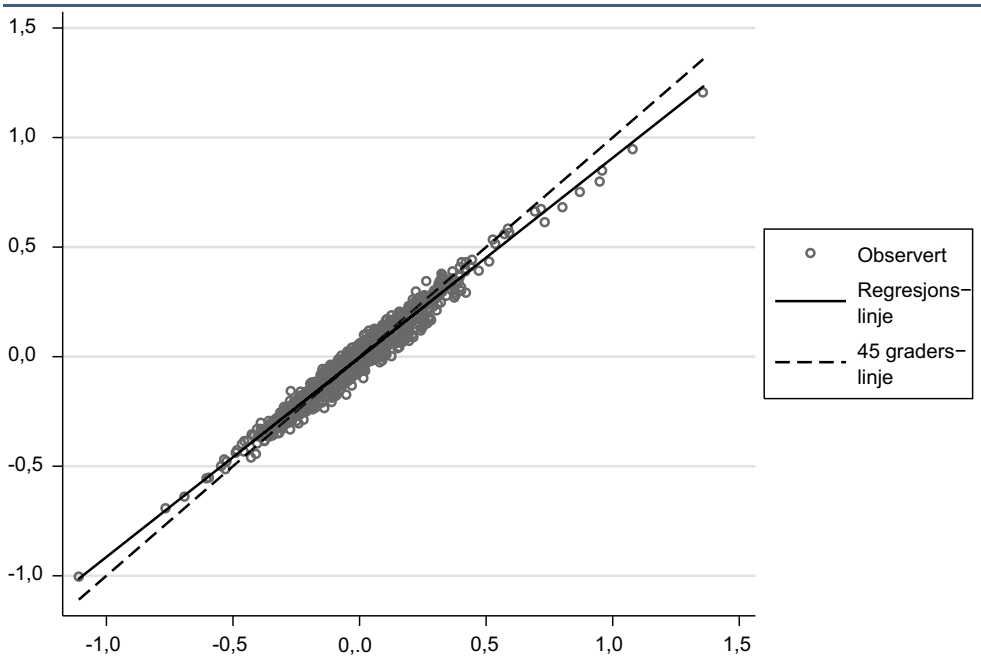

**Tabell 6.3. Korrelasjon mellom ulike indikatorer basert på gjennomsnittskår nasjonale prøver 8. trinn**

| | Ujust snitt | Snitt, FB | NP8-NP5 | NP8-NP5, FB | Kont NP5 | Kont NP5, FB | Kont alle NP5 | Kont alle NP5, FB |
|---|---|---|---|---|---|---|---|---|
| Ujustert gjennomsnitt ..... | 1 | | | | | | | |
| Kontroll for familiebakgrunn (1) ..... | 0,851 | 1 | | | | | | |
| Diff. NP8-NP5 (2) ..... | 0,346 | 0,391 | 1 | | | | | |
| Diff. NP8-NP5, fam.bak.(3) .... | 0,268 | 0,353 | 0,994 | 1 | | | | |
| Kontroll NP5 (4) ..... | 0,587 | 0,640 | 0,862 | 0,830 | 1 | | | |
| Kontroll NP5, fam.bak. (5) ..... | 0,478 | 0,634 | 0,881 | 0,872 | 0,975 | 1 | | |
| Kontroll alle NP5 (6) ..... | 0,567 | 0,631 | 0,839 | 0,809 | 0,979 | 0,958 | 1 | |
| Kontroll alle NP5, fam.bak. (7) | 0,468 | 0,623 | 0,855 | 0,846 | 0,956 | 0,980 | 0,980 | 1 |
| Antall skoler ..... | 1 211 | | | | | | | |

Så langt har vi sett at det har relativt stor betydning å kontrollere for tidligere resultater, og at det ikke er likegyldig hvordan man kontrollerer for disse. Men det er ikke opplagt hvilken spesifikasjon man bør velge. Siden de tidligere resultater i de ulike fagene på 8. trinn har til dels forskjellig effekt på resultater tre år senere (jfr. Tabell 6.1) foretrekker vi alternativet med den frieste spesifikasjonen der vi inkluderer resultatene fra de ulike emnene i NP 5. trinn separat. Denne spesifikasjonen hviler på færre teoretiske forutsetninger, og utnytter også informasjonen om tidligere prestasjoner bedre. Spørsmålet som gjenstår, er hvorvidt det er viktig å kontrollere for familiebakgrunn i tillegg. Den høye korrelasjonen mellom (6) og (7) ovenfor indikerer at dette ikke er viktig i det store bildet, verken for gjennomsnittsskolen eller vurderingen av den samlete variasjonen mellom skoler. Men selv om korrelasjonen er høy kan det ha stor betydning for enkeltskolers indikatorer om man gjør dette eller ikke. I Figur 6.2 ser vi at for de aller fleste skoler har det svært liten betydning. Dette trekker i retning av å benytte en beregningsmodell der man ikke kontrollerer for familiebakgrunn. Dette gjør modellen noe enklere, samt at

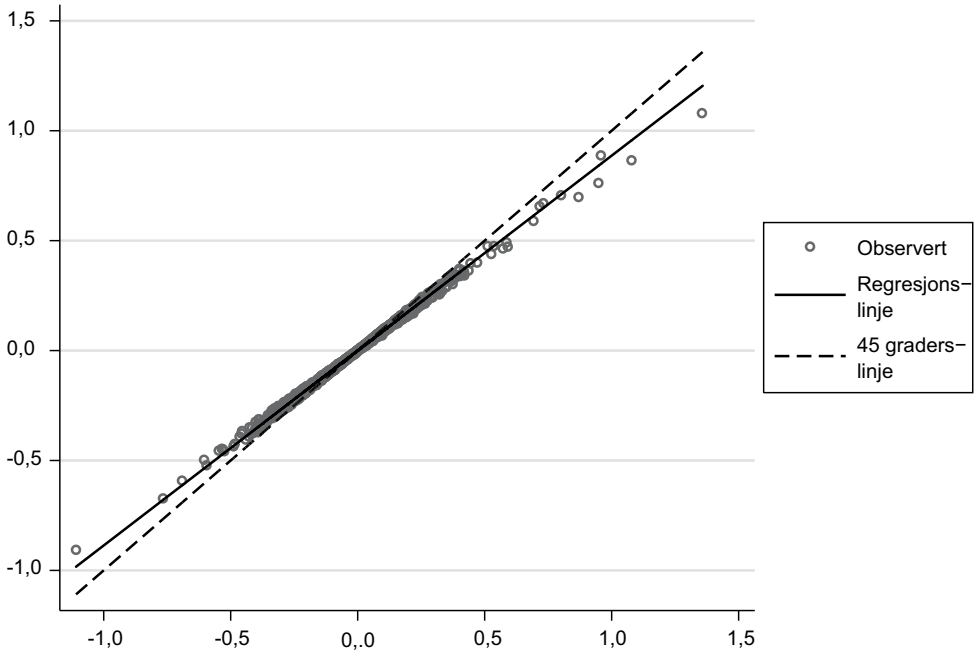
arbeidet med tilrettelegging av data blir vesentlig mindre. Indikatorene kan da beregnes med utgangspunkt i data utdanningsmyndighetene selv rår over. Figur 6.3 og Figur 6.4 viser at det har minimal betydning for indikatorene om vi spesifiserer skole-effekten som en fast effekt eller som en tilfeldig effekt, eller om vi ikke pålegger noen sammenhenger mellom elever fra samme skole, men bare regner ut skoleeffekten på bakgrunn av residualer basert på estimering ved hjelp av minste kvadraters metode.

**Figur 6.2. Sammenheng mellom indikatorer med og uten korreksjon for familiebakgrunn, nasjonale prøver 8. trinn**




1211 skoler med 50809 elever. Elevvektet reg.koeff: 0.910; korrelasjon=0.980; rangkorr.=0.976

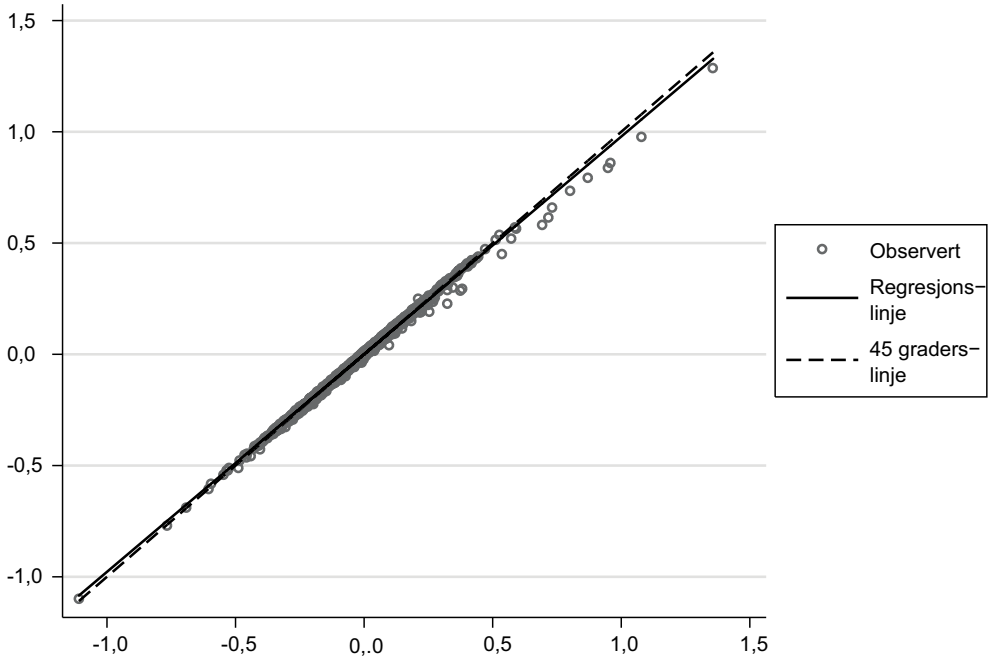
**Figur 6.3. Sammenheng mellom indikatorer basert på tilfeldig effekt ("random effect") og fast effekt ("fixed effect"), nasjonale prøver 8. trinn**




1211 skoler med 50809 elever. Elevvektet reg.koeff: 0.886; korrelasjon=0.999; rangkorr.=1.000

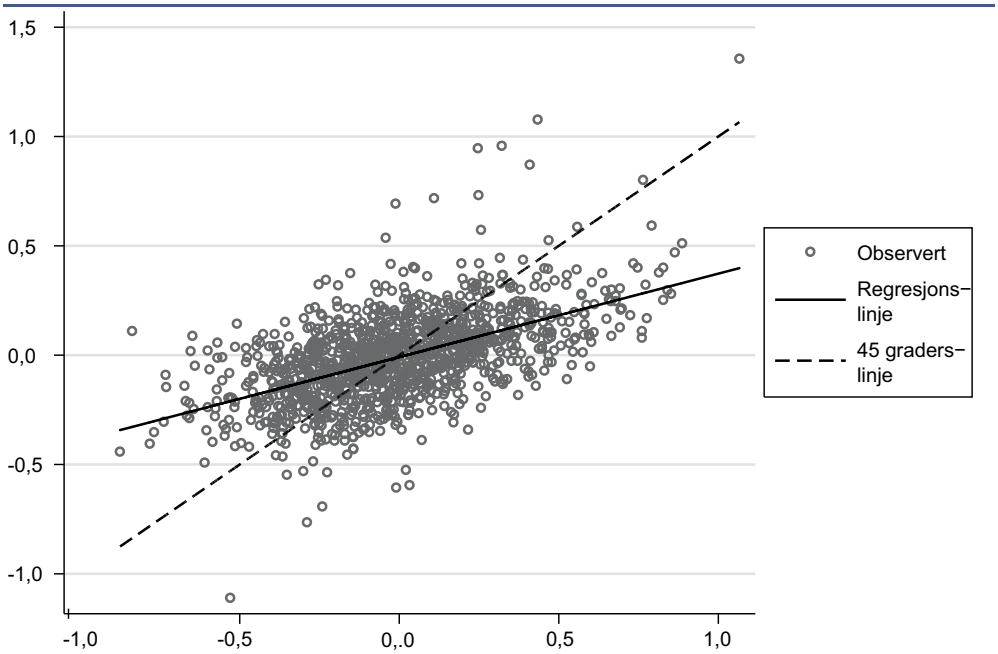
Avslutningsvis i dette avsnittet illustrerer Figur 6.5 og Figur 6.6 at det å kontrollere for tidligere resultater fører til et endret bilde av forskjeller mellom skoler. Dessuten skjer det endringer i hvilke skoler som bidrar mye og lite til elevenes læring. På den ene side finner vi at skoler som skårer høyt på nasjonale prøver på 8.trinn også tenderer til å skåre høyt når vi kontrollerer for tidligere resultater. Men det er også slik at det å kontrollere for tidligere resultater for mange skoler gir et vesentlig annet bilde av skolens bidrag enn både ujusterte resultater og resultater korrigert for forskjeller i familiebakgrunn. Mange punkter i Figur 6.6 ligger langt unna 45 graders linjen.

**Figur 6.4. Sammenheng mellom indikatorer basert på "OLS-residualer" og "fixed effect", nasjonale prøver 8. trinn**



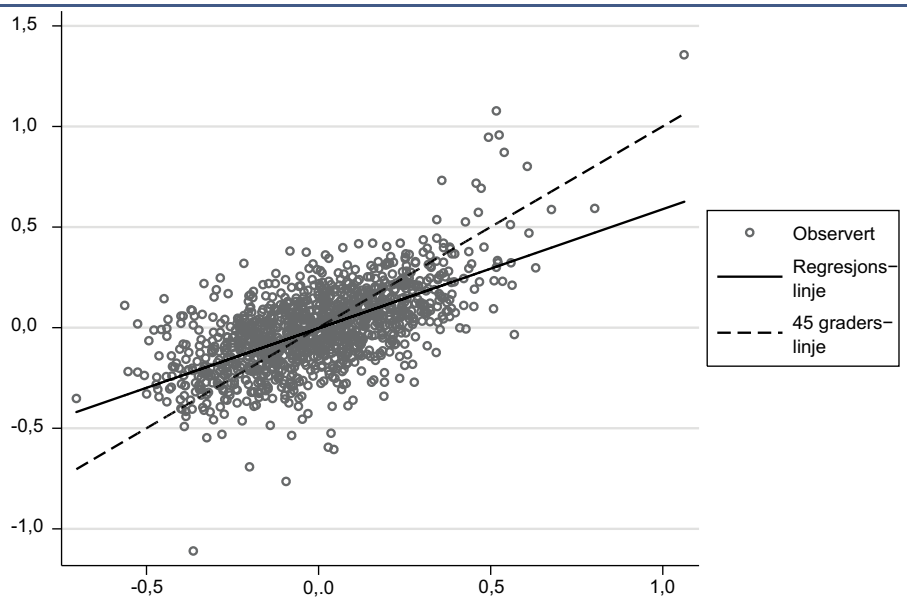

1211 skoler med 50809 elever. Elevvektet reg.koeff: 0.978; korrelasjon=0.998; rangkorr.=0.999

**Figur 6.5. Sammenheng mellom foretrukket indikator og ujustert resultat, nasjonale prøver 8. trinn**




1211 skoler med 50809 elever. Elevvektet reg.koeff: 0.381; korrelasjon=0.567; rangkorr.=0.563

**Figur 6.6. Sammenheng mellom foretrukket indikator og indikator bare med korreksjon for familiebakgrunn, nasjonale prøver 8. trinn**




1211 skoler med 50809 elever. Elevvektet reg.koeff: 0.591; korrelasjon=0.631; rangkorr.=0.587

## 6.2. Indikatorer for enkeltfag

I dette avsnittet gjentar vi store deler av analysen fra forrige avsnitt, men nå for lesing, regning og engelsk separat. Tabell 6.4, Tabell 6.5 og Tabell 6.6 gir regresjonsresultatene for de ulike modellspesifikasjonene. Mønstrene er om lag de samme som de vi så i Tabell 6.1, men modellenes forklaringskraft er jevnt over noe lavere. Vi merker oss også at i de tilfellene der vi kontrollerer for tidligere resultater i alle fag separat, er sammenhengen klart sterkest mellom nåværende og tidligere resultat innenfor samme fag. Likevel har tidligere resultater i andre fag også har betydning selv når vi betinger på tidligere resultater innen samme fag. Et annet fellestrekk er at koeffisienten for resultatet på 5.trinn er langt fra 1, hvilket igjen forteller oss at value added-modellen med full persistens (λ=1) i kolonne (2)-(3) representerer en feilspesifikasjon, og ikke bør anvendes.

**Tabell 6.4. Regresjonsresultater, basert på resultat nasjonale prøver 8. trinn, lesing**

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| | Lesing NP8 | Lesing NP8-Lesing NP5 | Lesing NP8-Lesing NP5 | Lesing NP8 | Lesing NP8 | Lesing NP8 | Lesing NP8 |
| Mangler observasjon fra tidligere år | -0,590*** | 0,0113 | -0,0131 | -0,323*** | -0,265*** | -0,134*** | -0,112** |
| | (0,0225) | (0,0257) | (0,0263) | (0,0255) | (0,0257) | (0,0345) | (0,0344) |
| Jente | 0,241*** | 0,0299*** | 0,0295*** | 0,106*** | 0,118*** | 0,177*** | 0,181*** |
| | (0,00761) | (0,00603) | (0,00602) | (0,00601) | (0,00590) | (0,00571) | (0,00564) |
| Mangler resultat NP5 lesing | | -0,464*** | -0,465*** | -0,555*** | -0,503*** | -0,366*** | -0,341*** |
| | | (0,0206) | (0,0206) | (0,0204) | (0,0201) | (0,0201) | (0,0199) |
| Poeng lesing NP5 | | | | 0,684*** | 0,639*** | 0,433*** | 0,412*** |
| | | | | (0,00325) | (0,00333) | (0,00426) | (0,00425) |
| Poeng engelsk NP5 | | | | | | 0,168*** | 0,169*** |
| | | | | | | (0,00380) | (0,00377) |
| Poeng regning NP5 | | | | | | 0,247*** | 0,224*** |
| | | | | | | (0,00394) | (0,00394) |
| Mangler resultat NP5 engelsk | | | | | | -0,192*** | -0,175*** |
| | | | | | | (0,0228) | (0,0225) |
| Mangler resultat NP5 regning | | | | | | -0,183*** | -0,159*** |
| | | | | | | (0,0206) | (0,0203) |
| Konstantledd | 2,935*** | 0,437*** | 0,438*** | 3,027*** | 3,012*** | 2,993*** | 2,984*** |
| | (0,00540) | (0,00429) | (0,00432) | (0,00427) | (0,00422) | (0,00408) | (0,00406) |
| Familiebakgrunnsvar inkl. | Ja | Nei | Ja | Nei | Ja | Nei | Ja |
| Forklaringskraft ($R^2$) | 0,151 | 0,0228 | 0,0255 | 0,475 | 0,496 | 0,537 | 0,549 |
| Antall elever | 56 982 | 56 982 | 56 982 | 56 982 | 56 982 | 56 982 | 56 982 |

Estimerte standardfeil i parentes. Statistisk signifikans: * p<0,05, ** p<0,01, *** p<0,001

**Tabell 6.5. Regresjonsresultater, basert på resultat nasjonale prøver 8. trinn, regning**

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| | Regning NP8 | Regning NP8 -Regning NP5 | Regning NP8 -Regning NP5 | Regning NP8 | Regning NP8 | Regning NP8 | Regning NP8 |
| Mangler observasjon fra tidligere år | -0,365*** | 0,00566 | -0,0413 | -0,146*** | -0,157*** | -0,0395 | -0,0738* |
| | (0,0216) | (0,0222) | (0,0229) | (0,0230) | (0,0234) | (0,0320) | (0,0319) |
| Jente | -0,127*** | -0,0504*** | -0,0511*** | 0,0709*** | 0,0725*** | -0,105*** | -0,102*** |
| | (0,00748) | (0,00515) | (0,00513) | (0,00534) | (0,00523) | (0,00533) | (0,00524) |
| Mangler resultat NP5 regning | | 0,277*** | 0,280*** | -0,445*** | -0,411*** | -0,378*** | -0,356*** |
| | | (0,0184) | (0,0183) | (0,0190) | (0,0186) | (0,0194) | (0,0191) |
| Poeng regning NP5 | | | | 0,768*** | 0,727*** | 0,665*** | 0,642*** |
| | | | | (0,00291) | (0,00299) | (0,00368) | (0,00367) |
| Poeng engelsk NP5 | | | | | | 0,0421*** | 0,0386*** |
| | | | | | | (0,00356) | (0,00351) |
| Poeng lesing NP5 | | | | | | 0,133*** | 0,114*** |
| | | | | | | (0,00399) | (0,00396) |
| Mangler resultat NP5 engelsk | | | | | | -0,0630** | -0,0506* |
| | | | | | | (0,0214) | (0,0211) |
| Mangler resultat NP5 lesing | | | | | | -0,117*** | -0,104*** |
| | | | | | | (0,0186) | (0,0183) |
| Konstantledd | 2,720*** | -0,255*** | -0,253*** | 2,709*** | 2,706*** | 2,728*** | 2,723*** |
| | (0,00530) | (0,00366) | (0,00367) | (0,00379) | (0,00374) | (0,00380) | (0,00377) |
| Familiebakgrunnsvar inkl. | Ja | Nei | Ja | Nei | Ja | Nei | Ja |
| Forklaringskraft ($R^2$) | 0,138 | 0,0134 | 0,0209 | 0,562 | 0,580 | 0,576 | 0,591 |
| Antall elever | 59 047 | 59 047 | 59 047 | 59 047 | 59 047 | 59 047 | 59 047 |

Estimerte standardfeil i parentes. Statistisk signifikans: * p<0,05, ** p<0,01, *** p<0,001

