

*Taryn Ann Galloway, Lars J. Kirkebøen og
Marte Rønning*

Karakterpraksis i grunnskoler

Sammenheng mellom standpunkt- og
eksamenskarakterer

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

© Statistisk sentralbyrå, februar 2011 Ved bruk av materiale fra denne publikasjonen skal Statistisk sentralbyrå oppgis som kilde.	Standardtegn i tabeller	Symbol
ISBN 978-82-537-8037-5 Trykt versjon	Tall kan ikke forekomme	.
ISBN 978-82-537-8038-2 Elektronisk versjon	Oppgave mangler	..
ISSN 0806-2056	Oppgave mangler foreløpig	...
Emne: 04.02.20	Tall kan ikke offentliggjøres	:
Trykk: Statistisk sentralbyrå	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

Denne rapporten inngår i et prosjekt for analyse av læringsresultater, og er finansiert av Utdanningsdirektoratet. Rapporten bygger i noen grad på tidligere analyser av forskjeller mellom standpunkt- og eksamenskarakterer, presentert i Hægeland, Kirkebøen og Raaum (2005) og Gravaas m.fl. (2008). Takk til Torbjørn Hægeland, Oddbjørn Raaum og Bjarne Strøm for kommentarer til et tidligere utkast.

Sammendrag

Denne rapporten analyserer hvorvidt det er systematiske forskjeller mellom skoler i hvordan de setter standpunkt karakterer for sine elever. Måten dette gjøres på er å undersøke om det er systematiske avvik (på skolenivå) mellom standpunkt- og eksamens karakterer.

Standpunkt karakterene settes av læreren i det enkelte fag, og kan potensielt inneholde systematiske feil på skolenivå. Disse feilene kan enten være i form av at nivået er generelt for høyt eller lavt (dvs. skolen er for snill eller streng i sin karakterpraksis), eller i form av at det er lite samsvar mellom den underliggende ferdigheten og standpunkt karakteren (dvs. skolen klarer ikke å skille mellom gode og dårlige prestasjoner). Tidligere analyser gir indikasjoner på at det kan være innslag av relativ karaktersetning i standpunkt karakterer, det vil si at det er vanskeligere å oppnå en god karakter når man blir sammenlignet med flinke elever. Siden det er systematiske forskjeller mellom skoler med hensyn til elevsammensetning, vil dette kunne gi utslag på skolenivå, ved at noen skoler blir strengere enn andre.

For skriftlig eksamens karakter er det per forutsetning ingen systematiske avvik på skolenivå mellom ferdighet og karakter. Denne forutsetningen er rimelig fordi eksamens karakteren fastsettes av lærere utenfor skolen (sensorer). Disse sensorene kjenner ikke elevene, har ingen spesielle sympatier eller antipatier og er i mindre grad preget av medelevenes ferdighetsnivå. Gitt at eksamens karakteren ikke inneholder noen systematiske feil, vil systematiske avvik mellom ekstern og intern sensur utelukkende skyldes feil i standpunkt karakteren.

Resultatene tyder på at det er betydelig variasjon mellom grunnskoler i nivået på karaktersetningen. Noen skoler – særlig de små – ser ut til å overvurdere nivået til elevene på sin skole, mens andre – særlig store skoler – ser ut til å undervurdere. Vi finner også indikasjoner på at skoler med lave gjennomsnittlige eksamens karakterer har en tendens til å overvurdere elevenes nivå ved fastsetting av standpunkt karakterer, mens skoler med høye gjennomsnittlige eksamens karakterer heller tenderer mot å undervurdere prestasjonsnivået til elevene på sin skole. Det er også høy samvariasjon i karaktersetningen på tvers av fag, noe som tilsier at funnene drives av persistente skolekjennetegn, heller enn for eksempel enkeltlærere. Vi finner ingen tydelige forskjeller mellom skoler i samsvaret mellom den underliggende ferdigheten og standpunkt karakteren. Forskjellen mellom standpunkt- og eksamens karakterer er dermed tilsvarende for alle nivåer av eksamens karakter. Vi forsøker også å vurdere forskjellene mellom grunnskoler ved hjelp av elevenes standpunkt karakterer fra videregående, men resultatene av dette blir for upresise til at de kan tolkes.

Vi finner at det har en viss betydning å kontrollere for elevkjennetegn i analysene. Utover dette er resultatene lite følsomme for endringer i hvordan analysene gjøres. Resultatene må likevel tolkes med varsomhet. Grunnen til det er at vi ikke kan utelukke at relevante kontrollvariable, som bidrar til å forklare oppnådde resultater, er utelatt fra spesifikasjonene. Videre, standpunkt og eksamen kan i noen grad måle foreskjellige ferdigheter, og en elevs ferdighet komme til uttrykk i eksamens karakteren. Dette bidrar ytterligere til å komplisere sammenhengen mellom standpunkt- og eksamens karakterer.

Abstract

This report studies the extent to which systematic differences in classroom grading exist by analyzing systematic deviations (at the school level) between classroom grades and standardized final exam results.

Classroom grades are given by teachers in each subject and can conceivably be affected by systematic school-wide errors in grading practices. Such errors can result either because the school's grading practice is generally too lenient (or too strict) or because there is only a weak correspondence between the underlying skills and classroom grades, i.e. that the school fails to adequately distinguish between weak and strong students. Previous studies have suggested that classroom grading may be influenced by relative performance, i.e. that good grades are harder to obtain when a student is being compared to many other strong students. This can occur because the composition of the student body at individual schools can differ greatly and influence teachers' conceptions of the distribution of skill levels among students in the student population at large.

We assume, by definition, that there are no systematic deviations in standardized exam grades and skill levels. This assumption is reasonable because standardized exams are graded by randomly assigned external teachers (examiners). These external examiners do not know the students they are grading and are thereby not prejudiced by the relative skill level of the school's student body. Given that exam grades are not influenced by systematic grading errors, any systematic deviations between internal (classroom) grading and external (exam) grading is exclusively the result of systematic errors in classroom grading.

The results indicate substantial variation among middle schools in the levels for grading practices. Some schools – particularly small ones – tend to overestimate the level of their students whereas other schools – in particular larger ones – are more likely to underestimate their students. We also find evidence that schools with weaker average exam performances tend to overestimate their students while strong schools do the opposite. Grading practices between subjects within the same school are highly correlated, a fact which suggests that the results are largely driven by persistent school-level characteristics rather than individual teacher effects. We are unable to uncover clear differences among schools in the relative evaluation of weaker and stronger students. Thus, the differences between classroom and exam grades are largely differences in levels affecting grades at all levels of the grading scale. We also attempt to study differences in middle schools' grading practices by examining later grades in upper secondary school, but the results from that analysis are too imprecise to permit interpretation.

As expected, controlling for relevant characteristics of the individual students has some influence on the results, but the results are otherwise largely robust to changes in specification. Still, the results of this study should be interpreted with some caution, since we cannot entirely rule out the possibility that relevant control variables are excluded from the analysis. Furthermore, classroom grades and exam performance may measure slightly different skills, i.e. if a student has a bad day when taking his exam and thereby fails to show his true skill level. This complicates the relationship between classroom grading and exam results.

Innhold

Forord	3
Sammendrag	4
Abstract	5
1. Innledning	7
2. Analytisk rammeverk	9
2.1. Sammenligning mellom standpunkt- og eksamenskarakterer i grunnskolen.....	12
2.2. Sammenligning mellom standpunktkarakterer fra grunnskolen og standpunktkarakterer fra videregående skole	13
2.3. Mer om tolkning av resultatene	14
3. Data	16
3.1. Elevprestasjoner.....	16
3.2. Data for sosioøkonomisk bakgrunn	17
3.3. Tilrettelegging av datamaterialet	18
4. Resultater	20
4.1. Skriftlig eksamen som ekstern sensur.....	20
4.2. Standpunktkarakter fra videregående som ekstern sensur	31
5. Konklusjon	36
Referanser	37
Vedlegg A: En matematisk framstilling av sammenhengen mellom standpunkt og eksamenskarakter	38
Vedlegg B: Robusthet: Samvariasjon mellom forskjellige estimater	40
Figurregister	42
Tabellregister	43