**Tabell 6.6. Regresjonsresultater, basert på resultat nasjonale prøver 8. trinn, engelsk**

| | Engelsk NP8 | Engelsk NP8 -Engelsk NP5 | Engelsk NP8 -Engelsk NP5 | Engelsk NP8 | Engelsk NP8 | Engelsk NP8 | Engelsk NP8 |
|---|---|---|---|---|---|---|---|
| Mangler observasjon fra tidligere år | -0,488*** | 0,0295 | 0,0564 | -0,193*** | -0,115*** | -0,105** | -0,103** |
| | (0,0228) | (0,0303) | (0,0311) | (0,0297) | (0,0299) | (0,0344) | (0,0347) |
| Jente | 0,0703*** | 0,0694*** | 0,0709*** | 0,0710*** | 0,0740*** | 0,00798 | 0,0104 |
| | (0,00775) | (0,00632) | (0,00630) | (0,00619) | (0,00606) | (0,00564) | (0,00561) |
| Mangler resultat NP5 engelsk | | 0,306*** | 0,327*** | -0,491*** | -0,434*** | -0,276*** | -0,264*** |
| | | (0,0259) | (0,0259) | (0,0254) | (0,0249) | (0,0230) | (0,0229) |
| Poeng engelsk NP5 | | | | 0,659*** | 0,628*** | 0,396*** | 0,394*** |
| | | | | (0,00334) | (0,00334) | (0,00376) | (0,00374) |
| Poeng regning NP5 | | | | | | 0,105*** | 0,0901*** |
| | | | | | | (0,00389) | (0,00392) |
| Poeng lesing NP5 | | | | | | 0,369*** | 0,356*** |
| | | | | | | (0,00422) | (0,00424) |
| Mangler resultat NP5 regning | | | | | | -0,0441* | -0,0291 |
| | | | | | | (0,0206) | (0,0204) |
| Mangler resultat NP5 lesing | | | | | | -0,272*** | -0,255*** |
| | | | | | | (0,0199) | (0,0198) |
| Konstantledd | 2,849*** | -0,360*** | -0,364*** | 2,864*** | 2,852*** | 2,895*** | 2,890*** |
| | (0,00549) | (0,00449) | (0,00450) | (0,00439) | (0,00433) | (0,00403) | (0,00403) |
| Familiebakgrunnsvar inkl. | Ja | Nei | Ja | Nei | Ja | Nei | Ja |
| Forklaringskraft ($R^2$) | 0,0977 | 0,0112 | 0,0197 | 0,423 | 0,449 | 0,537 | 0,543 |
| Antall elever | 58 378 | 58 378 | 58 378 | 58 378 | 58 378 | 58 378 | 58 378 |

Estimerte standardfeil i parentes. Statistisk signifikans: * p<0,05, ** p<0,01, *** p<0,001

Tabell 6.7, Tabell 6.8 og Tabell 6.9 gjengir beskrivende statistikk for indikatorene i de ulike fagene. Som for gjennomsnittet for alle tre emnene er spredningen mellom skoler er betydelig mindre når vi ser på de ulike skolebidragsindikatorene, sammenliknet med det ukorrigerte gjennomsnittet. Det er små forskjeller i standardavviket for en gitt indikator på tvers av fag. Tabell 6.10, Tabell 6.11 og Tabell 6.12 viser hvordan indikatorer fra ulike spesifikasjoner er korrelert. Igjen

finner vi at det er høy korrelasjon mellom alle indikatorer hvor man kontrollerer for tidligere resultater, men at det har en viss betydning *hvordan* vi kontrollerer for disse. Gitt hvordan man kontrollerer for tidligere resultater, så har det liten betydning for indikatorene om man i tillegg kontrollerer for familiebakgrunn. Konklusjonene fra forrige avsnitt om valg av foretrukket modell står seg i stor grad også når vi ser på enkeltfag.

**Tabell 6.7. Beskrivende statistikk indikatorer basert på resultat, nasjonale prøver 8. trinn, lesing (elevvektet)**

| | Gj.snitt | Std. avvik | Min. | 10. per-entil | 25. per-sentil | 50. per-sentil | 75. per-sentil | 90. per-sentil | Maks. |
|---|---|---|---|---|---|---|---|---|---|
| Ujustert gjennomsnitt ..... | 0,014 | 0,309 | -0,868 | -0,358 | -0,200 | 0,002 | 0,195 | 0,422 | 1,016 |
| Kontroll for familiebakgrunn | 0,001 | 0,232 | -0,819 | -0,282 | -0,164 | 0,001 | 0,151 | 0,311 | 0,831 |
| Diff. NP8-NP5 ..... | -0,005 | 0,209 | -0,780 | -0,270 | -0,129 | 0,003 | 0,133 | 0,259 | 0,765 |
| Diff. NP8-NP5, fam.bak. .... | -0,007 | 0,209 | -0,769 | -0,273 | -0,133 | -0,002 | 0,126 | 0,252 | 0,782 |
| Kontroll NP5 ..... | -0,001 | 0,226 | -0,763 | -0,268 | -0,148 | -0,005 | 0,136 | 0,271 | 1,329 |
| Kontroll NP5, fam.bak. ..... | -0,006 | 0,205 | -0,716 | -0,256 | -0,139 | -0,007 | 0,120 | 0,246 | 1,102 |
| Kontroll alle NP5 ..... | -0,003 | 0,222 | -0,840 | -0,269 | -0,148 | -0,004 | 0,128 | 0,271 | 1,325 |
| Kontroll alle NP5, fam.bak. | -0,006 | 0,208 | -0,731 | -0,256 | -0,139 | -0,009 | 0,116 | 0,260 | 1,138 |
| Antall skoler ..... | 1 155 | | | | | | | | |

**Tabell 6.8. Beskrivende statistikk indikatorer basert på resultat, nasjonale prøver 8. trinn, regning (elevvektet)**

| | Gj.snitt | Std. avvik | Min. | 10. per-entil | 25. per-sentil | 50. per-sentil | 75. per-sentil | 90. per-sentil | Maks. |
|---|---|---|---|---|---|---|---|---|---|
| Ujustert gjennomsnitt ..... | 0,011 | 0,327 | -0,928 | -0,401 | -0,213 | -0,008 | 0,222 | 0,467 | 1,022 |
| Kontroll for familiebakgrunn | -0,005 | 0,253 | -0,769 | -0,324 | -0,181 | -0,008 | 0,167 | 0,315 | 0,993 |
| Diff. NP8-NP5 ..... | -0,008 | 0,234 | -1,342 | -0,310 | -0,152 | -0,007 | 0,149 | 0,291 | 1,003 |
| Diff. NP8-NP5, fam.bak. .... | -0,011 | 0,232 | -1,311 | -0,309 | -0,159 | -0,007 | 0,143 | 0,284 | 1,016 |
| Kontroll NP5 ..... | -0,005 | 0,244 | -1,187 | -0,299 | -0,157 | -0,012 | 0,147 | 0,314 | 1,214 |
| Kontroll NP5, fam.bak. ..... | -0,011 | 0,229 | -1,051 | -0,295 | -0,158 | -0,017 | 0,129 | 0,282 | 1,050 |
| Kontroll alle NP5 ..... | -0,006 | 0,238 | -1,209 | -0,301 | -0,150 | -0,012 | 0,139 | 0,306 | 1,096 |
| Kontroll alle NP5, fam.bak. | -0,012 | 0,227 | -1,081 | -0,297 | -0,155 | -0,016 | 0,131 | 0,284 | 0,968 |
| Antall skoler ..... | 1185 | | | | | | | | |

**Tabell 6.9. Beskrivende statistikk indikatorer basert på resultat nasjonale prøver 8. trinn, engelsk (elevvektet)**

| | Gj.snitt | Std. avvik | Min. | 10. per-entil | 25. per-sentil | 50. per-sentil | 75. per-sentil | 90. per-sentil | Maks. |
|---|---|---|---|---|---|---|---|---|---|
| Ujustert gjennomsnitt ..... | 0,016 | 0,304 | -1,107 | -0,362 | -0,191 | -0,003 | 0,213 | 0,425 | 0,992 |
| Kontroll for familiebakgrunn | 0,004 | 0,239 | -0,863 | -0,312 | -0,162 | 0,013 | 0,167 | 0,300 | 1,023 |
| Diff. NP8-NP5 ..... | 0,005 | 0,248 | -1,314 | -0,298 | -0,135 | 0,011 | 0,156 | 0,309 | 0,927 |
| Diff. NP8-NP5, fam.bak. .... | 0,003 | 0,244 | -1,278 | -0,287 | -0,132 | 0,007 | 0,154 | 0,303 | 0,892 |
| Kontroll NP5 ..... | 0,007 | 0,246 | -1,202 | -0,302 | -0,146 | 0,001 | 0,155 | 0,313 | 1,036 |
| Kontroll NP5, fam.bak. ..... | 0,002 | 0,218 | -1,087 | -0,268 | -0,137 | -0,001 | 0,144 | 0,270 | 0,888 |
| Kontroll alle NP5 ..... | 0,001 | 0,218 | -1,390 | -0,268 | -0,135 | 0,005 | 0,142 | 0,255 | 1,086 |
| Kontroll alle NP5, fam.bak. | -0,001 | 0,208 | -1,314 | -0,256 | -0,131 | -0,002 | 0,127 | 0,249 | 0,978 |
| Antall skoler ..... | 1173 | | | | | | | | |

**Tabell 6.10. Korrelasjon mellom ulike indikatorer, nasjonale prøver 8. trinn, lesing**

| | Ujust snitt | Ujust snitt, FB | NP8-NP5 | NP8-NP5, FB | Kont NP5 | Kont NP5, FB | Kont alle NP5 | Kont alle NP5, FB |
|---|---|---|---|---|---|---|---|---|
| Ujustert gjennomsnitt ..... | 1 | | | | | | | |
| Kontroll for familiebakgrunn . | 0,851 | 1 | | | | | | |
| Diff. NP8-NP5 ..... | 0,341 | 0,445 | 1 | | | | | |
| Diff. NP8-NP5, fam.bak. ..... | 0,309 | 0,423 | 0,997 | 1 | | | | |
| Kontroll NP5 ..... | 0,713 | 0,762 | 0,803 | 0,787 | 1 | | | |
| Kontroll NP5, fam.bak. ..... | 0,597 | 0,771 | 0,851 | 0,844 | 0,965 | 1 | | |
| Kontroll alle NP5 ..... | 0,621 | 0,694 | 0,794 | 0,784 | 0,954 | 0,936 | 1 | |
| Kontroll alle NP5, fam.bak. .. | 0,526 | 0,699 | 0,824 | 0,820 | 0,918 | 0,956 | 0,978 | 1 |
| Antall skoler ..... | 1 155 | | | | | | | |

**Tabell 6.11. Korrelasjon mellom ulike indikatorer, nasjonale prøver 8. trinn, regning**

| | Ujust snitt | Ujust snitt, FB | NP8-NP5 | NP8-NP5, FB | Kont NP5 | Kont NP5, FB | Kont alle NP5 | Kont alle NP5, FB |
|---|---|---|---|---|---|---|---|---|
| Ujustert gjennomsnitt ..... | 1 | | | | | | | |
| Kontroll for familiebakgrunn . | 0,878 | 1 | | | | | | |
| Diff. NP8-NP5 ..... | 0,412 | 0,491 | 1 | | | | | |
| Diff. NP8-NP5, fam.bak. ..... | 0,369 | 0,472 | 0,995 | 1 | | | | |
| Kontroll NP5 ..... | 0,689 | 0,718 | 0,887 | 0,869 | 1 | | | |
| Kontroll NP5, fam.bak. ..... | 0,596 | 0,725 | 0,905 | 0,903 | 0,975 | 1 | | |
| Kontroll alle NP5 ..... | 0,658 | 0,709 | 0,891 | 0,876 | 0,992 | 0,978 | 1 | |
| Kontroll alle NP5, fam.bak. .. | 0,573 | 0,710 | 0,900 | 0,901 | 0,965 | 0,994 | 0,981 | 1 |
| Antall skoler ..... | 1185 | | | | | | | |

**Tabell 6.12. Korrelasjon mellom ulike indikatorer, nasjonale prøver 8. trinn, engelsk**

| | Ujust snitt | Ujust snitt, FB | NP8-NP5 | NP8-NP5, FB | Kont NP5 | Kont NP5, FB | Kont alle NP5 | Kont alle NP5, FB |
|---|---|---|---|---|---|---|---|---|
| Ujustert gjennomsnitt ..... | 1 | | | | | | | |
| Kontroll for familiebakgrunn . | 0,903 | 1 | | | | | | |
| Diff. NP8-NP5 ..... | 0,366 | 0,367 | 1 | | | | | |
| Diff. NP8-NP5, fam.bak. ..... | 0,291 | 0,339 | 0,990 | 1 | | | | |
| Kontroll NP5 ..... | 0,729 | 0,721 | 0,839 | 0,791 | 1 | | | |
| Kontroll NP5, fam.bak. ..... | 0,625 | 0,726 | 0,861 | 0,852 | 0,959 | 1 | | |
| Kontroll alle NP5 ..... | 0,610 | 0,677 | 0,774 | 0,750 | 0,915 | 0,922 | 1 | |
| Kontroll alle NP5, fam.bak. .. | 0,550 | 0,667 | 0,778 | 0,770 | 0,886 | 0,929 | 0,991 | 1 |
| Antall skoler ..... | 1173 | | | | | | | |

## 6.3. Sammenhenger mellom indikatorer på tvers av fag

Tabell 6.13 viser sammen med Figur 6.7 og Figur 6.8 sammenhenger på tvers av fag. Vi ser fra Tabell 6.13 at skoler som gjør det bra på nasjonale prøver tenderer til å gjøre det godt i alle fag, korrelasjonene på tvers av fag er opp mot 0,8. Dette bildet holder seg når vi ser på vår foretrukne indikator, selv om korrelasjonen da går ned til om lag 0,6. På tross av denne relativt høye korrelasjonen viser figurene at inntrykket av en del skoler med hensyn til om de bidrar mye eller lite til elevenes læring i betydelig grad avhenger av hvilket fag man ser på. En slik forskjell på tvers av fag kan skyldes flere ting. Det kan selvsagt skyldes tilfeldig variasjon og "støy" som slår ulikt ut på tvers av fag. Det kan imidlertid også skyldes at når vi snakker om en skoleeffekt, så behøver ikke den nødvendigvis være den samme på tvers av fag. Dersom virkeligheten var slik at det bare var hvilken lærer man hadde som betød noe for skolens bidrag, hver lærer underviste i bare ett fag og lærere var spredt tilfeldig rundt på skoler, skulle korrelasjonen mellom indikatorer på tvers av fag være lik null (forutsatt at man hadde klart å korrigere fullt ut for forskjeller i elevsammensetning). At korrelasjonen er såpass høy som den vi finner her, tyder likevel på at det er meningsfullt å snakke om en felles skole- eller lærerteameffekt, selv om data-situasjonen ennå ikke tillater å undersøke om denne er persistent over tid. Man må imidlertid også holde muligheten åpen for at en høy korrelasjon reflekterer elevfaktorer som virker på samme måte på tvers av fag, og som man ikke har kontrollert fullgodt for. At korrelasjonen på tvers av fag ikke påvirkes nevneverdig av hvor detaljert vi kontrollerer for familiebakgrunn, er likevel en indikasjon på at ulik uobservert elevsammensetning ikke påvirker resultatene i stor grad.

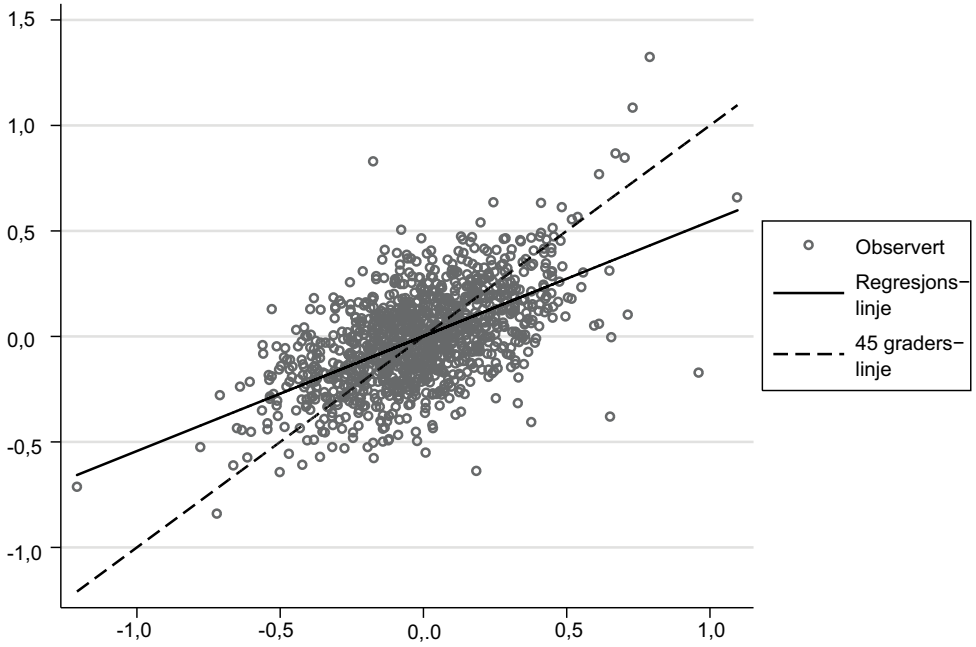
**Tabell 6.13. Korrelasjoner på tvers av fag – prøveresultater nasjonale prøver 8. trinn**

| | Gjennomsnitt NP8 | Lesing NP8 | Regning NP8 | Engelsk NP8 |
|---|---|---|---|---|
| Gjennomsnitt NP8 ... | 1 | | | |
| Lesing NP8 ..... | 0,916 | 1 | | |
| Regning NP8 ..... | 0,918 | 0,764 | 1 | |
| Engelsk NP8 ..... | 0,918 | 0,793 | 0,765 | 1 |
| Antall skoler ..... | 1 211 | | | |

**Tabell 6.14. Korrelasjoner på tvers av fag – indikatorer nasjonale prøver 8. trinn**

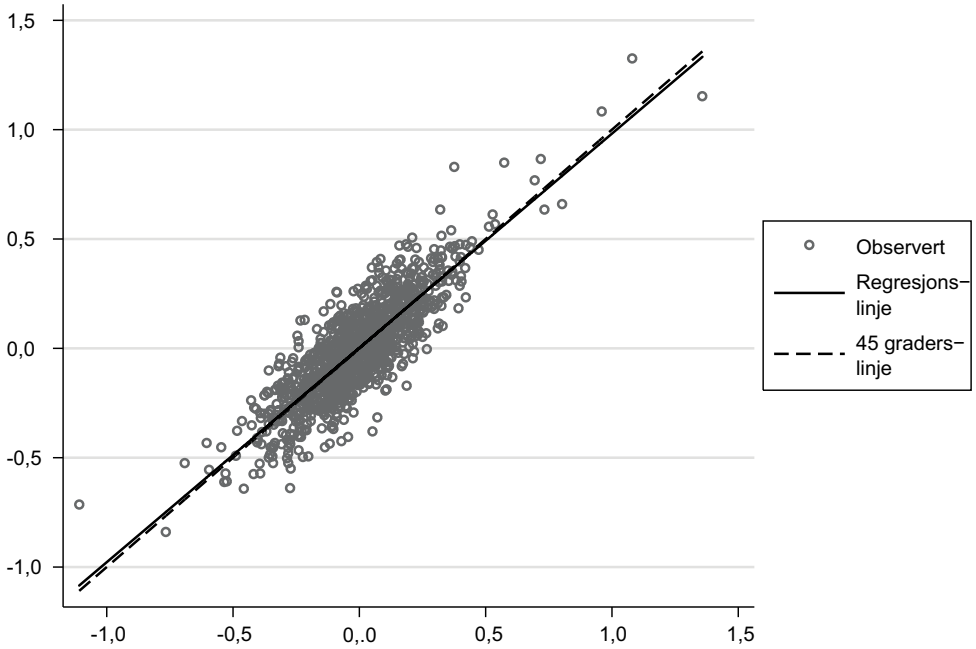
| | Gjennomsnitt NP8 | Lesing NP8 | Regning NP8 | Engelsk NP8 |
|---|---|---|---|---|
| Gjennomsnitt NP8 ... | 1 | | | |
| Lesing NP8 ..... | 0,852 | 1 | | |
| Regning NP8 ..... | 0,848 | 0,579 | 1 | |
| Engelsk NP8 ..... | 0,851 | 0,624 | 0,580 | 1 |
| Antall skoler ..... | 1211 | | | |

**Figur 6.7. Sammenheng mellom indikatorer – lesing versus regning, nasjonale prøver 8. trinn**




1145 skoler med 49144 elever. Elevvektet reg.koeff: 0.544; korrelasjon=0.579; rangkorr.=0.535

**Figur 6.8. Sammenheng mellom indikatorer – lesing versus gjennomsnitt, nasjonale prøver 8. trinn**




1155 skoler med 49364 elever. Elevvektet reg.koeff: 0.980; korrelasjon=0.852; rangkorr.=0.815

## 6.4. Usikkerhet ved indikatorene

Det er statistisk usikkerhet knyttet til forskjeller i skolebidrag mellom skoler på grunn av tilfeldig variasjon i resultatene og fordi vi anslår betydningen av ulik elevsammensetning på tvers av skoler. Denne usikkerheten med medfører at ikke alle forskjeller mellom skoler er statistisk signifikante, dvs. at vi med en rimelig

sikkerhet kan avvise at de skyldes tilfeldigheter. Når man sammenligner indikatorene til to skoler, er det ikke alltid grunn til å legge stor vekt på forskjellen mellom dem. Den er ikke nødvendigvis hva vi kaller statistisk signifikante. Forskjellen er signifikant dersom vi med stor sikkerhet kan avvise at forskjellene skyldes tilfeldigheter. Det er kun i disse tilfellene vi vil legge vekt på forskjellen vi finner mellom to skoler. En brukbar pekepinn på signifikans får man ved å sammenligne konfidensintervallene til skolebidragsindikatorene. La oss tenke oss to skoler, der skolebidragsindikatoren for skole A har høyere verdi enn for skole B. Dersom den nedre grensen i konfidensintervallet for A ligger høyere enn den høyeste for skole B har vi en klar indikasjon på en signifikant forskjell. Dette gir imidlertid ingen eksakt test. Konfidensintervallene til en enkelt parameter avhenger kun av variansen til parameteren selv, mens et konfidensintervall til en forskjell mellom to parametre også avhenger av kovariansen til de to parametrene.

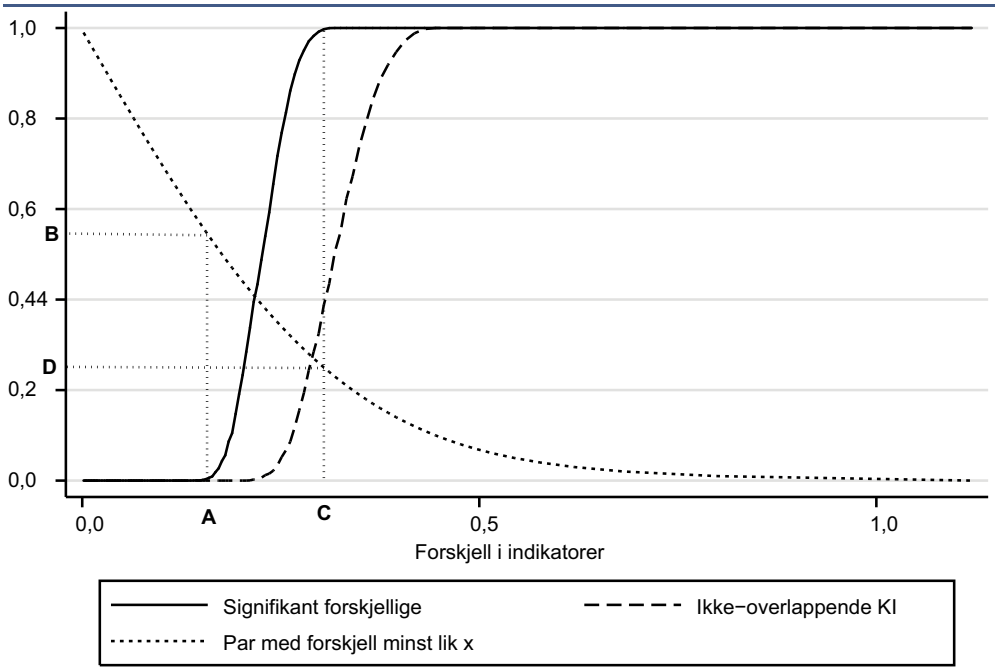
Figur 6.9 oppsummerer usikkerheten ved å se på alle mulige parvise sammenlikninger av skoler, både hvor store forskjeller mellom skoler er i tillegg til usikkerheten ved anslagene. Den finstiplede kurven viser hvor stor andel av alle parvise forskjeller som minst er av en viss størrelse. (Alle – dvs. andel lik 1 – forskjellene er minst så store som 0, mens nesten ingen er større enn 1).[8] Ved en forskjell på 0,25 standardavvik er det ca 30 prosent av skoleparene der forskjellen er større en dette. Den heltrukne kurven viser andelen av forskjellene som er statistisk signifikante (på 95 prosent nivå), avhengig av størrelsen på forskjellene. Selv om figuren er relativt komplisert kan hovedbildet enkelt oppsummeres: Når forskjellen mellom skoler er på 0,15 standardavvik eller mindre, dvs. inntil punktet A på den horisontale aksen, er denne aldri signifikant forskjellig fra null. Vi kan lese av andelen skoler dette omfatter fra den finstiplede kurven. Når vi gjør dette, som vist med de prikkede linjene, kommer vi til punktet B på den vertikale aksen, som svarer til en verdi mellom 0,5 og 0,6. Det er altså i overkant av halvparten av alle parvise sammenligninger som er minst så store. Knapt halvparten er mindre enn dette, og dermed – som vi ser fra den heltrukne kurven – aldri signifikant forskjellige. På den annen side er store forskjeller større enn ca 0,30 standardavvik (punkt C på den horisontale aksen, og nær 75-25 differansen i Tabell 6.2) alltid signifikante. Når vi leser av andelen fra den finstiplede kurven kommer til punktet D på den vertikale aksen, og ser at dette utgjør knappe 30 prosent av de parvise forskjellene. Når forskjellen ligger mellom 0,15 og 0,30 standardavvik kan den være signifikant eller insignifikant – begge deler forekommer, men med økende forskjell er en økende andel signifikante. Omkring 20 prosent av skoleparene ligger i dette intervallet.

Den grovstiplede linjen viser hvor mange av de estimerte skolebidragsindikatorene som har ikke-overlappende konfidensintervall. Å vurdere hvorvidt en forskjell er signifikant krever informasjon om de beregnede indikatorene, samt usikkerheten i indikatorene. Generelt er det også nødvendig med informasjon om sammenhengen mellom usikkerheten i indikatorene som sammenlignes. På grunn av den ekstra informasjonen er dette mer krevende, skolebidragsindikatorer har tradisjonelt sett også bare blitt publisert med informasjon om usikkerheten, og ikke om samvariasjonen mellom forskjellige skolers indikatorer. Den enklere sammenligningen, der vi ser bort fra samvariasjonen, svarer til å undersøke hvorvidt to skolers konfidensintervall overlapper. I figuren ser vi at den grovstiplede linjen, som altså angir andel skoler med ikke-overlappende konfidensintervall, ligger lavere enn den heltrukne linjen. For en gitt forskjell i indikatorer er det altså like mange eller flere skolepar som er statistisk signifikante enn det er som har ikke-overlappende konfidensintervall. Det vil si at ikke-overlappende konfidensintervall er en konservativ ”test” av signifikante forskjeller, vi kan altså ha at en forskjell mellom to skoler er statistisk signifikant, selv om konfidensintervallene overlapper, mens det motsatte ikke skjer.

[8] Figuren viser resultater for persentiler av parvis forskjell, dvs. hundre omtrent like store grupper av skolepar, sortert etter forskjellen i indikatorer.

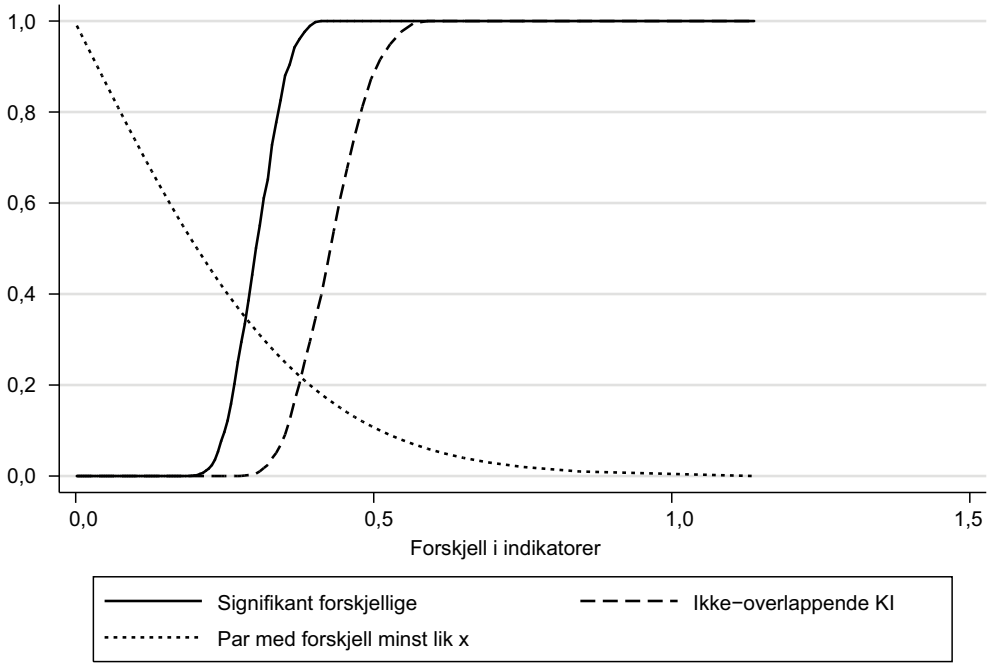
Eller, alle skolepar med ikke-overlappende konfidensintervall er signifikant forskjellige, og det er også noen skolepar med overlappende konfidensintervall.[9]

**Figur 6.9. Fordeling av skoleforskjeller og statistisk usikkerhet. Gjennomsnitt nasjonale prøver 8. trinn**




Andeler per persentil, 1211 skoler (732655 par)

**Figur 6.10. Fordeling av skoleforskjeller og statistisk usikkerhet. Nasjonale prøver 8. trinn, lesing**



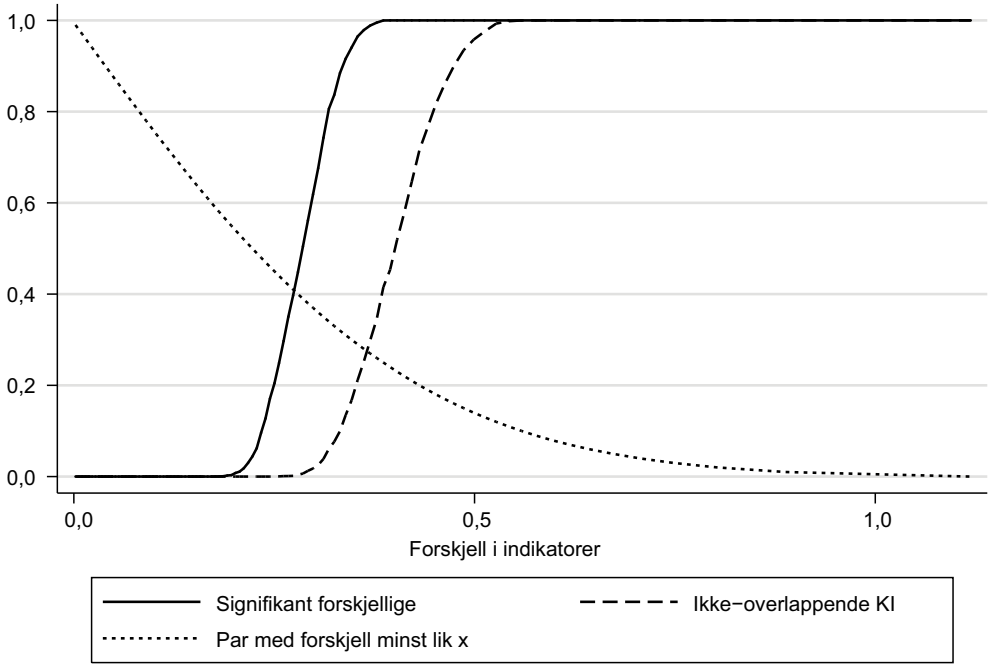

Andeler per persentil, 1155 skoler (666435 par)

[9] Denne konklusjonen er egentlig litt sterkere enn hva som kan leses ut av figuren. Selv om det er en høyere andel som er signifikante enn det er som har ikke-overlappende konfidensintervall, kan det i prinsippet likevel finnes enkeltpar som ikke er signifikant forskjellige, men har ikke-overlappende konfidensintervall. Dette krever i så fall en negativ samvariasjon. En nærmere inspeksjon av de estimerte samvariasjonene viser at de nesten utelukkende er positive. De få negative som finnes er av liten størrelse, og har ingen betydning for signifikans av forskjeller.

Figur 6.10, Figur 6.11 og Figur 6.12 viser tilsvarende beregninger for enkeltfag. Riktignok er noe færre skoleforskjeller signifikante, men igjen er mønsteret for enkeltfag svært likt hva vi finner for gjennomsnittet av de nasjonale prøvene i 8.klasse.

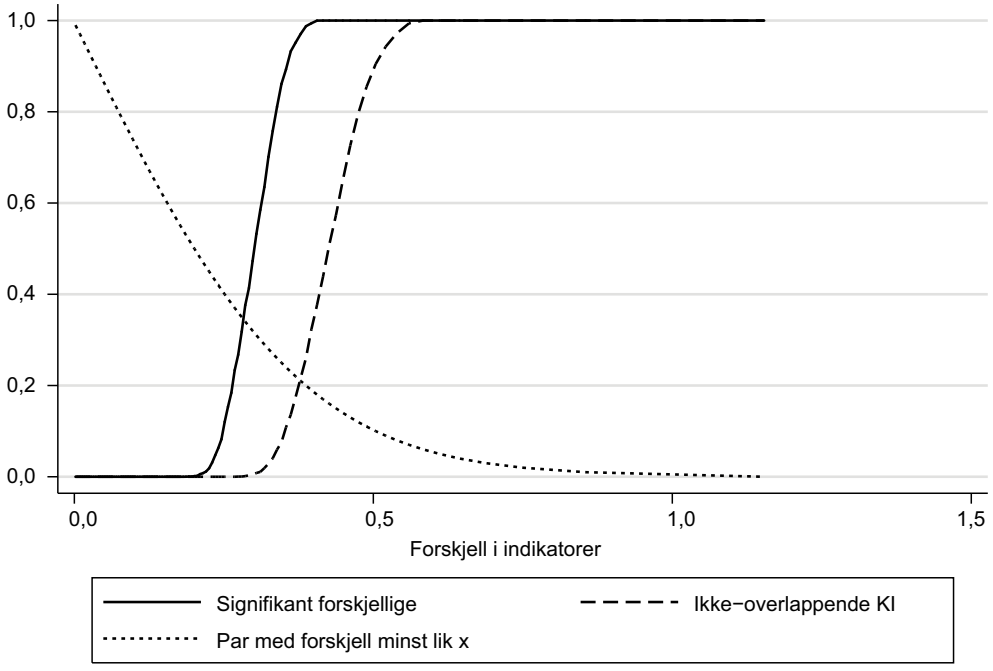
Figur 6.10, Figur 6.11 og Figur 6.12 viser tilsvarende beregninger for enkeltfag. Riktignok er noe færre skoleforskjeller signifikante, men igjen er mønsteret for enkeltfag svært likt hva vi finner for gjennomsnittet av de nasjonale prøvene på 8. trinn.

**Figur 6.11. Fordeling av skoleforskjeller og statistisk usikkerhet. Nasjonale prøver 8. trinn, regning**




Andeler per persentil, 1185 skoler (701520 par)

**Figur 6.12. Fordeling av skoleforskjeller og statistisk usikkerhet. Nasjonale prøver 8. trinn, engelsk**




Andeler per persentil, 1173 skoler (687378 par)

# 7. Indikatorer for ungdomstrinnet

I dette kapitlet ser vi på indikatorer for ungdomstrinnet basert på avgangsresultater fra 10.trinn. Læringsutbyttet for ungdomstrinnet blir tilknyttet avslutningsskolen og vi ser således bort fra at elever bytter skole fra 8. til 10.trinn.[10] Her er de relevante tidligere resultater de nasjonale prøvene på 8. trinn. Fra avgangsresultatene benytter vi to ulike resultatmål: Karakter på skriftlig eksamen (som er enten i norsk, matematikk eller engelsk) eller gjennomsnittlig standpunktkarakter i det vi benevner som basisfag (norsk hovedmål, matematikk og engelsk). Analysen følger i stor grad opplegget for nasjonale prøver 8. trinn i forrige kapittel, men drøftingen er mer summarisk. Dette skyldes at mange av problemstillingene knyttet til valg av indikator er uavhengig av resultatmål, slik at en full drøfting vil gi mye gjentakelse. Vi har imidlertid gjennomført de samme analysene også for resultatmålet som omtales i dette kapitlet, der vi ikke rapporterer fullt ut er mønstrene om lag tilsvarende de vi finner i forrige kapittel.

På ett viktig punkt har vi imidlertid en mer utførlig drøfting. For avgangsresultater fra ungdomstrinnet har vi i prinsippet to ulike resultatmål som er ment å måle de samme ferdighetene, nemlig standpunkt- og eksamenskarakterer.[11] Mens eksamen er en ekstern vurdering med anonymitet vil standpunktkarakteren lettere påvirkes av kunnskapsnivået i gruppen (relativ karaktersetting) og relasjoner mellom lærer og elev. På den annen side er eksamensutfallet langt sterkere påvirket av tilfeldigheter, siden grunnlaget for standpunkt er observasjon og interaksjon gjennom hele året. Dersom sammenhengen mellom underliggende kunnskapsnivå og resultat er den samme for standpunkt- og eksamenskarakterer, ville man forvente å finne en ganske høy korrelasjon mellom indikatorer målt ved hver av dem. Vi finner imidlertid en relativt lav korrelasjon. Avslutningsvis i dette kapitlet drøfter vi mulige forklaringer på dette og hvilke implikasjoner det gir for valg av indikator.

## 7.1. Resultater

Tabell 7.1 og Tabell 7.2 rapporterer resultater fra regresjonsanalyser på elevnivå for henholdsvis skriftlig eksamen og standpunktkarakterer. I begge tabellene presenterer vi resultater fra i alt fem ulike spesifikasjoner. I kolonne (1) kontrollerer vi kun for familiebakgrunn og ikke for tidligere resultater. Vi ser som ventet at karakterene er høyere i engelsk og lavere i matematikk, sammenliknet med norsk. Jenter har bedre karakterer og kjønnsforskjellene er størst for standpunktkarakterer. Elever uten resultater fra de nasjonale prøvene har svakere resultater tre år senere. Skoleeffektene som følger fra denne regresjonen svarer til skolebidragsindikatorene som tidligere er beregnet i bl.a. Hægeland, Kirkebøen, Raaum og Salvanes (2005a), men her med en noe enklere spesifikasjon av familiebakgrunn. I kolonne (2) viser vi resultater fra regresjonsanalyser med mer detaljert kontroll for familiebakgrunn, i form av den mer omfattende familiebakgrunnsvektoren fra Kirkebøen, Raaum og Salvanes (2005a). I kolonne (3) og (4) kontrollerer for tidligere resultater enten ved å inkludere resultatet fra nasjonale prøver, 8. trinn i det samme faget som eleven var oppe i til skriftlig eksamen (Tabell 7.1) eller gjennomsnittlig skår nasjonale prøver, 8. trinn (Tabell 7.2). I kolonne (5) og (6) kontrollerer for tidligere resultater ved å inkludere resultatene fra nasjonale prøver, 8. trinn i hvert emne for seg. For hver av value added-spesifikasjonene, dvs. i kolonnene (3)-(6), rapporterer vi resultater med og uten kontroll for familiebakgrunn.

Raden $R^2$ uttrykker hvor stor andel av variansen i det aktuelle resultatmålet som blir forklart av kjennetegnene i modellen. Tilsvarende som for nasjonale prøver for 8. trinn, ser vi at det å inkludere tidligere resultater bidrar til å forklare mye av

[10] Dette gjelder såpass få elever at det neppe skaper noen systematiske skjevheter, men som målefeil vil det bidra til å viske ut forskjeller mellom skoler.

[11] Nasjonale prøver på 8. trinn har også flere resultatmål, men de dekker hver sine emner målt på samme måte.

variasjonen i individuelle resultater, men at det har mindre betydning hvordan dette gjøres. Tidligere resultater er dermed – og kanskje ikke så overraskende – svært viktig som forklaringsfaktor for elevresultater. Bare å basere seg på informasjon om nåværende resultater og familiebakgrunnsvariable når man beregner skolebidragsindikatorer kan dermed være utilstrekkelig. Dessuten måler value added-indikatorer, nettopp ved å kontrollere for tidligere resultater, skolebidraget over et mer presist definert tidsrom som i dette tilfellet er tre klassetrinn. Resultater på et gitt tidspunkt fanger opp skoleeffektene fra alle tidligere klassetrinn. Å inkludere familiebakgrunnsvariable når man først har kontrollert for tidligere resultater har relativt liten betydning, det har heller ikke særlig stor betydning om vi kontrollerer for mer eller mindre detaljert familiebakgrunn.

**Tabell 7.1. Regresjonsresultater, basert på skriftlig eksamen**

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Skr. eks.kar | Skr. eks.kar | Skr. eks.kar | Skr. eks.kar | Skr. eks.kar | Skr. eks.kar |
| Skriftlig eksamen engelsk ..... | 0,249*** | 0,252*** | 0,760*** | 0,719*** | 0,236*** | 0,236*** |
| | (0,0113) | (0,0111) | (0,00971) | (0,00955) | (0,00911) | (0,00903) |
| Skriftlig eksamen matematikk . | -0,263*** | -0,263*** | 0,217*** | 0,178*** | -0,274*** | -0,274*** |
| | (0,0112) | (0,0111) | (0,00963) | (0,00947) | (0,00908) | (0,00900) |
| Mangler observasjon fra tidligere år ..... | -0,349*** | | -0,470*** | -0,350*** | -0,322*** | -0,282*** |
| | (0,0286) | | (0,0242) | (0,0255) | (0,0354) | (0,0361) |
| Jente ..... | 0,341*** | 0,344*** | 0,305*** | 0,309*** | 0,309*** | 0,311*** |
| | (0,00812) | (0,00802) | (0,00673) | (0,00660) | (0,00679) | (0,00673) |
| Poeng NP08 i samme fag som skriftlig eksamen | | | 0,813*** | 0,749*** | | |
| | | | (0,00424) | (0,00436) | | |
| Mangler res NP08 i samme fag som skriftlig eksamen | | | -0,219*** | -0,189*** | | |
| | | | (0,0124) | (0,0122) | | |
| Poeng engelsk NP8 | | | | | 0,232*** | 0,224*** |
| | | | | | (0,00471) | (0,00469) |
| Poeng regning NP8 | | | | | 0,368*** | 0,342*** |
| | | | | | (0,00493) | (0,00495) |
| Poeng lesing NP8 | | | | | 0,237*** | 0,219*** |
| | | | | | (0,00559) | (0,00563) |
| Mangler res NP8 engelsk | | | | | -0,0774*** | -0,0692*** |
| | | | | | (0,0183) | (0,0181) |
| Mangler res NP8 regning | | | | | -0,172*** | -0,155*** |
| | | | | | (0,0193) | (0,0191) |
| Mangler res NP8 lesing | | | | | -0,124*** | -0,112*** |
| | | | | | (0,0196) | (0,0195) |
| Konstantledd ..... | 3,347*** | 3,339*** | 3,010*** | 3,028*** | 3,391*** | 3,384*** |
| | (0,00902) | (1,001) | (0,00781) | (0,00768) | (0,00741) | (0,00738) |
| Familiebakgrunnsvar inkl. ..... | Ja | Detaljert | Nei | Ja | Nei | Ja |
| Forklaringskraft ($R^2$) ..... | 0,177 | 0,202 | 0,433 | 0,456 | 0,460 | 0,470 |
| Antall elever ..... | 58 870 | 58 870 | 58 870 | 58 870 | 58 870 | 58 870 |

Estimerte standardfeil i parentes. Statistisk signifikans: * p<0,05, ** p<0,01, *** p<0,001

Tabell 7.3 og Tabell 7.4 gjengir beskrivende statistikk for skolebidrags-indikatorene som følger fra estimeringene i de ulike kolonnene i Tabell 7.1 og Tabell 7.2, i tillegg til ujustert gjennomsnitt. Som vi ser, er spredningen i alle indikatorene lavere enn for de ujusterte resultatene. For eksamen faller standardavviket fra 0,31 for de ujusterte skolegjennomsnittene til under 0,2 når vi kontrollerer for resultatene fra nasjonale prøver for 8.trinn. Indikatorene som kontrollerer for tidligere resultater har lavere spredning enn der hvor vi bare kontrollerer for forskjeller i familiebakgrunn. Særlig gjelder dette eksamen der eksempelvis karakterforskjellen mellom gode (90-persentilen) og dårlige skoler (10-persentilen) går ned fra 0,575 til 0,464. For variasjon i standpunktkarakterer mellom skoler er den svært lite påvirket av å korrigere for resultater på 8.trinn, dersom vi i utgangspunktet har tatt hensyn til forskjeller i familiebakgrunn.