1. Innledning

Lik vurdering av elevenes kunnskaper er viktig, av minst to årsaker. For det første har karakterene betydelige konsekvenser for elevene, gjennom å påvirke fremtidige utdanningsmuligheter og derigjennom karrieremuligheter i arbeidsmarkedet. For det andre blir karakterer og prøveresultater – og indikatorer avledet av disse - i økende grad brukt i vurderinger av skolen som virksomhet, både mer uformelt og gjennom formelle kvalitetsvurderingssystemer. For å sikre legitimiteten til karakterer som et sorteringskriterium på individnivå og som grunnlag for resultatvurdering av skoler, er det viktig at det ikke er *systematiske* forskjeller i sammenhengen mellom prestasjoner og karakterer på tvers av enheter. En elevs prestasjoner bør vurderes uavhengig av hva slags medelever han har eller hvilken skole han går på. Endelig er det viktig for følelsen av likebehandling og motivasjon hos den enkelte elev at det er godt samsvar mellom prestasjoner og karakterer. Denne rapporten analyserer hvorvidt det er *systematiske* forskjeller mellom skoler i hvordan de setter karakterer for sine elever.

En vesentlig del av de karakterene elevene får på sine vitnemål, er standpunkt-karakterer. For grunnskolen utgjør disse en svært stor del av beregningsgrunnlaget for grunnskolepoeng, som er opptaksgrunnlaget til videregående skoler. Standpunkt-karakterene settes av læreren i det enkelte fag. Det finnes retningslinjer for hva slags prestasjoner som skal kreves for å oppnå en gitt karakter, men disse er ikke veldig detaljerte. Det vil ofte ikke være mulig å gi så entydige og klare retningslinjer at de alene sikrer en enhetlig karaktersetting. I tillegg til retningslinjer og kompetanse hos den enkelte lærer, er det trolig at resultater fra for eksempel tidligere eksamener og kartleggingsprøver (for andre elever og årskull) kan gi en pekepinn på hvordan karakterene bør settes. For den enkelte lærer eller skole er det likevel betydelig grad av frihet når standpunkt-karakterer skal settes. Det er da en risiko for at det utvikler seg systematiske forskjeller mellom skoler mht. hva slags prestasjoner som kreves for å få en gitt karakter.

Tidligere analyser gir indikasjoner på at det kan være innslag av systematiske forskjeller mellom skoler i karaktersetting. Hægeland, Raaum og Salvanes (2004) finner at når man kontrollerer for elevens sosioøkonomiske bakgrunn, er sjansen for å få en høy standpunkt-karakter lavere jo mer fordelaktig den sosioøkonomiske sammensetningen ved skolen er (målt ved forventet karakternivå). Denne sammenhengen er ikke tilstede for eksamens-karakterer. Dette kan tyde på en form for relativ karaktersetting i standpunkt-karakterer, det vil si at det er vanskeligere å oppnå en god karakter når man blir sammenlignet med flinke elever. Siden det er systematiske forskjeller mellom skoler med hensyn til elevsammensetning, vil dette kunne gi utslag på skolenivå, ved at noen skoler blir strengere enn andre. Det er selvsagt også mulig at det vil være forskjeller i karakterpraksis selv i fravær av relativ karaktersetting, rett og slett fordi noen skoler legger til grunn en strengere vurdering enn andre. Gravaas, Hægeland, Kirkebøen og Steffensen (2008) finner i en foreløpig analyse tegn til at noen skoler peker seg ut som ”snille” eller ”strengt” gjennom å ha systematiske avvik mellom eksamens- og standpunkt-karakterer over tid.