**Tabell 7.2. Regresjonsresultater, basert på standpunktkarakterer**

| | (1) Standpunkt basisfag | (2) Standpunkt basisfag | (3) Standpunkt basisfag | (4) Standpunkt basisfag | (5) Standpunkt basisfag | (6) Standpunkt basisfag |
|---|---|---|---|---|---|---|
| Mangler observasjon fra tidligere år | -0,318*** | | | | -0,186*** | -0,160*** |
| | (0,0251) | | | | (0,0263) | (0,0260) |
| Jente | 0,382*** | 0,385*** | 0,323*** | 0,328*** | 0,357*** | 0,359*** |
| | (0,00681) | (0,00664) | (0,00481) | (0,00465) | (0,00490) | (0,00475) |
| Poeng NP08, gjennomsnitt | | | 0,754*** | 0,696*** | | |
| | | | (0,00258) | (0,00270) | | |
| Mangler res NP08 | | | -0,624*** | -0,565*** | | |
| | | | (0,0165) | (0,0171) | | |
| Poeng engelsk NP8 | | | | | 0,225*** | 0,214*** |
| | | | | | (0,00340) | (0,00330) |
| Poeng regning NP8 | | | | | 0,411*** | 0,377*** |
| | | | | | (0,00356) | (0,00349) |
| Poeng lesing NP8 | | | | | 0,244*** | 0,225*** |
| | | | | | (0,00404) | (0,00398) |
| Mangler res NP8 engelsk | | | | | -0,105*** | -0,0952*** |
| | | | | | (0,0133) | (0,0129) |
| Mangler res NP8 regning | | | | | -0,200*** | -0,180*** |
| | | | | | (0,0139) | (0,0135) |
| Mangler res NP8 lesing | | | | | -0,148*** | -0,138*** |
| | | | | | (0,0142) | (0,0138) |
| Konstantledd | 3,584*** | 3,359*** | 3,613*** | 3,607*** | 3,615*** | 3,608*** |
| | (0,00483) | (0,608) | (0,00339) | (0,00329) | (0,00349) | (0,00341) |
| Familiebakgrunnsvar inkl. | Ja | Detaljert | Nei | Ja | Nei | Ja |
| Forklaringskraft ($R^2$) | 0,230 | 0,272 | 0,616 | 0,642 | 0,626 | 0,649 |
| Antall elever | 58 956 | 58 956 | 58 956 | 58 956 | 58 956 | 58 956 |

Estimerte standardfeil i parentes. Statistisk signifikans: * p<0,05, ** p<0,01, *** p<0,001

Det samme mønsteret er illustrert i Figur 7.1 og Figur 7.2, hvor den heltrukne grafen viser fordelingen til ujusterte resultater, den stiplede til indikatoren som følger av å kontrollere for familiebakgrunn, mens den prikkede viser fordelingen til indikatoren som følger av å kontrollere for alle resultater fra NP 8. trinn. Som vi ser er spredningen mindre enn for ujusterte resultater når man kontrollerer for familiebakgrunn, og enda mindre når man kontrollerer for tidligere resultater. Men som både tabellene og figurene viser, det er særlig for skriftlig eksamen at spredningen reduseres når vi kontrollerer for tidligere resultater. Dette gjenspeiler at familiebakgrunn fanger opp en større del av variasjonen i standpunktkarakterer enn i skriftlig eksamen.

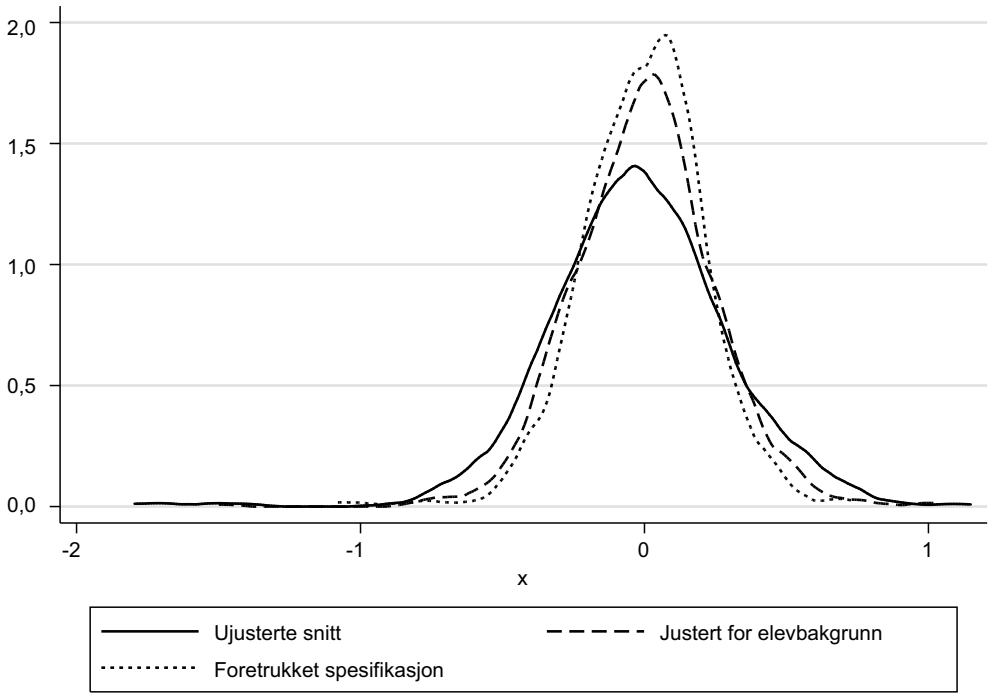
**Tabell 7.3. Beskrivende statistikk indikatorer basert på skriftlig eksamen (elevvektet)**

| | Gj.snitt | Std. avvik | Min. | 10. per-sentil | 25. per-sentil | 50. per-sentil | 75.per-sentil | 90.per-sentil | Maks. |
|---|---|---|---|---|---|---|---|---|---|
| Ujustert gjennomsnitt | 0,006 | 0,310 | -1,796 | -0,367 | -0,183 | -0,005 | 0,188 | 0,400 | 1,149 |
| Kontroll for familiebakgrunn (1) | -0,003 | 0,220 | -1,499 | -0,296 | -0,132 | 0,012 | 0,131 | 0,279 | 1,035 |
| Kontroll for familiebakgrunn, stor (2) | -0,003 | 0,213 | -1,328 | -0,280 | -0,131 | 0,010 | 0,131 | 0,265 | 0,986 |
| Kontroll tilsv. fag NP8 (3) | 0,002 | 0,201 | -1,213 | -0,259 | -0,126 | 0,009 | 0,131 | 0,239 | 0,822 |
| Kontroll tilsv. fag NP8, fam.bak. (4) | -0,002 | 0,184 | -1,058 | -0,239 | -0,114 | 0,010 | 0,109 | 0,218 | 0,802 |
| Kontroll alle fag NP8 (5) | 0,001 | 0,192 | -1,080 | -0,237 | -0,118 | 0,018 | 0,124 | 0,227 | 0,739 |
| Kontroll alle fag NP8, fam.bak (6) | -0,001 | 0,185 | -1,003 | -0,227 | -0,115 | 0,008 | 0,117 | 0,217 | 0,738 |
| Antall skoler | 777 | | | | | | | | |

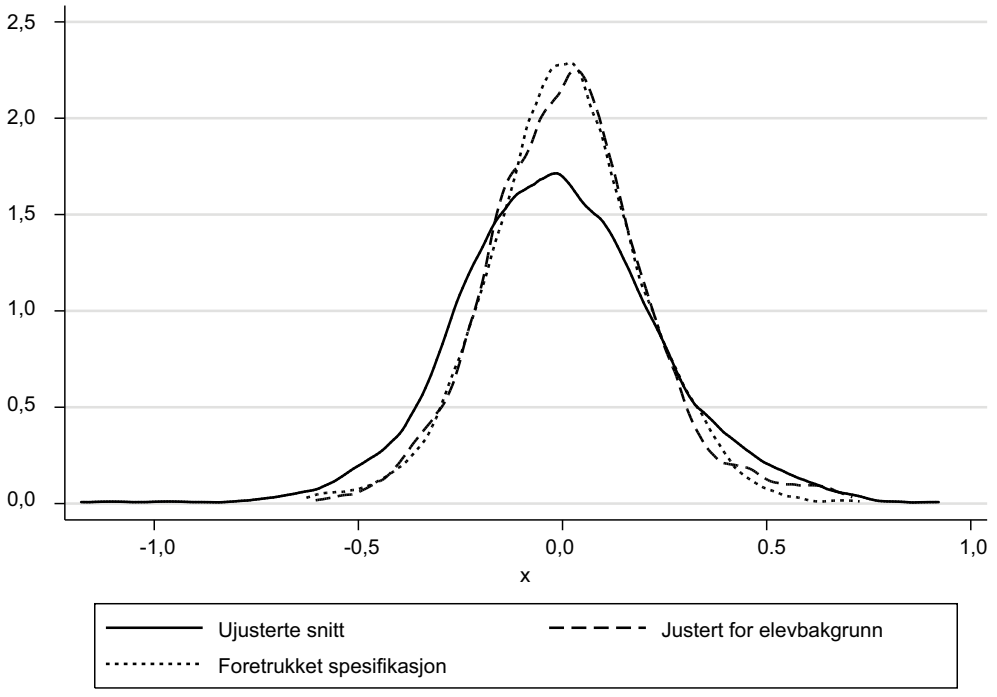
**Tabell 7.4. Beskrivende statistikk indikatorer basert på standpunktkarakterer (elevvektet)**

| | Gj.snitt | Std. avvik | Min. | 10. per-sentil | 25. per-sentil | 50. per-sentil | 75.per-sentil | 90.per-sentil | Maks. |
|---|---|---|---|---|---|---|---|---|---|
| Ujustert gjennomsnitt ..... | -0,001 | 0,236 | -1,179 | -0,277 | -0,162 | -0,006 | 0,157 | 0,300 | 0,922 |
| Kontroll for familiebakgrunn (1) ..... | -0,010 | 0,173 | -0,605 | -0,208 | -0,119 | -0,009 | 0,094 | 0,204 | 0,710 |
| Kontroll for familiebakgrunn, stor (2) .... | -0,010 | 0,170 | -0,532 | -0,206 | -0,122 | -0,011 | 0,097 | 0,199 | 0,674 |
| Kontroll tilsv. fag NP8 (3) .... | -0,006 | 0,169 | -0,600 | -0,221 | -0,114 | -0,010 | 0,099 | 0,203 | 0,735 |
| Kontroll tilsv. fag NP8, fam.bak. (4) ..... | -0,009 | 0,177 | -0,655 | -0,229 | -0,119 | -0,010 | 0,102 | 0,207 | 0,650 |
| Kontroll alle fag NP8 (5) ..... | -0,006 | 0,170 | -0,627 | -0,221 | -0,114 | -0,008 | 0,099 | 0,204 | 0,728 |
| Kontroll alle fag NP8, fam.bak (6) ..... | -0,009 | 0,178 | -0,666 | -0,222 | -0,123 | -0,010 | 0,098 | 0,209 | 0,687 |
| Antall skoler ..... | 776 | | | | | | | | |

**Figur 7.1. Fordeling av ulike resultatmål basert på skriftlig eksamen**




**Figur 7.2. Fordeling av ulike resultatmål basert på standpunktkarakterer**



Figur 7.1 og Figur 7.2 viser at forskjeller mellom skoler blir mindre ved justering, men ingenting om hvor mye vår vurdering av hvilke skoler som bidrar lite eller mye til elevenes læring i ungdomsårene endres.

For å kunne vurdere om de ulike indikatorene gir forskjellige svar på hvilke skoler som bidrar mye eller lite til elevenes læring, må vi se på samvariasjonen mellom dem. Hvis justeringen reduserte forskjellene mellom skolene, men ikke påvirket de relative forskjellene (dette ville også medført at rangeringen av skolene ville vært den samme) ville korrelasjonen vært 1. Tabell 7.5 og Tabell 7.6 viser korrelasjonsmatrisene for de ulike indikatorene for henholdsvis skriftlig eksamen og standpunkt. Som vi ser av tabellene, er det en meget høy korrelasjon mellom alle indikatorer hvor man kontrollerer for tidligere resultater, omtrent 0,9 eller høyere. Som i forrige kapittel er det også slik at gitt hvordan man kontrollerer for tidligere resultater, så har det liten betydning for indikatorene om man i tillegg kontrollerer for familiebakgrunn – her er korrelasjonene omtrent 0,96-0,98. Vi merker oss også at det å kontrollere for familiebakgrunn har mer å si for standpunktkarakterer enn for skriftlig eksamenskarakter. Korrelasjonen mellom indikatoren med familiekontroll og den som kontrollerer for resultater fra 8. trinn (men ikke familiebakgrunn) er på 0,804 for eksamenskarakterer, mens den er 0,674 for standpunkt.

**Tabell 7.5: Korrelasjon mellom ulike indikatorer basert på skriftlig eksamen**

| | Ujust snitt | Ujust snitt, FB | Ujust snitt, FB, det. | Kontr. tilsv. fag | Kontr. tilsv. fag, FB | Kontr. alle fag | Kontr. alle fag, FB |
|---|---|---|---|---|---|---|---|
| Ujustert gjennomsnitt ..... | 1 | | | | | | |
| Kontroll for familiebakgrunn (1) | 0,867 | 1 | | | | | |
| Kontroll for familiebakgrunn, det. (2) ..... | 0,840 | 0,987 | 1 | | | | |
| Kontroll tilsv. fag NP8 (3) ..... | 0,814 | 0,833 | 0,827 | 1 | | | |
| Kontroll tilsv. fag NP8, fam.bak. (4) ..... | 0,680 | 0,834 | 0,841 | 0,957 | 1 | | |
| Kontroll alle fag NP8 (5) ..... | 0,725 | 0,804 | 0,810 | 0,933 | 0,932 | 1 | |
| Kontroll alle fag NP8, fam.bak (6) ..... | 0,632 | 0,791 | 0,805 | 0,890 | 0,944 | 0,982 | 1 |
| Antall skoler ..... | 777 | | | | | | |

**Tabell 7.6. Korrelasjon mellom ulike indikatorer basert på standpunktkarakterer**

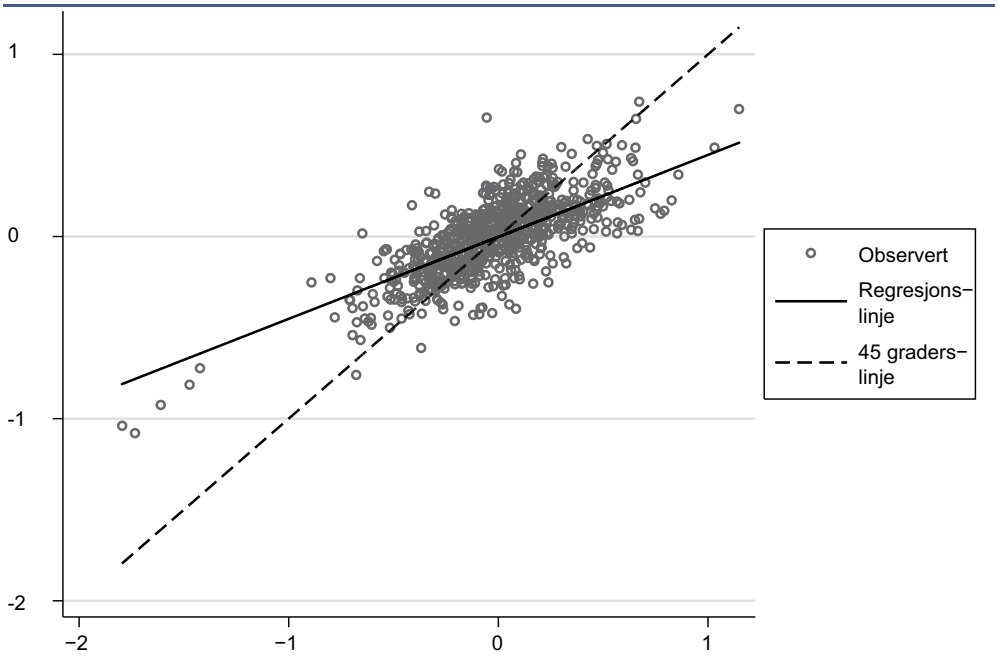
| | Ujust snitt | Ujust snitt, FB | Ujust snitt, FB, det. | Kontr. tilsv. fag | Kontr. tilsv. fag, FB | Kontr. alle fag | Kontr. alle fag, FB |
|---|---|---|---|---|---|---|---|
| Ujustert gjennomsnitt ..... | 1 | | | | | | |
| Kontroll for familiebakgrunn (1) | 0,684 | 1 | | | | | |
| Kontroll for familiebakgrunn, det. (2) ..... | 0,621 | 0,973 | 1 | | | | |
| Kontroll tilsv. fag NP8 (3) ..... | 0,442 | 0,691 | 0,712 | 1 | | | |
| Kontroll tilsv. fag NP8, fam.bak. (4) ..... | 0,259 | 0,688 | 0,723 | 0,955 | 1 | | |
| Kontroll alle fag NP8 (5) ..... | 0,416 | 0,674 | 0,698 | 0,986 | 0,949 | 1 | |
| Kontroll alle fag NP8, fam.bak (6) ..... | 0,245 | 0,673 | 0,711 | 0,943 | 0,990 | 0,960 | 1 |
| Antall skoler ..... | 776 | | | | | | |

På samme måte som for indikatorer basert på nasjonale prøver 8. trinn, foretrekker vi den modellspesifikasjonen med den frieste spesifikasjonen av tidligere resultater, med andre ord der hvor vi inkluderer resultatene fra NP 8. trinn hver for seg. Begrunnelsen er den samme: Spesifikasjonen hviler på færre teoretiske forutsetninger, og den utnytter også mer av informasjonen om tidligere prestasjoner. Å kontrollere for familiebakgrunn i tillegg har begrenset informasjonsverdi, og praktiske hensyn tilsier igjen å benytte en modell hvor man ikke kontrollerer for familiebakgrunn.

Figur 7.3 og Figur 7.4 utdyper informasjonen om hvilke skoler som bidrar mye og lite til læring ut fra sammenhengen mellom foretrukket indikator og ujusterte resultater for henholdsvis skriftlig eksamen og standpunktkarakter. Det å kontrollere for tidligere resultater fører til at bildet av forskjeller mellom skoler og av hvilke skoler som bidrar mye og lite til elevenes læring endres betydelig. Selv

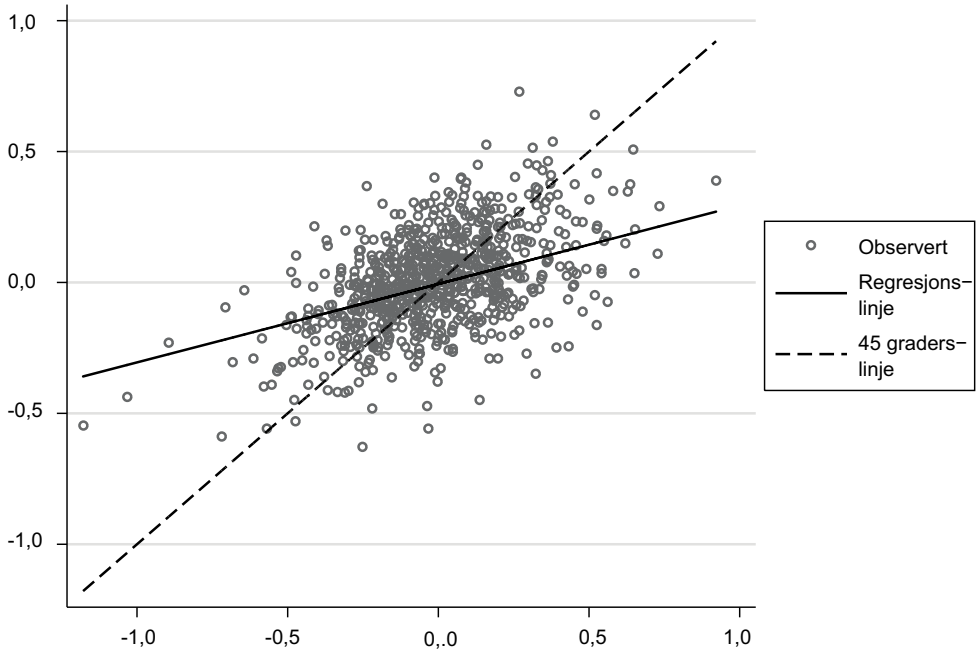
om det er slik at skoler som skårer høyt i karakterfordelingen tenderer til å skåre høyt også når vi kontrollerer for tidligere resultater, gir det å kontrollere for tidligere resultater et annet bilde av skolenes bidrag.

**Figur 7.3. Sammenheng mellom foretrukket indikator og ujustert resultat, skriftlig eksamen**




777 skoler med 55660 elever. Elevvektet reg.koeff: 0.450; korrelasjon=0.725; rangkorr.=0.717

**Figur 7.4. Sammenheng mellom foretrukket indikator og ujustert resultat, standpunktkarakterer**

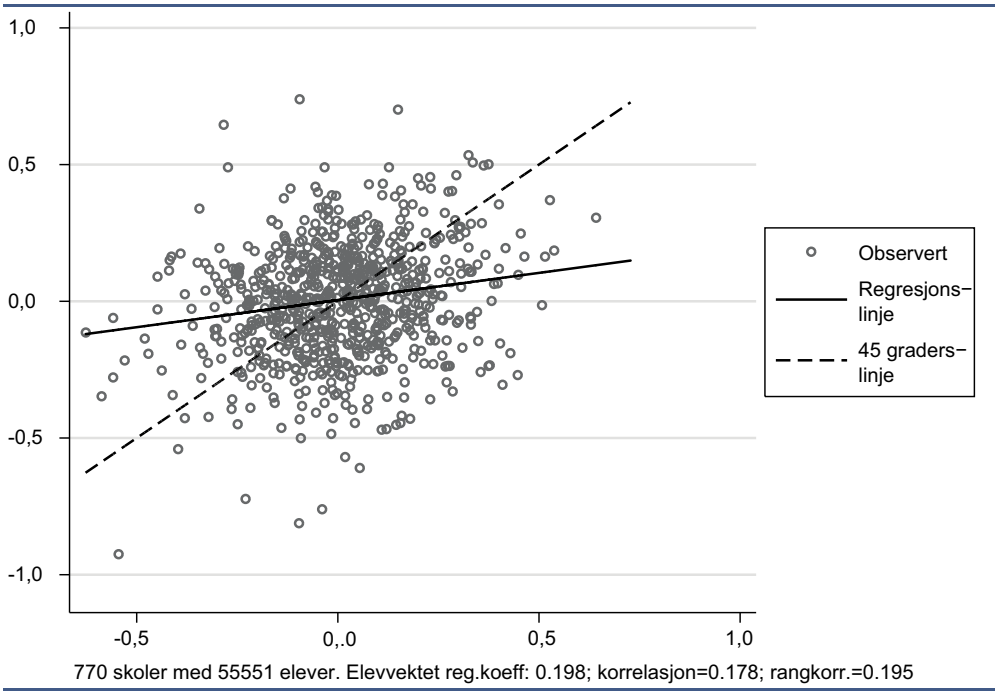



776 skoler med 55714 elever. Elevvektet reg.koeff: 0.299; korrelasjon=0.416; rangkorr.=0.474

Vi merker oss imidlertid at sammenhengen mellom ujusterte resultater og vår foretrukne indikator er vesentlig svakere for standpunktkarakterer enn for skriftlig eksamen. I Figur 7.5 sammenligner vi de to indikatorene. Som vi ser av figuren, er sammenhengen positiv, men svak, med en korrelasjon på om lag 0,2. I utgangspunktet skulle man forvente at korrelasjonen mellom dem var høy. Begge indikatorene baserer seg på resultater som måler ferdighetsoppnåelse i forhold til samme læreplan, datamaterialet omfatter de samme elevene og det er den samme statistiske modellen som ligger til grunn for beregningen. Spørsmålet er da om det

er ulike mekanismer som ligger til grunn for hvordan standpunkt- og eksamenskarakterer settes, og om disse varierer mellom skoler. Eksamenskarakter fastsettes ved ekstern sensur, mens standpunktkarakter settes av elevens lærer. Det kan tenkes at praksis ved setting av standpunktkarakterer varierer mellom skoler. I avsnitt 0 vurderer vi hvorvidt forskjeller i karakterpraksis mellom skoler kan ligge bak den lave korrelasjonen mellom indikatorer basert på eksamens- og standpunktkarakterer. Hvis dette er tilfelle, innebærer det at man skal være mer varsomme med å tolke skoleforskjeller fra indikatorer basert på standpunktkarakterer som faktiske forskjeller i skolenes bidrag til elevenes læring.

**Figur 7.5. Sammenheng mellom foretrukne indikatorer, skriftlig eksamen vs standpunktkarakterer**




## 7.2. Usikkerhet i indikatorene

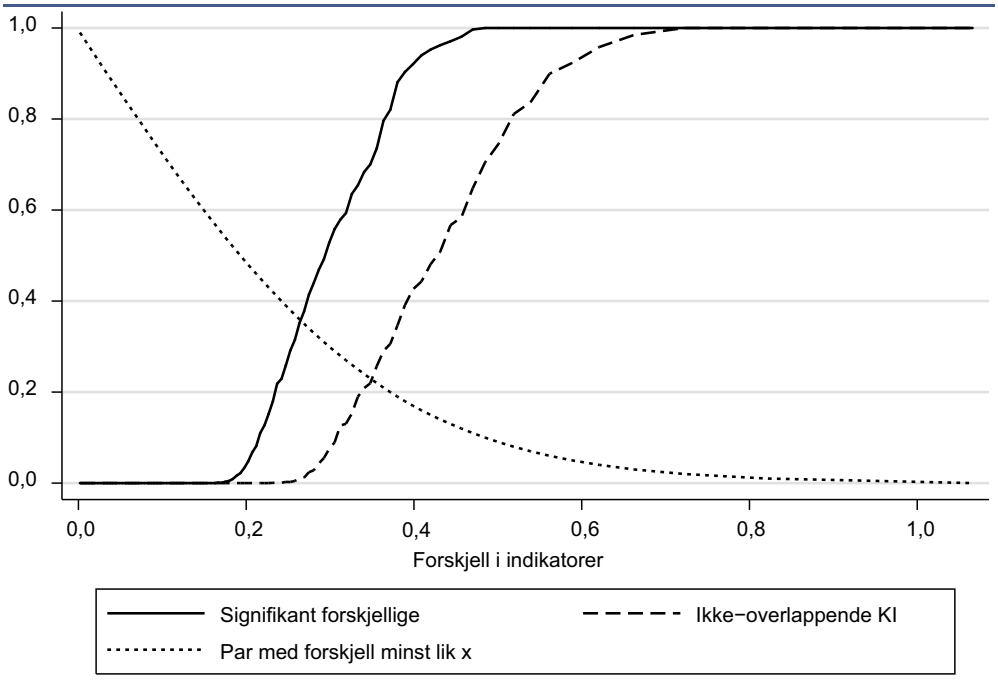
Som beskrevet i omtalen av indikatorene for mellomtrinnet, vil det alltid være statistisk usikkerhet i de estimerte indikatorene. Denne usikkerheten med medfører at ikke alle forskjeller mellom skoler er statistisk signifikante, dvs. at vi ikke med en rimelig grad av sikkerhet kan avvise at de skyldes tilfeldigheter. En brukbar pekepinn på signifikans får man ved å sammenligne konfidensintervallene til skolebidragsindikatorene.

Figur 7.6 og Figur 7.7 oppsummerer usikkerheten ved å se på alle mulige parvise sammenlikninger av henholdsvis indikatorer basert på skriftlig eksamen og standpunkt. Figurene viser hvor store forskjellene mellom skoler, samt usikkerheten ved anslagene. Den finstiplede kurven viser hvor stor andel av alle parvise sammenligninger som minst er av en viss størrelse. Denne er én ved forskjell på 0, dvs. alle forskjeller er minst 0 karakterpoeng store, mens nesten ingen forskjeller er større enn 0,8 karakterpoeng – da er kurven bare så vidt over den horisontale aksen. Mellom 40 og 50 prosent av de parvise forskjellene er større enn 0,2 karakterpoeng.

Den heltrukne kurven viser andelen av forskjellene som er statistisk signifikante (på 95 prosent nivå), avhengig av størrelsen på forskjellene. Selv om figuren er relativt komplisert kan hovedbildet i Figur 7.6 enkelt oppsummeres: Når forskjellen mellom skoler i indikatorer basert på skriftlig eksamen er på omtrent 0,2 karakterpoeng eller mindre er denne aldri signifikant – vi kan ikke legge noen større vekt på slike forskjeller. Dette gjelder drøyt halvparten av alle parvise sammenlikninger (jf. den finstiplede linjen). På den annen side er store forskjeller på ca 0,4 karakterpoeng omtrent alltid signifikante. Dette utgjør knappe 20 prosent av de parvise forskjellene.

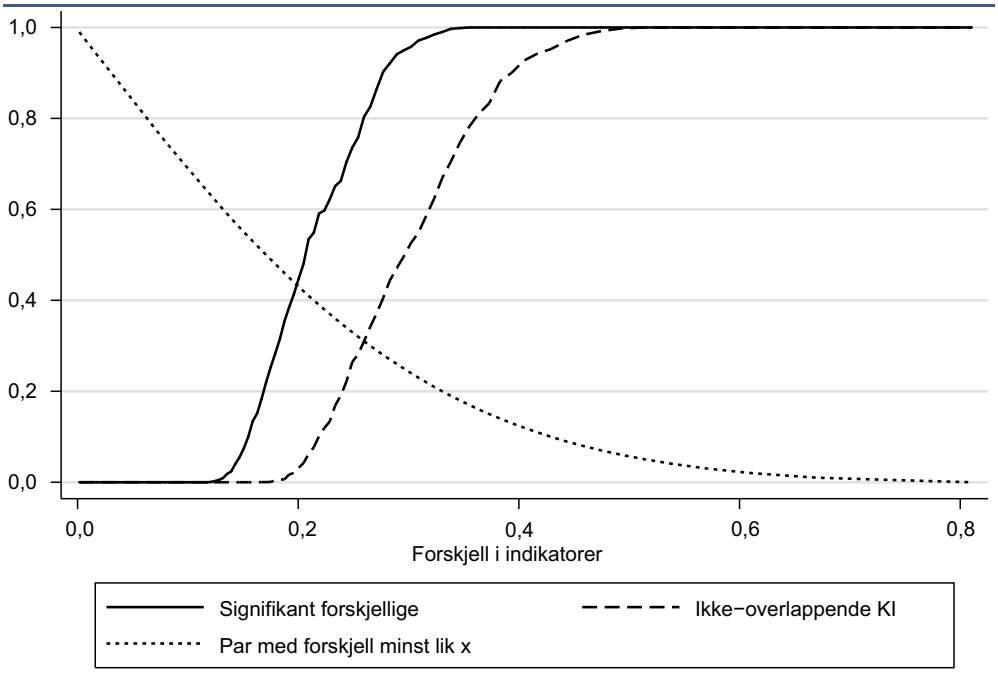
Når forskjellen ligger mellom 0,2 og 0,4 karakterpoeng finnes det både signifikante og insignifikante forskjeller, men andelen signifikante øker med størrelsen på forskjellen. Dette gjelder i overkant av 20 prosent av skoleparene. Den grovstiplede linjen viser hvor mange av de estimerte skolebidragsindikatorene som har ikke-overlappende konfidensintervall. Denne ligger lavere enn den heltrukne linjen, slik at det er færre skolepar med ikke-overlappende konfidensintervall enn det er signifikant forskjellige skolepar. Det vil si at ikke-overlappende konfidensintervall er en konservativ "test" av signifikante forskjeller. Det må være en forskjell på nesten 0,3 karakterpoeng før noen forskjeller har ikke-overlappende konfidensintervall, og det er bare de få parene med forskjeller større enn ca 0,7 karakterpoeng som aldri overlapper.

**Figur 7.6. Fordeling av skoleforskjeller og statistisk usikkerhet. Value added-indikatorer basert på skriftlig eksamen**




Andeler per persentil, 777 skoler (301476 par)

**Figur 7.7. Fordeling av skoleforskjeller og statistisk usikkerhet. Value added-indikatorer basert på standpunktkarakterer**




Andeler per persentil, 776 skoler (300700 par)

Figur 7.7 viser tilsvarende forskjeller for indikatorer basert på standpunkt-karakterer. Vi ser at de er mer presist beregnet enn indikatorene basert på skriftlig eksamen: Forskjeller mellom 0,1 og 0,2 karakterpoeng er i en del tilfeller signifikante, slik at gruppen uten noen signifikante forskjeller utgjør omtrent 40 prosent av parene, mot over halvparten i Figur 7.6. Omtrent halvparten av forskjellene på omkring 0,2 karakterpoeng er signifikante, og alle forskjeller større enn omtrent 0,3 karakterpoeng er signifikante – den siste gruppen utgjør i overkant av 20 prosent av skoleparene. Ikke-overlappende konfidensintervall gir igjen en konservativ test for signifikans, men det er flere skolepar med ikke-overlappende konfidensintervaller enn i Figur 7.6.

## 7.3. Sammenheng mellom ferdigheter, skolekvalitet og karakterpraksis

Indikatorene basert på henholdsvis eksamen og standpunkt gir ulike svar på hvilke skoler som bidrar lite eller mye til elevenes læring. På den ene side vil standpunkt-karakterer mer presist oppsummere de underliggende ferdighetene hos elevene ettersom faglærerne observerer elevene over lang tid, og langs ulike dimensjoner og kompetanse. En femtimers eksamen gir begrensninger når det gjelder å teste alle læringsmålene, og tilfeldigheter kan lettere påvirke utfallet. Den alvorlige innvendingen mot standpunkt er at vurderingen er lokal og ikke anonym. Dette medfører at karakterpraksis kan variere mellom elever, og mest bekymringsfullt i vår sammenheng, mellom skoler.

I dette avsnittet formulerer vi en enkel modell, basert på stiliserte antakelser, som illustrerer hvordan forskjeller i karakterpraksis og skolekvalitet kan separeres.[12] Vi viser hvordan forskjeller i karakterpraksis og skolekvalitet, og hvordan de eventuelt varierer sammen, vil avspeiles i samvariasjonen mellom beregnede indikatorer basert på standpunkt- og eksamenskarakterer. Modellen inneholder tre ligninger. Den første relaterer ferdighet ($F$) til kjennetegn ved eleven ($X$, dette kan for eksempel være familiebakgrunn eller tidligere ferdigheter), skolekvalitet ($\mu_j$, $j$-indeksen indikerer at denne er felles for alle elever ved skole $j$)[13] og tilfeldig variasjon:

$$F = X\gamma + \mu_j + \varepsilon \tag{8}$$

$\gamma$ uttrykker sammenhengen mellom observerbare elevkjennetegn og ferdighet, og $\mu_j$ er en skoleeffekt (eller skolekvalitet, som i gjennomsnitt er null). Restleddet ($\varepsilon$) representerer alle andre forhold som ikke fanges opp av observerte elevkjennetegn eller skolekvalitet. Det vesentlige i denne sammenheng er at restleddet er ukorrelert med begge disse, dvs. at uobserverte faktorer som påvirker elevresultater ikke samvarierer systematisk med skolekvalitet eller elevkjennetegn. Restleddet har også en forventning lik null.

Eksamensresultatet antas å avspeile underliggende ferdigheter, men ikke perfekt. Forskjellen representeres ved et tilfeldig restledd ($\upsilon$):

$$\begin{aligned} E &= F + v \\ &= X\gamma + \mu_j + (\varepsilon + v) \end{aligned} \tag{9}$$

Forskjellen mellom ferdigheter og eksamensresultat ($v$) kan blant annet skyldes at en eksamen neppe kan måle elevenes ferdigheter innen alle deler av faget. Dermed kan elevene på eksamen bli testet i noe de kan dårligere eller bedre enn øvrige

[12] Modellen er delvis basert på Galloway, Kirkebøen og Rønning (2011).

[13] For en enklere framstilling setter vi ikke indekser på variable for å vise at de varierer mellom elever, men indekserer heller eksplisitt de variablene som kun varierer mellom skoler, men er felles for alle elever innen hver skole.

deler av pensum. En elev kan også ha en dårlig dag på eksamen, og dermed prestere dårligere enn vedkommendes ferdigheter skulle tilsi. Vi antar imidlertid at det ikke er noen systematiske forskjeller mellom elever i så henseende, det er for eksempel ikke slik at elever med bedre ferdigheter eller på skoler med høyere kvalitet presterer bedre eller dårligere relativt til sine ferdigheter på eksamen.

Tilsvarende avpeiler også standpunktkarakterene ferdighetene, men i tillegg til tilfeldige avvik mellom ferdighet og karakter finnes også et skolespesifikt ledd, som fanger karakterpraksisen ved skolen:

$$\begin{aligned} S &= F + \theta_j + \omega \\ &= X\gamma + (\mu_j + \theta_j) + (\varepsilon + \omega) \end{aligned} \tag{10}$$

Vi antar at ferdighet er den samme som måles ved eksamen, og – som ved eksamen – kan det være usystematiske avvik mellom ferdighet og karakter. Men i tillegg til slik tilfeldig variasjon er det også systematiske forskjeller mellom skoler. En gitt ferdighet bedømmes ikke likt ved alle skoler og det finnes en (nivå)forskjell i karakterpraksis, jf. Galloway, Kirkebøen og Rønning (2011). Dette innebærer at den totale skoleeffekten, dvs. forskjellen i standpunktkarakterer mellom elever med tilsvarende kjennetegn på forskjellige skoler, fanger opp både kvalitet og karakterpraksis. Videre er det systematiske forskjeller mellom skoler, for eksempel finner Galloway, Kirkebøen og Rønning (2011) at skoler med faglig sterkere elever har en strengere karakterpraksis, dvs. gir en lavere karakter for en gitt prestasjon. Dette betyr at $\theta_j$ ikke varierer tilfeldig mellom skoler, men er (negativt) korrelert med elevenes ferdigheter (og muligens korrelert med skolekvalitet).

I prinsippet kan det finnes variasjon i karakterpraksis mellom skoler også ved skriftlig eksamen. Så lenge antall elever er begrenset kan en sensor som stiller spesielt høye eller lave krav bidra til at mange elever ved en skole får henholdsvis lave eller høye eksamenskarakterer i forhold til deres ferdigheter. Dette bør imidlertid jevne seg ut når vi har mange elever, enten via flere sensorer per skole og år, eller gjennom flere årskull bak hvert skoleresultat. Imidlertid, i den grad en enkelt sensor retter besvarelser fra bare en eller noen få skoler, vil elevene i stor grad sammenlignes med sine medelever. Dette kan gi opphav til relativ karaktersetting og forskjeller i karakterpraksis også på eksamen, selv om det er mange elever på hver skole. Det er likevel flere grunner til å tro at eventuelle systematiske skjevheter ved skriftlig eksamen er mindre enn ved fastsetting av standpunktkarakterer. Ved skriftlig eksamen vurderer to eksterne sensorer den samme skriftlige besvarelsen uavhengig av hverandre. Sensureringen avsluttes med en fellessensur, der sensorene møtes og setter endelig karakter. Dersom de to sensorene ikke er enige om karakteren, er det en tredje uavhengig sensor som involveres for å vurdere og beslutte karakteren. Til alle skriftlige eksamener foreligger en vurderingsveiledning med kjennetegn på måloppnåelse, og det gjennomføres sensorskoleringer i alle fag. Dette gir mindre rom for skjønn. Sensor kjenner heller ikke elevene, slik at utenomfaglige forhold i mindre grad kan spille noen rolle, og sensor er vant med andre elever og et annet faglig nivå, som er uavhengig av besvarelsene sensor vurderer. I den grad det er en skolekomponent også i skriftlig eksamenskarakter, kan vi tenke på $\theta_j$ som differansen mellom skolekomponenten i standpunktkarakter og skolekomponenten i skriftlig eksamenskarakter.

Det er verdt å påpeke at modellen som (8)-(10) utgjør ikke er noen presis eller fullstendig beskrivelse av sammenhengen mellom kjennetegn, ferdigheter og skoleresultater. Relevante innvendinger er for eksempel at det ikke nødvendigvis er eksakt de samme ferdighetene som ligger til grunn for standpunkt- og eksamenskarakterer, samt at det kan være en mer komplisert sammenheng mellom ferdighet og resultat (og potensielt forskjellig sammenhenger for standpunkt og eksamen). Det kan også være mer systematisk variasjon i resultater, for eksempel

dersom det er tilfeldige begivenheter som påvirker en hel klasse eller skoles prestasjoner, som når undervisningen i mindre grad har vektlagt temaene som kommer til eksamen. Modellen er ment å forenkle sammenhengen mellom kjennetegn, ferdighet og resultater mest mulig, uten at det går på bekostning av evnen til å illustrere hva dette kan bety.

I de neste avsnittene ser vi nærmere på hvordan vi kan bruke observert variasjon i standpunkt og eksamen, både på elev- og skolenivå som grunnlag for vurdering av hvilken rolle forskjeller i karaktersetting spiller. Vi viser først hva modellen impliserer for spredning i resultater for enkeltelever. Variansen til hhv. *E* og *S* på elevnivå er vist i (11) og (12):

(11) $\text{var}(E) = \text{var}(F) + \text{var}(\nu)$

(12) $\text{var}(S) = \text{var}(F) + \text{var}(\theta_j) + \text{var}(\omega) + \text{cov}(F, \theta_j)$

Vi ser dermed at

(13) $\text{var}(E) > \text{var}(S) \Leftrightarrow [\text{var}(\nu) - \text{var}(\omega)] > [\text{var}(\theta_j) + \text{cov}(F, \theta_j)]$

Spredningen i eksamenskarakter skyldes spredning i ferdighet og tilfeldigheter i evalueringen av ferdighet til eksamen. Variansen til standpunktkarakterene har tilsvarende ledd, men betydningen av tilfeldigheter i karaktersettingen kan være ulik. I tillegg kommer spredning fra karakterpraksis. Det er rimelig å anta at standpunktkarakterer mer presist oppsummerer den underliggende ferdigheten, om vi ser bort fra forskjeller i karakterpraksis, fordi faglærerne i mye større grad observerer elevene over lang tid, og har større mulighet til å danne seg et bilde av deres ferdigheter enn hva en eksamenssensor har. Altså vil spredningen være størst ved eksamen med mindre det finnes innflytelse fra ulik karakterpraksis mellom skoler. I data viser det seg at standpunkt- og eksamenskarakterer har omtrent like stor spredning, hvilket gir oss en klar pekepinn om at det finnes forskjeller i karakterpraksis mellom skoler. Men utfra forskjeller i spredning på elevnivå er det umulig å si hvor viktig den er.