Når man skal analysere hvorvidt karakterpraksis mht standpunkt-karakterer er systematisk forskjellig på tvers av skoler, er man avhengig av at det finnes en eller annen form for ”ekstern sensur”, for eksempel en eksamens-karakter, hvor prestasjonene til elever fra ulike skoler vurderes likt. Under en antakelse om at standpunkt-karakterene og karakterene ved ”ekstern sensur” i gjennomsnitt er satt etter samme skala på grunnlag av den samme prestasjon eller underliggende kunnskaps- og ferdighetsnivå, kan forskjeller mellom disse to målene, aggregert til skolenivå, benyttes som en sjekk på for grunnskolenes fastsetting av standpunkt-karakterer. Tabellen nedenfor viser et stilisert eksempel med en karakterskala med tre nivåer og en internt satt karakter og en ekstern karakter. For elever som har begge karakterer, vil det være ni mulige utfall, som vist i tabellen. Skoler som har stor

grad av samsvar mellom sine interne karakterer og de som følger av en ekstern sensur, vil ha en relativt stor del av sine elevobservasjoner langs diagonalen i tabellen nedenfor (i cellene merket ”Lik vurdering”), og ingen systematikk i om avvikene går i den ene eller andre retningen. Skoler som bedømmer sine elever for strengt, vil ha en relativt overhyppighet av elever i cellene merket ”Streng”, og tilsvarende for skoler som gir en for snill bedømming.

Ekstern sensur	Intern sensur		
	1	2	3
1	Lik vurdering (1,1)	Snill (1,2)	Snill (1,3)
2	Streng (2,1)	Lik vurdering (2,2)	Snill (2,3)
3	Streng (3,1)	Streng (3,2)	Lik vurdering (3,3)

I tillegg til avvik i form av streng eller snill praksis, er det også mulig at skoler er dårlige til å bedømme elevens prestasjoner, uten at det er noen spesielle tendenser til at de er spesielt snille eller strenge. Her vil elevobservasjonene være spredt mer jevnt ut over hele tabellen. Disse skolene vil ha relativt like gjennomsnittresultater på eksterne og interne karakterer, men samvariasjonen på elevnivå mellom ekstern og intern karakter vil være lav. Dette skaper i og for seg ikke problemer i et kvalitetsvurderingssystem for skoler som baserer seg på gjennomsnittresultater, men liten grad av samsvar mellom prestasjoner og karakterer på individnivå vil selvsagt være et stort problem for enkeltelever, og særlig de som blir bedømt strengt.

En utfordring ved en slik analyse er i hvilken grad ekstern og intern sensur faktisk måler samme underliggende ferdighet. Vi kan ikke nødvendigvis forvente at det skal være fullt samsvar mellom de to resultatene. Det trenger heller ikke være et likt samsvar i alle grupper av elever, ettersom forskjellige grupper av elever fordeler seg forskjellig på skoler er dette viktig å ta hensyn til i analysene.

I kapittel 2 presenterer vi et rammeverk for analyser av karakterpraksis i grunnskolen, og i kapittel 3 beskriver vi datamaterialet vi bruker. I kapittel 4 presenterer vi resultater fra analysen, og i kapittel 5 oppsummerer vi.

2. Analytisk rammeverk

For å analysere karakterpraksis trenger vi to resultater (intern og ekstern sensur) som i noen grad måler samme ferdighet. Som antydnet i navnet og innledningsvis benytter vi standpunkt karakteren fra grunnskolen som et mål på intern sensur, og i hovedsak eksamens karakteren fra grunnskolen som et mål på ekstern sensur. Standpunkt karakterene fra videregående kan også fungere som et tilleggsmål på ekstern sensur. Det analytiske rammeverket er i hovedsak felles, uavhengig av hva vi bruker som ekstern sensur. Vi bruker begrepene intern og ekstern sensur i den generelle omtalen.

Det ene målet, standpunkt karakter eller intern sensur, inneholder potensielt systematiske feil på skolenivå. Det er disse som er fokus i dette prosjektet. Disse feilene kan enten være i form av at nivået er generelt for høyt eller lavt (dvs. skolen er for snill eller streng i sin karakterpraksis), eller i form av at det er lite samsvar mellom den underliggende ferdigheten og standpunkt karakteren (dvs. skolen klarer ikke å skille mellom gode og dårlige prestasjoner).