Vårt fokus er hva forskjeller i karakterpraksis kan bety for de estimerte value added-indikatorene. Skolebidrag basert på eksamenskarakter ( $SBI_j^E$ er dette bidraget for skole *j*) estimeres fra en ligning som (14):

(14) $E = \alpha + X\beta + SBI_j^E + u$

Under nærmere antakelser er vårt beregnede SBI basert på eksamen *forventningsrett*, noe som innebærer at i gjennomsnitt og i den grad vi har mange elevobservasjoner ved hver enkelt skole gir disse et riktig bilde av skolekvaliteten, $\mu_j$:[14] $\text{E}(\hat{SBI}_j^E) = \mu_j$

For SBI basert på standpunktkarakter er bildet mer komplisert, denne fanger i forventning *summen* av skolekvalitet og karakterpraksis: $\text{E}(\hat{SBI}_j^S) = \mu_j + \theta_j$

Spredningen i de estimerte SBI-ene vil kunne gi oss pekepinn om betydningen av karakterpraksis, dvs. hvor stor del av standpunktindikatorene som kan tilskrives karakterpraksis. Variansene blir, når vi antar at (gjennomsnittet av) restleddene $\varepsilon$, $\nu$ og $\omega$ er ukorrelerte med kvalitet og karakterpraksis lik:

[14] I vedlegg A viser vi under hvilke antagelser dette gjelder. Merk at E(*X*) er forventningen til *X*, og ikke har noe med eksamensresultatet, *E*, å gjøre.

(15) $$\text{var}(\hat{SBI}_j^E) = \text{var}(\mu_j) + \text{var}(\bar{v}_j) + \text{var}(\bar{\varepsilon}_j)$$

(16) $$\text{var}(\hat{SBI}_j^S) = \text{var}(\mu_j) + \text{var}(\theta_j) + \text{var}(\bar{\omega}_j) + \text{var}(\bar{\varepsilon}_j) + 2\,\text{cov}(\mu_j, \theta_j)$$

(17) $$\text{var}(\hat{SBI}_j^E) - \text{var}(\hat{SBI}_j^S) = [\text{var}(\bar{v}_j) - \text{var}(\bar{\omega}_j)] - [\text{var}(\theta_j) + 2\,\text{cov}(\mu_j, \theta_j)]$$

Variansen til de gjennomsnittlige restleddene $\bar{v}_j$, $\bar{\omega}_j$ og $\bar{\varepsilon}_j$ avtar raskt med antall elevobservasjoner ved skolen. Når vi ser på skoler med et visst antall elever skulle leddene med gjennomsnittet av tilfeldigheter gå mot null. Ettersom vi observerer større spredning i SBI basert på eksamen enn basert på standpunkt må det bety det finnes en negativ samvariasjon mellom skolekvalitet og karakterpraksis, som er betydelig sammenlignet med spredningen i karakterpraksis. Alternativt kan variansen til $\bar{v}_j$ være større enn variansen til $\bar{\omega}_j$, men ettersom variansen til restleddene avtar raskt med antall observasjoner må spredningen (på elevnivå) i tilfeldigheter i eksamenskarakteren være svært mye større enn hva som er tilfelle for standpunkt for å gi merkbare utslag. Vi står dermed igjen med tilfeldigheter på skolenivå ved eksamen som et alternativ, eller at karakterpraksis i noen grad henger systematisk sammen med skolekvalitet.

Et utgangspunkt for drøftingen av karakterpraksis var den overraskende lave samvariasjonen mellom SBI beregnet ved henholdsvis eksamen og standpunkt. Hvordan kan den forstås i lys av rammeverket her? Samvariasjonen mellom de to beregnede SBI-ene er, igjen under en antagelse om at restleddene er uavhengige av andre variable:

(18) $$\text{cov}(\hat{SBI}_j^E, \hat{SBI}_j^S) = \text{var}(\mu_j) + \text{cov}(\mu_j, \theta_j)$$

Ellers i rapporten fokuserer vi på korrelasjoner, ettersom disse ikke avhenger av skalaen til variablene, og dermed er lettere å tolke. Vi finner korrelasjonen fra kovarians og varians, slik at korrelasjonen mellom beregnede SBI blir:

(19)

$$\begin{aligned} corr(\hat{SBI}_j^E, \hat{SBI}_j^S) &= \frac{\text{cov}(\hat{SBI}_j^E, \hat{SBI}_j^S)}{\sqrt{\text{var}(\hat{SBI}_j^E)\,\text{var}(\hat{SBI}_j^S)}} \\ &= \frac{\text{var}(\mu_j) + \text{cov}(\mu_j, \theta_j)}{\sqrt{\left(\text{var}(\mu_j) + \text{var}(\bar{v}_j) + \text{var}(\bar{\varepsilon}_j)\right)\left(\text{var}(\mu_j) + \text{var}(\theta_j) + \text{var}(\bar{\omega}_j) + \text{var}(\bar{\varepsilon}_j) + 2\,\text{cov}(\mu_j, \theta_j)\right)}} \\ &\approx \frac{\text{var}(\mu_j) + \text{cov}(\mu_j, \theta_j)}{\sqrt{\text{var}(\mu_j)\left(\text{var}(\mu_j) + \text{var}(\theta_j) + 2\,\text{cov}(\mu_j, \theta_j)\right)}} \end{aligned}$$

I det siste leddet har vi antatt at de leddene som avhenger av de gjennomsnittlige restleddene har en beskjeden betydning og satt lik null, ettersom disse avtar raskt når elevtallet øker. Vi ser at ved fravær av forskjeller i karakterpraksis vil variansen til indikatorer basert både på standpunkt- og eksamenskarakter, samt kovariansen mellom dem, være lik variansen til skolekvalitet. Da vil også korrelasjonen mellom SBI beregnet på grunnlag av eksamen og standpunkt være lik én, dvs. perfekt samsvariasjon.

Dersom vi har forskjeller i karakterpraksis vil dette imidlertid bidra til å redusere samvariasjonen. Hvis vi antar lik varians for kvalitet og karakterpraksis, men at disse er ukorrelert får vi en korrelasjonskoeffisient på 0,5. Dersom det er en negativ sammenheng mellom skolekvalitet og karakterpraksis, det vil si dersom skoler med lite bidrag gjennomgående gir bedre karakter for tilsvarende ferdigheter, vil korrelasjonen falle ytterligere.

For å komme nærmere et anslag på betydningen av forskjeller i karakterpraksis trenger vi et mål på skolekvalitet. Ettersom vi har argumentert tidligere for at SBI basert på eksamenskarakterer er forventningsrette, er disse et naturlig valg. Vi kan dermed bruke differansen mellom *S* og *E* til å estimere forskjeller i karakterpraksis. Fra (9) og (10) får vi at denne er:

$$S - E = \theta_j + \omega - \nu \quad (20)$$

Ettersom restleddene $\omega$ og $\upsilon$ – og dermed også differansen mellom dem – er antatt å være uavhengige av karakterpraksisen kan vi estimere karakterpraksisen direkte som en fast effekt i (20), dvs. som den gjennomsnittelige forskjellen mellom standpunkt og eksamen på hver skole. Denne uttrykker hvor store forskjeller det er i bedømming av en gitt ferdighet (målt ved eksamen) i form av karakterpoeng til standpunktkarakteren. Et positivt tall svarer dermed til høye karakterer for en gitt ferdighet.
Fra de tidligere resultatene har vi sett at standardavviket for estimerte indikatorer basert på or både skriftlig eksamen (som under forutsetningene over gir forventningsrette estimater) og standpunktkarakterer er i underkant av 0,2 karakterpoeng. Når vi estimerer ligning (20) finner vi til sammenligning at standardavviket til estimert karakterpraksis er 0,26, dvs., vi finner større forskjeller i karakterpraksis enn i skolekvalitet. Videre er karakterpraksisen klart negativt korrelert med skolekvalitet (målt ved estimert skolebidrag for skriftlig eksamen), med en korrelasjonskoeffisient på -0,54.[15] Når vi på bakgrunn av de beregnede variansene og kovariansene beregner den forventede korrelasjonen mellom estimerte indikatorer for skriftlig eksamen og standpunkt får vi en korrelasjonskoeffisient på 0,27. Dette er litt høyere enn hva vi faktisk finner for de beregnede indikatorene i forrige avsnitt (omtrent 0,2), men samsvarer likevel forholdsvis godt med den overraskende lave korrelasjonen.

Disse beregningene er basert på en del forutsetninger, blant annet at standpunkt og eksamen måler eksakt samme ferdighet, at det ikke er noen skolespesifikk karakterpraksis på eksamen og at vi fullt ut er i stand til å kontrollere for elevsammensetning. Av den grunn bør de tolkes med en viss varsomhet. Imidlertid illustrerer de at forskjeller i karakterpraksis kan være betydelige, og ha store konsekvenser når vi forsøker å måle skolekvalitet. Dersom en ønsker å bruke skoleresultater som grunnlag for måling av skolekvalitet, er det nødvendig å ta hensyn til elevsammensetning. Men når vi etter dette står igjen med relativt små forskjeller mellom skoler, vil ulik karakterpraksis kunne utgjøre en betydelig feilkilde. Ettersom skriftlig eksamenskarakter sannsynligvis er mindre utsatt for systematiske forskjeller i karakterpraksis, trekker dette i retning av kvalitetsmål basert på eksamenskarakterer, eller mer generelt, en form for ekstern evaluering med minst mulig skjønn. Dette fordrer at et tilstrekkelig antall elever faktisk avlegger eksamen ved hver enkelt skole, særlig ettersom eksamen sannsynligvis er mindre presis som mål på elevenes kunnskaper.

[15] Estimerte skolebidragsindikatorer for standpunkt er derimot positivt korrelert med estimert karakterpraksis, med en korrelasjonskoeffisient på 0,57.

# 8. Indikatorer for barnetrinnet

For nasjonale prøver på 5. trinn har vi ingen informasjon om tidligere resultater. Det er derfor ikke mulig å beregne value added-indikatorer med dette som resultatmål. Vi kan imidlertid beregne skolebidragsindikatorer hvor vi kontrollerer for familiebakgrunn, det vil si tverrsnittsindikatorer. Vi gjør separate analyser for gjennomsnittsskår for alle de tre prøvene (lesing, regning, engelsk) og separat for hvert emne. For hvert av disse resultatmålene sammenligner vi tre forskjellige indikatorer: Ujusterte gjennomsnitt, resultater korrigert for et enkelt sett med familiebakgrunnsvariable (tilsvarende det som har blitt benyttet i de andre analysene i denne rapporten) og korrigert for et detaljert sett med familiebakgrunnsvariable (tilsvarende det som har blitt benyttet ved tidligere beregninger av skolebidragsindikatorer, og som også er benyttet i noen av analysene i kapittel 7). Nedenfor presenterer vi resultatene gjennom tabeller og figurer på samme måte som i de foregående kapitlene.

Fra Tabell 8.1 ser vi at elevbakgrunn målt med familiebakgrunnskjennetegn forklarer relativt lite av variasjonen i elevresultater på nasjonale prøver for 5. trinn. Andel forklart variasjon er langt lavere enn for resultater målt lenger ut i skoleløpet. Et større sett med familiebakgrunnsvariable bidrar til å gi modellen økt forklaringskraft, men ikke veldig mye. Ettersom det ikke ser ut til å ha særlig betydning å kontrollere for detaljert bakgrunn er vår foretrukne spesifikasjon den enkle, dvs. med kontroll for et "lite" sett av familiebakgrunnsvariable.

**Tabell 8.1. Regresjonsresultater, basert på resultater nasjonale prøver 5. trinn**

| | (1) Snitt NP5 | (2) Snitt NP5 | (3) Lesing NP5 | (4) Lesing NP5 | (5) Regning NP5 | (6) Regning NP5 | (7) Engelsk NP5 | (8) Engelsk NP5 |
|---|---|---|---|---|---|---|---|---|
| Mangler res. NP5 engelsk .... | -0,475*** | -0,460*** | | | | | | |
| | (0,0146) | (0,0144) | | | | | | |
| Mangler res. NP5 regning .... | -0,0730*** | -0,0569*** | | | | | | |
| | (0,0153) | (0,0150) | | | | | | |
| Mangler res. NP5 lesing ..... | -0,169*** | -0,168*** | | | | | | |
| | (0,0113) | (0,0112) | | | | | | |
| Jente ..... | 0,0112* | 0,0133** | 0,181*** | 0,182*** | -0,158*** | -0,157*** | 0,0229*** | 0,0230*** |
| | (0,00460) | (0,00454) | (0,00575) | (0,00569) | (0,00491) | (0,00486) | (0,00587) | (0,00582) |
| Konstantledd .... | 2,979*** | 1,862*** | 2,838*** | 2,014*** | 2,657*** | 2,056*** | 3,392*** | 2,998*** |
| | (0,00405) | (0,349) | (0,00499) | (0,477) | (0,00421) | (0,387) | (0,00505) | (0,494) |
| Familiebakgrunn ..... | Enkel | Detaljert | Enkel | Detaljert | Enkel | Detaljert | Enkel | Stort |
| Forklaringskraft ($R^2$) ..... | 0,111 | 0,140 | 0,102 | 0,127 | 0,106 | 0,131 | 0,0760 | 0,0989 |
| Antall elever ..... | 117 865 | 117 865 | 110 609 | 110 609 | 114 851 | 114 851 | 114 459 | 114 459 |

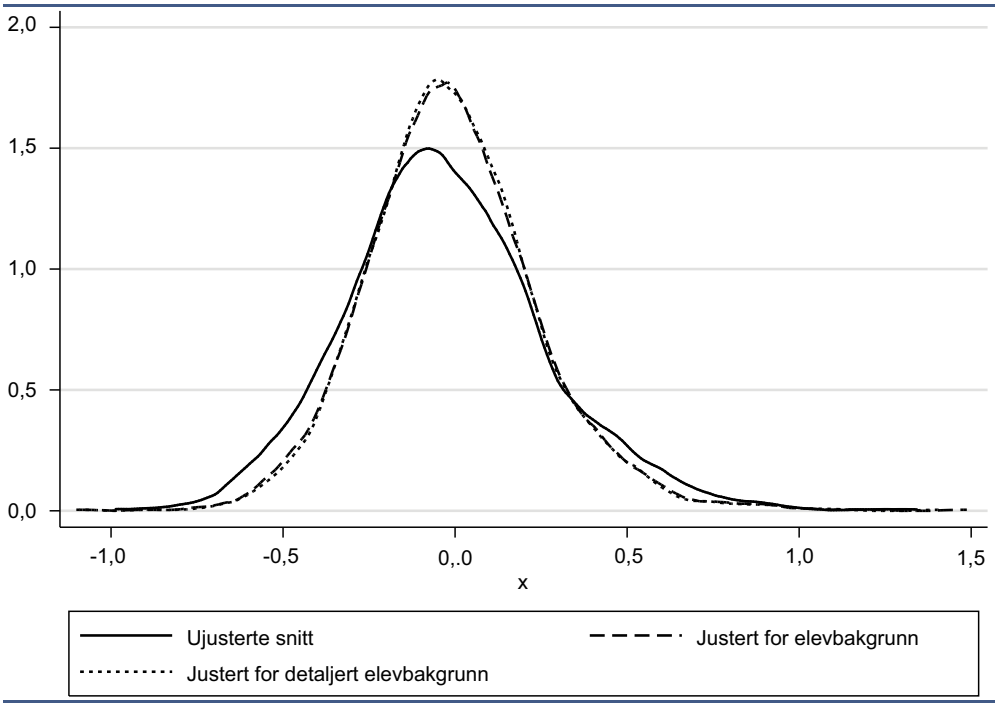
Estimerte standardfeil i parentes. Statistisk signifikans: * p<0,05, ** p<0,01, *** p<0,001

Tabell 8.2 og Figur 8.1 viser at resultatforskjellene mellom skoler blir noe mindre når vi kontrollerer for forskjeller i elevbakgrunn mellom skoler, standardavviket på skolenivå reduseres fra omtrent 0,3 (elevnivå) standardavvik til omtrent 0,25. Det har svært lite å si om vi bruker en enkel eller detaljert spesifikasjon av familiebakgrunn, fordelingene av indikatorene basert på disse to spesifikasjonene er omtrent identiske.

**Tabell 8.2. Beskrivende statistikk indikatorer basert på resultat nasjonale prøver 5. trinn (elevvektet)**

| | Gj. snitt | Std. avvik | Min. | 10. per-sentil | 25. per-sentil | 50. per-sentil | 75.per-sentil | 90. per-sentil | Maks. |
|---|---|---|---|---|---|---|---|---|---|
| **Snitt alle prøver** | | | | | | | | | |
| Snittresultat ..... | 0,008 | 0,290 | -0,986 | -0,340 | -0,184 | -0,013 | 0,176 | 0,406 | 1,342 |
| Enkel kontroll for FB ... | 0,003 | 0,236 | -1,091 | -0,286 | -0,153 | -0,010 | 0,143 | 0,290 | 1,487 |
| Detaljert kontroll for FB | 0,002 | 0,231 | -1,101 | -0,278 | -0,147 | -0,016 | 0,139 | 0,292 | 1,419 |
| Antall skoler ..... | 1 667 | | | | | | | | |
| **Lesing** | | | | | | | | | |
| Snittresultat ..... | 0,006 | 0,298 | -1,016 | -0,345 | -0,188 | -0,000 | 0,191 | 0,392 | 1,091 |
| Enkel kontroll for FB ... | 0,002 | 0,240 | -1,021 | -0,290 | -0,154 | -0,009 | 0,154 | 0,300 | 1,141 |
| Detaljert kontroll for FB | 0,001 | 0,233 | -1,032 | -0,283 | -0,152 | -0,005 | 0,148 | 0,296 | 1,208 |
| Antall skoler ..... | 1 603 | | | | | | | | |
| **Regning** | | | | | | | | | |
| Snittresultat ..... | 0,010 | 0,313 | -1,072 | -0,374 | -0,196 | -0,015 | 0,198 | 0,435 | 1,262 |
| Enkel kontroll for FB ... | 0,004 | 0,267 | -1,038 | -0,320 | -0,180 | -0,006 | 0,169 | 0,337 | 1,248 |
| Detaljert kontroll for FB | 0,004 | 0,264 | -1,031 | -0,316 | -0,175 | -0,006 | 0,170 | 0,337 | 1,210 |
| Antall skoler ..... | 1 639 | | | | | | | | |
| **Engelsk** | | | | | | | | | |
| Snittresultat ..... | 0,010 | 0,329 | -1,147 | -0,377 | -0,209 | -0,016 | 0,197 | 0,439 | 1,917 |
| Enkel kontroll for FB ... | 0,005 | 0,288 | -1,202 | -0,328 | -0,191 | -0,007 | 0,168 | 0,374 | 1,799 |
| Detaljert kontroll for FB | 0,004 | 0,283 | -1,235 | -0,325 | -0,187 | -0,009 | 0,162 | 0,364 | 1,687 |
| Antall skoler ..... | 1 636 | | | | | | | | |

**Figur 8.1. Fordeling av ulike resultatmål basert på gjennomsnittskår nasjonale prøver 5. trinn**



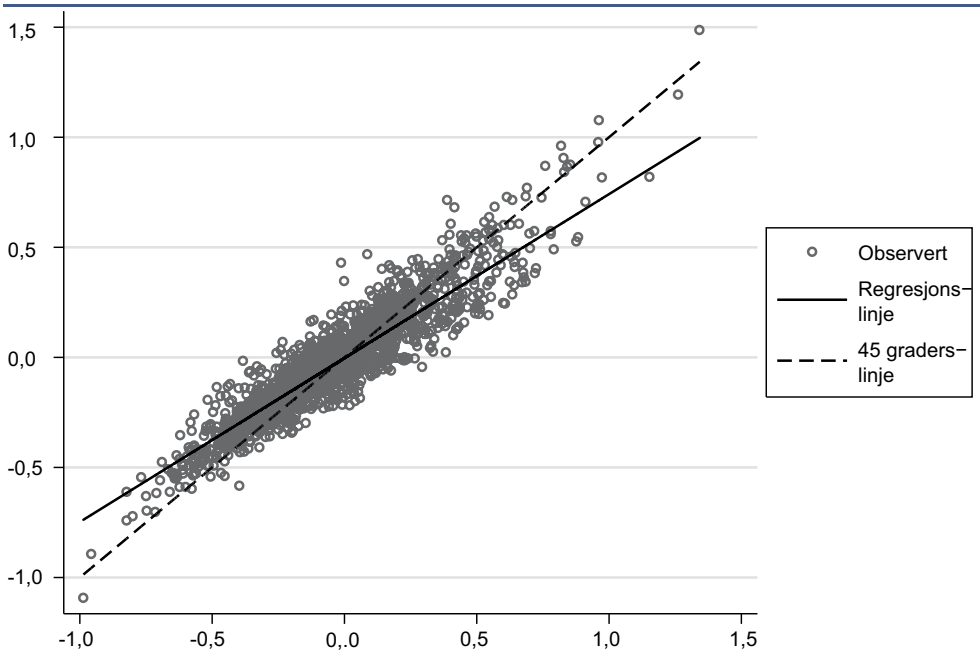

Bildet av hvilke skoler som gjør det godt og dårlig endres i noen grad når vi kontrollerer for familiebakgrunn, men noe mindre enn tilsvarende indikatorer beregnet for resultater lenger ut i skoleløpet. Fra Tabell 8.3 ser vi at korrelasjonen mellom de forskjellige indikatorene og ujusterte gjennomsnitt er omtrent 0,9, unntatt for engelsk, som har en høyere korrelasjon – omtrent 0,95. Figur 8.2 gir en visuell framstilling av dette for vår foretrukne indikator (enkel spesifikasjon av familiebakgrunn). Vi ser at det er en klar sammenheng mellom indikatoren og ujusterte gjennomsnitt.

Det har svært lite å si om vi bruker en enkel eller detaljert spesifikasjon av familiebakgrunn, jf. Tabell 8.3 og Figur 8.3, og de to indikatorene gir omtrent identiske resultater. Korrelasjonene er over 0,99 både for gjennomsnittsresultat og for hver av prøvene separat, jf. Tabell 8.3 og Figur 8.2.

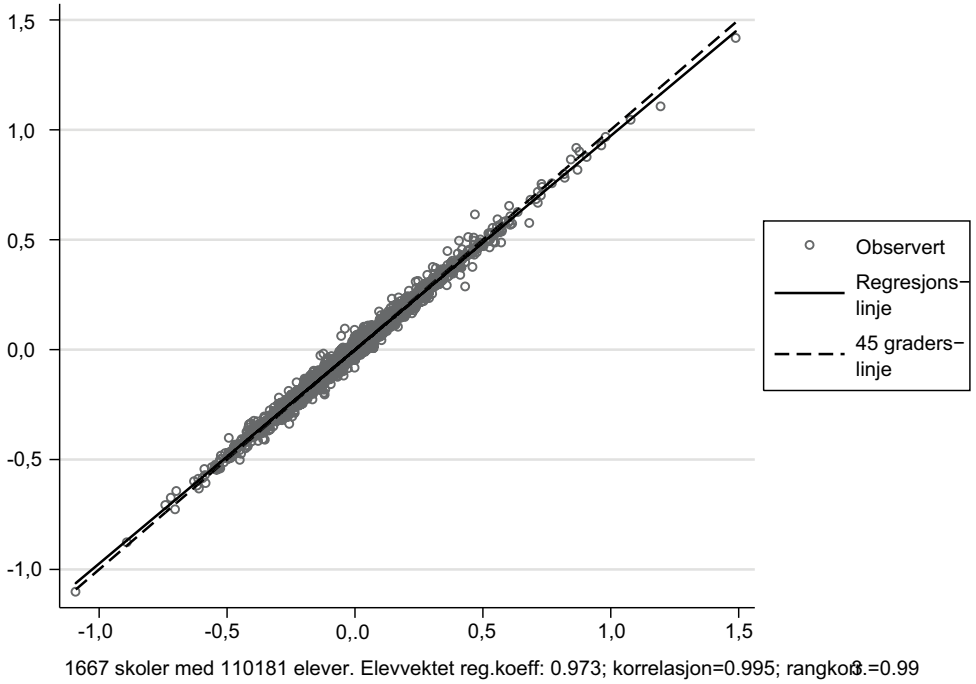
**Tabell 8.3. Korrelasjon mellom ulike indikatorer, innen resultatmål, nasjonale prøver 5. trinn**

| Korrelasjon | Snitt | Lesing | Regning | Engelsk |
|---|---|---|---|---|
| Snitt og indikator med kontroll for enkel elevbakgrunn ..... | 0,913 | 0,888 | 0,919 | 0,960 |
| Snitt og indikator med kontroll for detaljert elevbakgrunn ..... | 0,898 | 0,867 | 0,907 | 0,949 |
| Indikatorer med hhv. enkel og detaljert elevbakgrunn ..... | 0,995 | 0,993 | 0,996 | 0,996 |

**Figur 8.2 Sammenheng mellom foretrukket indikator og ujustert resultat, gjennomsnittskår nasjonale prøver 5. trinn**




**Figur 8.3. Sammenheng mellom indikatorer med stor og liten korreksjon for familiebakgrunn, gjennomsnittskår nasjonale prøver 5. trinn**




Det er relativt stor samvariasjon på tvers av fag, både for ujusterte resultater og for skolebidragsindikatorer. Fra Tabell 8.4 ser vi at korrelasjonskoeffisientene mellom de ujusterte gjennomsnittene for de forskjellige prøvene er omkring 0,75, og alle har en korrelasjon med snittet av alle prøver på omtrent 0,9. Fra Tabell 8.5 ser vi at tilsvarende verdier for de estimerte indikatorene er hhv. 0,6-0,7 og 0,8-0,9.

Indikatorene er altså lavere korrelert enn de ujusterte snittene, men fortsatt høyt korrelert, for eksempel sammenlignet med indikatorene for mellomtrinnet.

**Tabell 8.4. Korrelasjon mellom ujusterte gjennomsnitt, nasjonale prøver 5. trinn**

| | Snitt NP5 | Lesing NP5 | Regning NP5 | Engelsk NP5 |
|---|---|---|---|---|
| Snitt NP5 ..... | 1 | | | |
| Lesing NP5 ..... | 0,896 | 1 | | |
| Regning NP5 ..... | 0,918 | 0,756 | 1 | |
| Engelsk NP5 ..... | 0,911 | 0,731 | 0,747 | 1 |

**Tabell 8.5. Korrelasjon mellom foretrukket indikator på tvers av fag, nasjonale prøver 5. trinn**

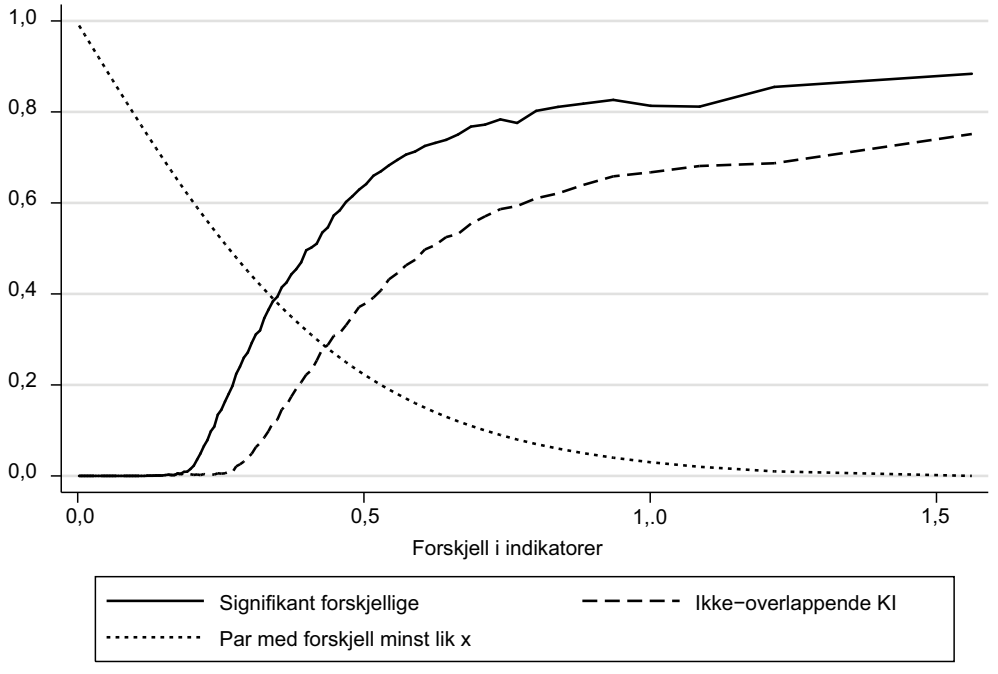
| | Snitt NP5 | Lesing NP5 | Regning NP5 | Engelsk NP5 |
|---|---|---|---|---|
| Snitt NP5 ..... | 1 | | | |
| Lesing NP5 ..... | 0,858 | 1 | | |
| Regning NP5 ..... | 0,880 | 0,666 | 1 | |
| Engelsk NP5 ..... | 0,897 | 0,681 | 0,680 | 1 |

## 8.1. Usikkerhet i indikatorene

I Figur 8.4 til Figur 8.7 presenterer vi den statistiske usikkerheten i de estimerte indikatorene for barnetrinnet, som vi tidligere har gjort for mellomtrinnet og ungdomstrinnet. Figur 8.4 viser usikkerheten i indikatorene basert på snittet av alle prøvene. Omtrent 40 prosent av skoleparene har en forskjell i estimert indikator på mindre enn 0,2 standardavvik, og ingen av disse er statistisk signifikante. Selv om andelen signifikante forskjeller øker med størrelsen på forskjellen, ser vi at det er betydelig usikkerhet også for de største forskjellene: Blant de omtrent 10 prosent av forskjellene større enn 0,7 standardavvik er det bare ca 80 prosent som er statistisk signifikante. Andelene med ikke-overlappende konfidensintervall er, som for tidligere presenterte indikatorer, enda lavere.

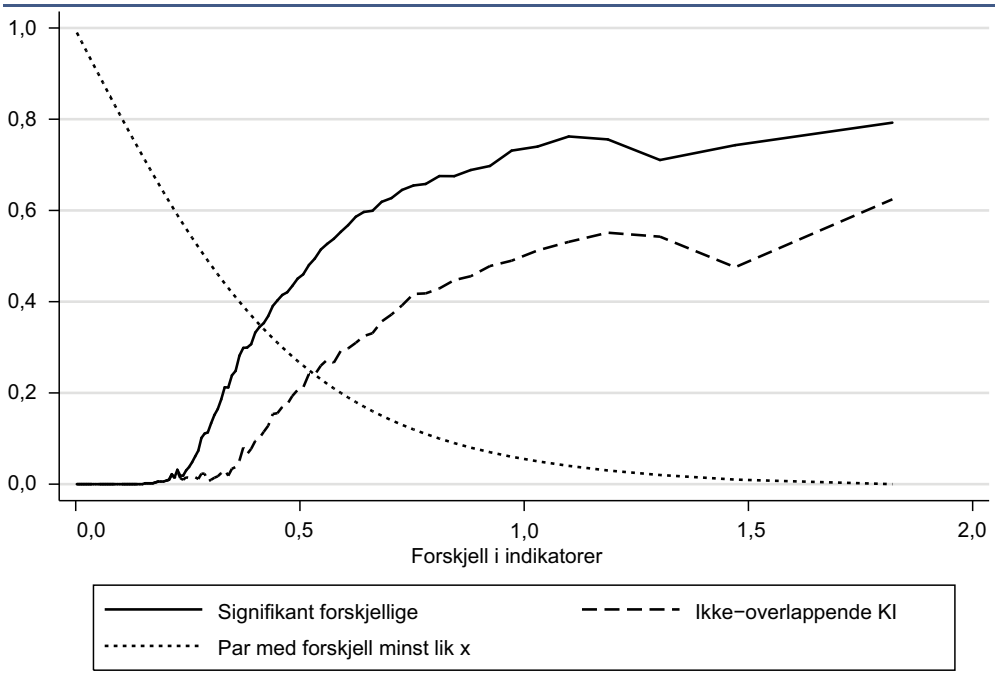
Figur 8.5, Figur 8.6 og Figur 8.7 viser den statistiske usikkerheten i indikatorene basert på hhv. lesing, regning og engelsk. Disse har et forløp som ligner det vi så i Figur 8.4, men særlig for lesing (Figur 8.5) og i noen grad for engelsk (Figur 8.7) er det en enda lavere andel av forskjellene som er signifikante.

**Figur 8.4. Fordeling av skoleforskjeller og statistisk usikkerhet. Indikatorer basert på gjennomsnittet av prøvene.**



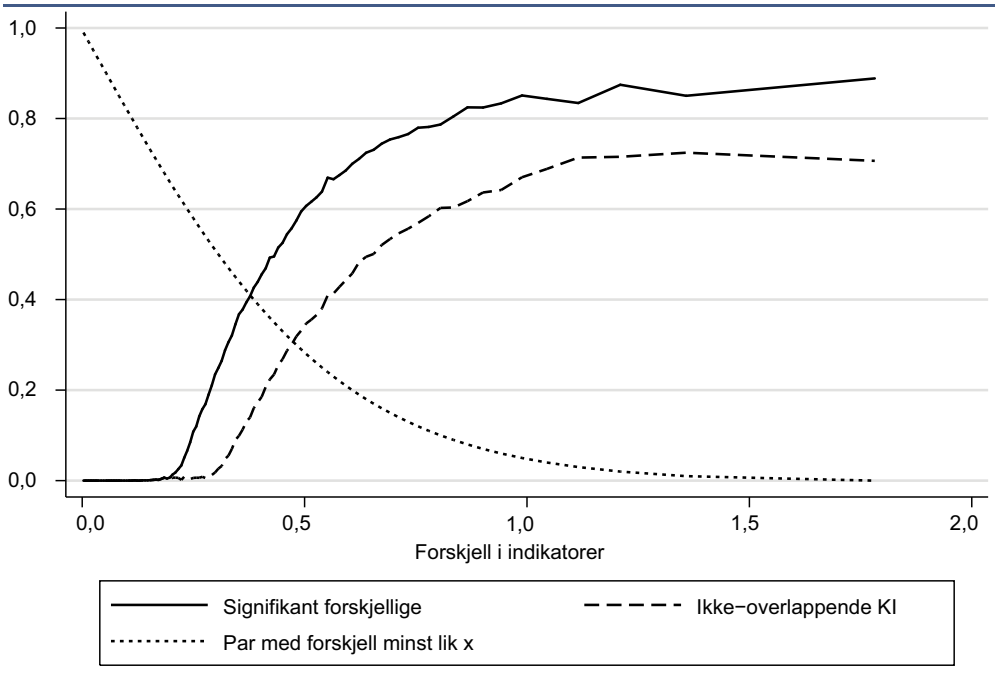

Andeler per persentil, 1619 skoler (1309771 par)

**Figur 8.5. Fordeling av skoleforskjeller og statistisk usikkerhet. Indikatorer basert på prøve i lesing**



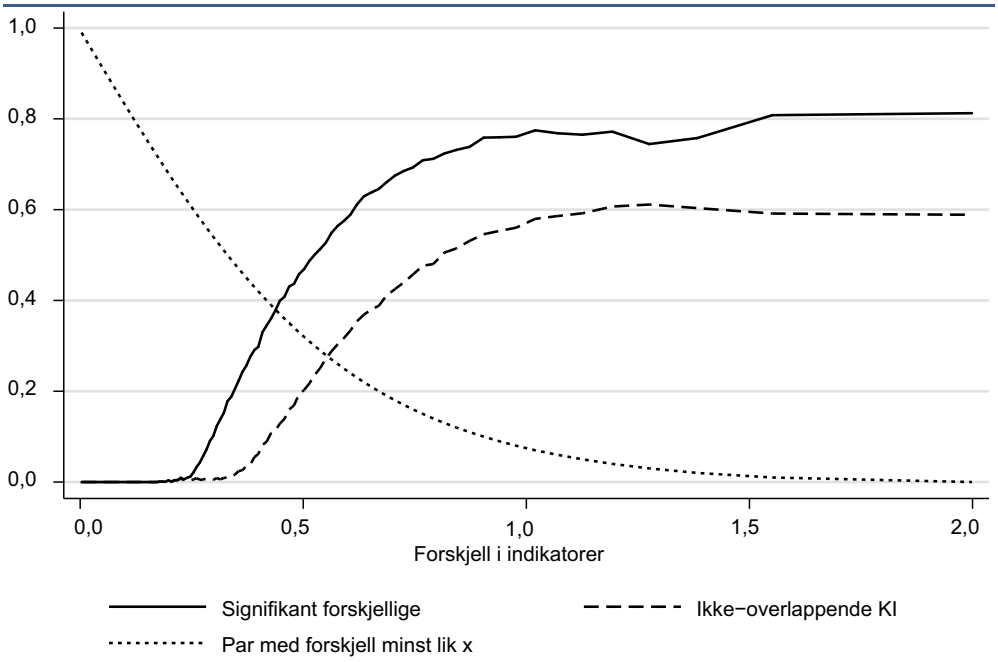

Andeler per persentil, 1556 skoler (1209790 par)

**Figur 8.6. Fordeling av skoleforskjeller og statistisk usikkerhet. Indikatorer basert på prøve i regning**




Andeler per persentil, 1591 skoler (1264845 par)

**Figur 8.7. Fordeling av skoleforskjeller og statistisk usikkerhet. Indikatorer basert på prøve i engelsk**




Andeler per persentil, 1588 skoler (1260078 par)

# 9. Videregående skoler

I en viss forstand er det relativt enkelt å lage resultatindikatorer for grunnskolen, siden alle skolene produserer det samme "produktet" (alle elevene fullfører per definisjon, og de har de samme fagene på vitnemålet). Videregående skoler kjennetegnes derimot av stor heterogenitet. Et kompliserende element for analyse av resultater er at elevene selv velger utdanningsprogram, og i noen grad enkeltfag innenfor det valgte programmet. Ettersom valgene avspeiler ulike yrkes- og karrierevalg vil det være forskjellige grupper av elever som velger de forskjellige programmene og fagene. Dette kommer klart fram i Hægeland, Kirkebøen, og Raaum (2006). Slike forhold må tas i betraktning når man lager indikatorer for videregående skole. Grovt sett finnes det to alternativer. En mulighet er å gjøre skolene så sammenlignbare som mulig, ved å kontrollere for fagsammensetning og lignende, og beregne indikatorer for et bredt sett av skoler under ett. I mange tilfeller vil imidlertid ikke dette være meningsfullt, fordi skolene er for ulike til at det er hensiktsmessig å sammenligne dem. Det vil da være mer hensiktsmessig å følge den andre fremgangsmåten der man beregner indikatorer separat for ulike sett av skoler som internt er mer sammenlignbare.

Settet av mulige resultatindikatorer er større for videregående skole enn for grunnskolen. Hægeland, Kirkebøen, Raaum og Salvanes (2006) så på indikatorer som var basert på karakterer i enkeltfag eller et sett av enkeltfag. Det er selvsagt interessant og relevant å studere resultatforskjeller mellom videregående skoler langs denne dimensjonen, men for videregående skole er imidlertid også problematikken knyttet til frafall og manglende fullføring svært relevant (i motsetning til for grunnskolen, hvor så å si alle fullfører). Eventuelle forskjeller mellom skoler langs denne dimensjonen fanges ikke godt nok opp via indikatorer som baserer seg utelukkende på karakterer.

Hægeland, Kirkebøen og Raaum (2010) analyserte value added-indikatorer for videregående skoler i Oslo, basert på karakterer og gjennomføring blant elever på studieforberedende og yrkesfaglige programmer. Rapporten drøftet noen sentrale egenskaper til indikatorene langs de samme linjene som denne rapporten vurderer indikatorer for grunnskolen, og drøftet i hvilken utstrekning indikatorene endres ved å inkludere familiebakgrunnsinformasjon. Konklusjonene med hensyn til hvilke variable det er hensiktsmessig å inkludere i modellene er i stor grad sammenfallende med hva vi har funnet for ulike trinn i denne rapporten. Siden Oslo er et stor kommune med opp mot ti prosent av det samlede elevtall i Norge, og i tillegg har en bredt sammensatt elevmasse, er det derfor grunn til å tro at funnene herfra knyttet til modellspesifikasjon lar seg generalisere til landet som helhet.

Med dette mener vi ikke at resten av landet er som Oslo, men at Oslo omfatter et stort spenn av skoler, og at de fleste forhold vi er interessert i å studere kan gjenfinnes innen Oslo-skolen. Hægeland, Kirkebøen, Raaum og Salvanes (2005b) studerer tverrsnittsindikatorer for grunnskolen, og finner at det er større spredning i elevsammensetning i Oslo enn i resten av landet. Oslo-skolen omfatter både skoler med svært fordelaktig elevbakgrunn med hensyn til resultater, skoler med utfordrende elevbakgrunn og mer gjennomsnittlige skoler, og Oslo-skolen omfatter skoler som presterer både sterkt og svakt. Samtidig finner Hægeland, Kirkebøen, Raaum og Salvanes (2005b) at det har svært liten betydning, både for estimerte sammenhenger mellom familiebakgrunn og resultater og for beregnede indikatorer, hvorvidt analysene gjøres for hele landet eller separat for Oslo-skolen.

Innenfor rammen av denne rapporten har vi derfor valgt å konsentrere de empiriske analysene om resultatmål for grunnskolen, men refererer hovedfunnene fra Hægeland, Kirkebøen og Raaum (2010) nedenfor.

Når vi sammenliknet resultater i videregående skoler, ga karakterene fra 10. trinn et godt mål på elevenes kunnskapsnivå ved skolestart, det er liten gevinst av å benytte ytterligere data om familiebakgrunn. Skolebidragsindikatorer der elevutfall er justert for kjønn, standpunkt basisfag og skriftlig eksamen i 10. trinn er en robust beregningsmåte som kan implementeres av skoleeier uten tilleggsinformasjon fra eksterne kilder. Enkeltskolers plassering i fordelingen av skolebidragsindikatorer påvirkes noe av beregningsmåte. Selv om hovedmønsteret er robust understreker dette at justeringen gir oss en indikator, men ikke et entydig, presist mål på skolens resultatkvalitet.

Når vi analyserer karakterdataene, finner vi at justeringen for ulik elevsammensetning på tvers av skoler med hensyn til resultater fra grunnskolen er svært viktig. Det er ingen klar sammenheng mellom skolenes karakterergjennomsnitt og deres skolebidragsindikator for karakterer i Vg1. Også for gjennomstrømningsindikatorene finner vi at justering for resultater fra grunnskolen er svært viktig ved sammenlikning av gjennomstrømning på tvers av skoler. Likevel finner vi en positiv samvariasjon mellom skolenes gjennomstrømning og den tilsvarende skolebidragsindikatoren. Det er betydelig usikkerhet knyttet til indikatorene for gjennomstrømning. Dette innebærer for en del indikatorer at vi også for sammenlikninger der skolebidragsindikatorene er svært ulike må ha kjennskap til presisjonen i anslaget for hver enkelt skole.

Skolebidragsindikatorer for karakterer og gjennomstrømning er ikke sammenfallende, men ser ut til å fange opp forskjellige kvaliteter ved skolene. Dermed finnes det ikke noen enkeltindikator som oppsummerer skolekvalitet, i stedet kan ulike indikatorer øke informasjonsmengden for skoleeiere, skoleledere, ansatte, elever og foreldre. For studieforberedende finner vi at skoler med gode (dårlige) karakterer har gjennomgående høy (lav) fullføring. Etter korreksjon for elevsammensetning i de tilsvarende skolebidragsindikatorene finner vi lavere korrelasjon, som også varierer mellom kohorter. På skolene med yrkesfag er det også en klar tendens til at de skolene som har høye ujusterte karaktersnitt også har en høy andel elever som fullfører første skoleår. Likevel er det ingen samvariasjon mellom SBI for karakterer og andel som fullfører første år. Dette står i sterk kontrast til studieforberedende, der skoler som lyktes godt med karakterer gjennomgående hadde høy fullføring.

I lys av diskusjonene rundt karakterpraksis i kapittel 7 er det imidlertid viktig å vise forsiktighet i tolkningen av resultatene basert på karakterer, da disse baserer seg på standpunktkarakterer. Dette er et grunnleggende problem for analyse av kvalitet i videregående skoler, elevene avlegger svært få eksamener. Videre er de som avlegges i stor grad i studieretningsfag med små og selekterte elevgrupper, og bare i svært beskjeden grad i de store fellesfagene, der elevtallet ellers tillater meningsfull statistisk analyse. Dermed er det vanskelig både å vite omfanget av og å ta hensyn til eventuelle forskjeller i karakterpraksis. Det er imidlertid verdt å merke seg at vi for videregående skole har tilgang til mer objektive – og kan hende mer relevante – mål, i form av gjennomstrømning/fullføring. Med objektive mener vi at de i mindre grad måler læreres/sensorers vurdering av en prestasjon, og i større grad det faktiske resultatet: Hvorvidt elevene fortsetter til neste trinn, og etter hvert fullfører og består. Disse målene er mindre utsatt for forskjeller i karakterpraksis, om enn ikke helt upåvirket, noe som er et argument for bruk av slik mål i vurdering av videregående skoler.

Forskjeller i karakterpraksis i grunnskolen kan også skape problemer for bruk av standpunktkarakterer til å ta hensyn til elevenes forutsetninger. Dette problemet er imidlertid mindre alvorlig enn evt. forskjeller i karakterpraksis på videregående. På elevnivå gir sannsynligvis standpunktkarakterer et bedre mål på ferdigheter og forutsetninger enn eksamenskarakterer, på grunn av den mer omfattende observasjonen som ligger bak standpunktkarakterene. Beregnede forskjeller i karakterpraksis er betydelige sammenlignet med beregnede forskjeller i skolekvalitet, men

små sammenlignet med forskjeller i elevers ferdigheter (standardavviket til de to første er omtrent 20-25 prosent av standardavviket resultater på elevnivå, jf. kapittel 5 og 7). Ved overgangen fra grunnskole til videregående fordeles elevene slik at videregående skoler sjelden er dominert av en grunnskole. Karakterpraksisen som ligger til grunn for de standpunktkarakterene elever ved en videregående skole har fra grunnskolen vil dermed variere, slik at utslagene på skolenivå blir mindre. Det kan imidlertid tenkes at noen videregående skoler har en stor andel elever fra skoler med avvikende karakterpraksis, slik at gjennomsnittlig grunnskolekarakter ikke er noe presist mål på forutsetninger. Dette er et argument for bruk av skriftlig eksamenskarakterer fra grunnskolen, som må veies mot den økte informasjonsverdien i å ha flere (og potensielt mer presise) kilder til informasjon.

# 10. Konklusjoner

I de fleste OECD-land har det i de senere år blitt lagt mer vekt på resultatkvalitet i skolen, og dokumentasjon av dette. Verdien av et slikt økt fokus avhenger imidlertid kritisk av at de måleinstrumentene man faktisk benytter kan si noe om skolens kvalitet. Gode indikatorer for variasjon i skolenes bidrag til elevenes resultater er fundamentalt for nytten av og tilliten til et kvalitetsvurderingssystem. Vurderinger og beslutninger med dette som en del av faktagrunnlaget er avhengig av skolebidragsindikatorer som faktisk inneholder den informasjon de er tiltenkt. Det er etter hvert allment anerkjent at ukorrigerte resultatgjennomsnitt på skolenivå kan være sterkt påvirket av faktorer som er utenfor skolens egen kontroll. Selv om slike resultatmål gir verdifull informasjon om elevenes kunnskapsnivå og prestasjoner, kan de gi et ufullstendig og misvisende bilde av skolekvalitet og hva som er skolens bidrag til resultatene. Mange studier har etter hvert vist at elevsammensetning og tilfeldig variasjon er viktige bidragsytere til resultatforskjeller mellom skoler. Resultatmål som ikke tar hensyn til disse faktorene, er med stor sikkerhet misvisende som mål på skolekvalitet. Spørsmålet er om man kan finne resultatmål som bedre reflekterer skolens bidrag til elevenes læring enn ukorrigerte skoleprestasjoner.

I denne rapporten har vi sett nærmere på value added-indikatorer for tre ulike deler av grunnopplæringen. En forutsetning for at vi har kunnet utføre disse beregningene, er at det har blitt tilgjengelig testresultater for de samme enkeltelevene på ulike trinn, satt sammen på en slik måte at resultatene kan ses i sammenheng. Konkret har vi sett på følgende value added-indikatorer:

1. Mellomtrinnet: Basert på nasjonale prøver for 8. og 5. trinn
2. Ungdomstrinnet: Basert på avgangskarakterer for 10. trinn og nasjonale prøver for 8. trinn
3. Videregående skole: Basert på karakterer/fullføring/frafall fra og avgangskarakterer for 10. trinn
4. I tillegg har vi beregnet skolebidragsindikatorer basert på tverrsnittsinformasjon for nasjonale prøver fra 5. trinn.

I norsk sammenheng er det tidligere blitt beregnet skolebidragsindikatorer for grunnskolen der resultater for den enkelte skole er korrigert for forskjeller i elevsammensetning målt ved elevenes sosioøkonomiske bakgrunn (tverrsnittsindikatorer).

Value added-indikatorer er resultatmål som bygger på samme tankegang, men som ved å ta hensyn til tidligere prøveresultater går lenger i korrigere for forskjeller i elevgrunnlag. Dermed oppnår man et mer nøyaktig resultatmål i forhold til å være indikatorer for skolens kvalitet eller bidrag til elevenes læring, i det de korrigerer for viktige forskjeller mellom skoler med hensyn til elevsammensetning som er utenfor deres kontroll, og som således ikke bør influere på indikatorer for skolens bidrag. Value added-indikatorer skiller seg fra andre indikatorer ved at de også benytter informasjon om elevenes resultater på et tidligere tidspunkt. Dermed får indikatoren en tydeligere tolkning som skolenes bidrag til endring i kunnskaper *i tidsrommet mellom de to målepunktene*. Gjennom valg av måletidspunkt kan man sikre seg at elevene (med unntak av ved flytting) er tilknyttet den samme skolen gjennom hele perioden.

Formålet med denne rapporten har vært å se nærmere på hvordan value added-indikatorer kan beregnes med de data som er tilgjengelig i Norge i dag, og drøfte hvordan de eventuelt kan implementeres innenfor Nasjonalt kvalitetsvurderingssystem (NKVS).

Value added- og skolebidragsindikatorer er et hjelpemiddel til å sammenligne resultatene til skoler med forskjellig elevsammensetning, og kan tolkes som det

resultatgjennomsnittet vi forventer at en skole ville hatt, om dens elevmasse var gjennomsnittlig i forhold til inkluderte elevkjennetegn. Således vil en indikator kunne bli påvirket av hvilken informasjon som er tilgjengelig og hvordan man teknisk gjennomfører korreksjonen. Indikatorene er et supplement til eksisterende informasjon om skoler og skolekvalitet, for eksempel på nettstedet Skoleporten. De kan ikke *erstatte* eksisterende informasjon, men kan bidra til å gi et mer utfyllende bilde av virksomheten som foregår på skolene.

Erkjennelsen av at skolekvalitet ikke kan oppsummeres i ett enkelt tall, gjør at value added-indikatorer bør presenteres sammen med annen relevant informasjon om skoler, slik at det er mulig å danne seg et mer helhetlig bilde av virksomheten ved den enkelte skole. Value added-indikatorer representerer en korrigering av ujusterte resultatgjennomsnitt, men det er ikke dermed sagt at slike ujusterte tall er uinteressante. Selv om ukorrigerte resultatmål som for eksempel skolens gjennomsnittsresultat ved skriftlig eksamen, eller andelen elever under et visst nivå på nasjonale prøver ikke nødvendigvis bare reflekterer skolens bidrag til elevenes læring, har slike mål selvsagt betydelig informasjonsverdi. Uavhengig av hvor godt skolens bidrag er, er det bekymringsfullt dersom mange elever ved en skole har resultater som vitner om et kunnskapsnivå som er for lavt i forhold til å skulle klare seg i videre utdanning og i arbeidslivet.

Value added-indikatorer representerer ikke en endelig ”sannhet” om enkeltskolers bidrag til elevenes læring. Man skal derfor være svært forsiktig med å knytte for eksempel insentiver eller sanksjoner overfor skoler direkte til indikatorene. Eksisterende resultatmål og tilhørende value added-indikatorer for enkeltskoler gir heller ingen oppskrift for hva som bør gjøres for å bedre læringsutnyttet. De må ses på verktøy for å *identifisere* god praksis i skolen, dvs. finne de skoler som bidrar mye eller lite til elevenes læring. For å *karakterisere* god praksis, dvs. finne hva som kjennetegner skoler med høyt bidrag eller enda mer ambisiøst *hvorfor* noen skoler bidrar mer enn andre, kreves andre data og andre analyseverktøy. Value added-indikatorene kan likevel være viktig input i mer overordnede analyser og som et første steg i arbeidet med å heve kvaliteten i norsk skole.

Selv om det er noen variasjoner mellom indikatorer for ulike trinn, er våre hovedfunn i kvalitativ forstand relativt robuste på tvers av trinnene. Våre hovedkonklusjoner kan oppsummeres slik:

1. Vi finner sterk samvariasjon mellom resultater på ulike trinn for enkeltelever. Det er naturlig nok slik at elever som lykkes godt på et trinn også skårer høyt tre år senere. For enkeltelever har de nasjonale prøvene på både 5. og 8. trinn sterk prediksjonskraft for henholdsvis nasjonale prøver på 8. trinn og avgangsresultater fra 10. trinn. Dette gir grunnlag for å bruke resultater fra nasjonale prøver i konstruksjonen av skolebidragsindikatorer for grunnskolen.
2. Resultatene viser gjennomgående at skoler som skårer høyt med hensyn til ujusterte resultater, også tenderer til å skåre høyt når vi ser på value added-indikatorer og skolebidragsindikatorer. Sammenhengene er imidlertid langt fra entydige, og det er mange skoler som får sine resultater betydelig justert. Resultatmål som tar hensyn til at skoler har ulik elevsammensetning gir betydelig tilleggsinformasjon sammenlignet med ujusterte resultater. Forskjellene mellom skoler er stort sett mindre når vi ser på indikatorer som tar hensyn til forskjeller i elevsammensetning (enten mht. tidligere resultater eller familiebakgrunn, eller begge deler), enn når vi ser på ujusterte resultater.
3. Plassering av enkeltskoler varierer og avhenger i noen grad av hvordan value added-indikatorene beregnes. Dessuten kan justeringen få svært store utslag for enkelte skoler. De mest ”ekstreme” skolene vil ofte avvike enda mer fra de andre når man benytter på value added-indikatorer. Dette kan reflektere flere forhold. For det første kan det skyldes at disse skolenes bidrag til elevenes læring er spesielt gode eller dårlige. Det kan imidlertid også skyldes ekstreme utslag pga tilfeldig variasjon (også for de tidligere resultatene) eller sterke

utslag i "halene" av fordelingen på grunn av den lineære spesifikasjonen av modellen. Slike og andre forhold illustrerer at value added-indikatorer ikke alene bør utgjøre datagrunnlaget for evaluering av skoler.

4. Både familiebakgrunnsvariable og tidligere elevprestasjoner forklarer en relativt stor del av den totale variasjonen i de resultatmålene vi ser på. Når vi allerede kontrollerer for tidligere resultater, har det imidlertid liten betydning (både for andelen forklart variasjon og for de estimerte skolebidrags-indikatorene) om vi kontrollerer for familiebakgrunn i tillegg.
5. Plasssering av enkeltskoler i fordelingen av indikatorer påvirkes i noen grad av beregningsmåte. Det er også forskjeller på tvers av fag. Selv om hoved-mønsteret er robust, understreker dette at justeringen gir oss en indikator, men ikke et entydig, presist mål på skolekvalitet.
6. Selv om skoletilhørighet forklarer en relativt liten andel av den totale variasjonen i elevers læringsresultater, viser de estimerte indikatorene betydelige forskjeller mellom skoler i deres bidrag til elevenes læring.

Få årganger med data gjør det vanskelig å avgjøre hvorvidt dette representerer tilfeldig variasjon (støy), transitoriske forskjeller mellom skoler eller mer persistente skoleforskjeller. Data for flere årskull med resultater for ulike trinn vil i prinsippet gjøre det mulig å skille ut skolebidrag som er stabile over tid.

***Konkret om implementering***

1. Indikatorer der elevutfall er justert for kjønn og tidligere resultater synes å være en robust beregningsmåte og har den åpenbare fordel at den kan implementeres av utdanningsmyndighetene, selv uten tilleggsinformasjon om eksempelvis familiekjennetegn fra eksterne kilder.
2. Usikkerhet bør rapporteres sammen med indikatorene som grunnlag for å vurdere om resultatforskjeller mellom skoler kan avvises som tilfeldige (statistisk signifikans).
3. Indikatorene bør beregnes med bakgrunn i data for flere årskull. Det vil redusere den statistiske usikkerheten knyttet til indikatorene gjennom at det blir flere observasjoner bak hver indikator, og ved å følge utviklingen i indikatorene over tid blir det mulig å se hvorvidt forskjellene mellom skoler representerer varige eller forbigående forhold.
4. Indikatorer basert på standpunktkarakterer frarådes. Innflytelsen fra variasjon i karakterpraksis på tvers av skoler kan være betydelig, og den er vanskelig å korrigere for.
5. For informasjonsverdien til value added-indikatorer, eller mer generelt, all sammenligning av elevers prestasjoner, er det åpenbart viktig at gjennomføring av nasjonale prøver foregår på mest mulig samme måte på forskjellige skoler

# Referanser

Galloway, T, L. J. Kirkebøen and M. Rønning (2011*). Karakterpraksis i grunnskoler. Sammenheng mellom standpunkt- og eksamenskaraktere,r* Rapport 2011/04. Statistisk sentralbyrå.

Hanushek, E.A. (2003), "The failure of input-based schooling policies," *Economic Journal*, 113 (February), F64–F98.

Hægeland, T og L. J. Kirkebøen (2008). School performance and value added indicators - what is the effect of controlling for socioeconomic background? Documents 2008/8. Statistisk sentralbyrå.

Hægeland, T., L. J. Kirkebøen og O. Raaum (2005). Skoleresultater 2004. En kartlegging av karakterer fra grunn- og videregående skoler i Norge, Notater 2005/31. Statistisk sentralbyrå.

Hægeland, T., L. J. Kirkebøen og O. Raaum (2006). *Resultatforskjeller mellom videregående skoler. En analyse basert på karakterdata fra skoleåret 2003-2004,* Rapport 2006/16. Statistisk sentralbyrå.

Hægeland, T., L. J. Kirkebøen, O. Raaum og K. G. Salvanes (2004). *Marks across lower secondary schools in Norway: What can be explained by the composition of pupils and school resources?* Rapport 2004/11. Statistisk sentralbyrå.

Hægeland, T., L. J. Kirkebøen, O. Raaum og K. G. Salvanes (2005a). *Skole-bidragsindikatorer. Beregnet for avgangskarakterer fra grunnskolen for skoleårene 2002-2003 og 2003-2004,* Rapport 2005/33. Statistisk sentralbyrå.

Hægeland, T., L. J. Kirkebøen, O. Raaum og K. G. Salvanes (2005b). *Skole-bidrags-indikatorer for Oslo-skoler. beregnet for avgangskarakterer fra grunnskolen for skoleårene 2002-2003 og 2003-2004,* Rapport 2005/36. Statistisk sentralbyrå.

Hægeland, T., L. J. Kirkebøen, O. Raaum and K. G. Salvanes (2005c). "Familiebakgrunn, skoleressurser og avgangskarakterer i norsk grunnskole." i Utdanning 2005 - ressurser, rekruttering og resultater, *Statistiske analyser* **74**. Statistisk sentralbyrå

Hægeland, T., L. J. Kirkebøen, O. Raaum og K. G. Salvanes (2007). *Skolebidrags-indikatorer for Oslo-skoler. Beregnet for avgangskarakterer fra grunnskolen for skoleårene 2004-2005 og 2005-2006,* Rapport 2007/28. Statistisk sentralbyrå.

Krueger, A.B. (2003), "Economic considerations and class size," *Economic Journal*, 113 (February), F34–F63.

OECD (2008) Measuring Improvements in Learning Outcomes: Best Practices to Assess the Value added of Schools. Paris: OECD.

Raudenbush, S.W. (2004). "What are value added models estimating and what does this imply for statistical practice?" *Journal of Educational and Behavioral Statistics*, 29(1), 121-129.

Rothstein, J. (2010): "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement", *The Quarterly Journal of Economics*, vol. 125(1), 175-214.

Todd, P.E. and K.I. Wolpin, (2003), "On the specification and estimation of the production function for cognitive achievement", *Economic Journal,* 113 (February), F3–F33.

# Vedlegg A: Mer om estimering av indikatorer og karakterpraksis

I dette vedlegget viser vi hvordan vi estimerer indikatorer for skolekvalitet, hvordan forskjeller i karakterpraksis fanges av de estimerte indikatorene, og under hvilke forutsetninger indikatorer basert på skriftlig eksamen vil være forventningsrette.

Indikatorene estimeres fra en ligning som dette:

$$E = \alpha + X\beta + SBI_j^E + u \tag{21}$$

Her antar vi at vi er i stand til å observere de kjennetengene som påvirker ferdigheter/prestasjoner, og som også potensielt varierer systematisk mellom skoler.[16] Skolebidrag basert på standpunktkarakterer estimeres fra en helt tilsvarende ligning. Med faste skoleeffekter for skolebidraget vil sammenhengen mellom kjennetegn og ferdigheter beregnes utfra (den empiriske) samvariasjonen mellom kjennetegn og gjennomsnittlige kjennetegn på skolen (M(*Y,Z*) angir den empiriske samvariasjonen mellom to variable *Y* og *Z*, M(*Y,Y*) er den empiriske variansen til *Y*, $\overline{X}_j$ angir gjennomsnittlig *X* på skole *j*, tilsvarende for *E*, og – i det følgende – andre variable) og tilsvarende forskjell i resultat:

$$\hat{\beta}^E = \frac{\mathrm{M}(X - \overline{X}_j, E - \overline{E}_j)}{\mathrm{M}(X - \overline{X}_j, X - \overline{X}_j)} \tag{22}$$

Når vi får svært mange observasjoner svarer disse empiriske samvariasjonene til teoretiske kovarianser:

$$p\lim \hat{\beta}^E = \frac{\mathrm{cov}(X - \overline{X}_j, E - \overline{E}_j)}{\mathrm{var}(X - \overline{X}_j)} \tag{23}$$

Skolebidragsindikatoren beregnes som forskjellen mellom det faktisk observerte snittresultatet på en skole, og det vi skulle vente basert på gjennomsnittlige kjennetegn til elevene ved skolen:

$$S\hat{B}I_j^E = \overline{E}_j - \overline{X}_j\hat{\beta}^E \tag{24}$$

Med ferdigheter og resultater bestemt fra (8)-(10) ser vi at SBI basert på henholdsvis eksamen og standpunkt blir:

$$\begin{aligned} S\hat{B}I_j^E &= \overline{E}_j - \overline{X}_j\hat{\beta}^E \\ &= (\overline{X}_j\gamma + \mu_j + \overline{\nu}_j + \overline{\varepsilon}_j) - \overline{X}_j\hat{\beta}^E \\ &= \overline{X}_j(\gamma - \hat{\beta}^E) + \mu_j + \overline{\nu}_j + \overline{\varepsilon}_j \end{aligned} \tag{25}$$

$$\begin{aligned} S\hat{B}I_j^S &= \overline{S}_j - \overline{X}_j\hat{\beta}^S \\ &= (\overline{X}_j\gamma + \mu_j + \theta_j + \overline{\omega}_j + \overline{\varepsilon}_j) - \overline{X}_j\hat{\beta}^S \\ &= \overline{X}_j(\gamma - \hat{\beta}^S) + \mu_j + \theta_j + \overline{\omega}_j + \overline{\varepsilon}_j \end{aligned} \tag{26}$$

[16] Dette inkluderer tidligere prestasjoner og elevbakgrunnsvariable. Kjennetegn som ikke varierer systematisk mellom skoler påvirker ikke beregnede indikatorer.

I det følgende antar vi at vi klarer å fange den korrekte sammenhengen mellom kjennetegn og ferdigheter ( $\hat{\beta}^E = \hat{\beta}^S = \gamma$ ). Videre er forventningsverdien til restleddene lik null: $\mathrm{E}(\overline{v}_j) = \mathrm{E}(\overline{\omega}_j) = \mathrm{E}(\overline{\varepsilon}_j) = 0$ Dette betyr at SBI basert på eksamen er *forventningsrett*, noe som innebærer at i gjennomsnitt og i den grad vi har mange elevobservasjoner ved hver enkelt skole gir disse et riktig bilde av skolekvaliteten, $\mu_j$: $\mathrm{E}(\hat{SBI}_j^E) = \mu_j$

For SBI basert på standpunktkarakterer er bildet mer komplisert, denne fanger i forventning *summen* av skolekvalitet og karakterpraksis: $\mathrm{E}(\hat{SBI}_j^S) = \mu_j + \theta_j$

# Figurregister

# Tabellregister