For ekstern sensur (for eksempel eksamens karakter) er det per forutsetning ingen systematiske avvik på skolenivå mellom ferdighet og karakter. Denne forutsetningen blir rettfærdiggjort av at den eksterne sensuren settes utenfor skolen av tilfeldige andre lærere (sensorer). Disse sensorene kjenner ikke elevene, har ingen spesielle sympatier eller antipatier og er i mindre grad preget av medelevenes ferdighetsnivå (jf. diskusjonen om relativ karaktersetting). I prinsippet kan et hvilket som helst resultat som tilfredsstillende dette kravet (og som i noen grad måler samme ferdighet) brukes i analysene.

Denne rapporten har til hensikt å analysere hvorvidt det er systematiske forskjeller i karakterpraksis i grunnskolen. Måten dette gjøres på er å undersøke om det systematiske avvik (på skolenivå) mellom ekstern og intern sensur. Gitt at eksamens karakteren ikke inneholder noen systematiske feil, vil systematiske avvik mellom ekstern og intern sensur utelukkende skyldes feil i intern karaktersetting.

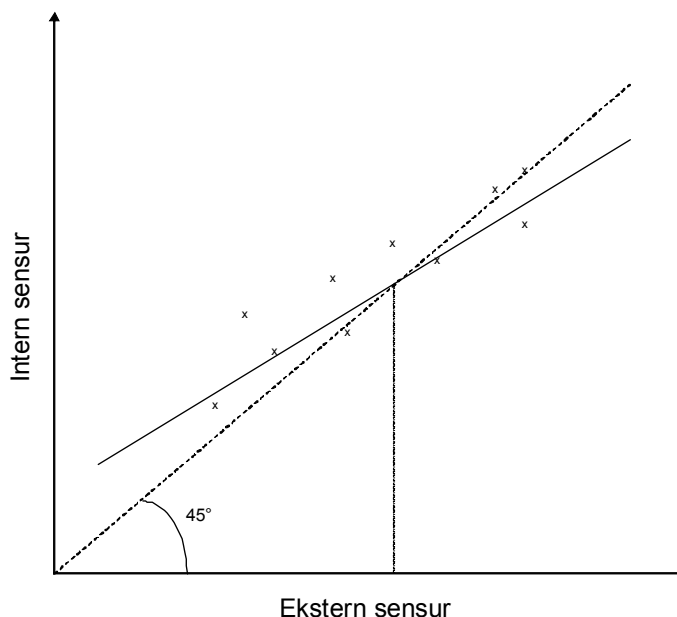
Kort oppsummert antar vi at intern sensur avhenger av ferdighet, som kommer til uttrykk gjennom ekstern sensur, andre karakteristika ved elevene og karakterpraksis. Fra hele datamaterialet, uten å ta hensyn til enkelt skoler, kan vi estimere en generell sammenheng mellom intern (I) og ekstern sensur (E), hensyn tatt til andre karakteristika ved elevene (F):¹

$$(1) \quad I = \alpha + \beta E + \gamma F + \varepsilon$$

I figur 2.1 viser vi denne sammenhengen grafisk. Kryssene representerer elever, hvor hver elev fremkommer som en kombinasjon av ekstern og intern sensur. Helningskoeffisienten, β , sier noe om hvor sterk samvariasjonen mellom intern og ekstern sensur er. Dette er representert ved den heltrukne linjen som er den faktiske sammenhengen mellom intern og ekstern sensur. Hvis det er perfekt samsvar mellom ekstern og intern sensur forventer vi at β er lik én. Dette er representert ved den stiplede 45-graderslinjen, der standpunkt er lik eksamen. Vi måler E og F på en slik måte at konstantleddet (α) uttrykker forventet intern sensur for en elev med gjennomsnittlig ekstern sensur og andre karakteristika. Dette er markert ved en stippet vertikal linje i figur 2.1. β og α vil fange opp en form for gjennomsnittlig karakterpraksis for alle skolene i landet.

¹ I vedlegg A gir vi en kortfattet beskrivelse av hvordan sammenhengen mellom intern og ekstern sensur kan modelleres matematisk for å komme fram til ligning (1), og kommenterer litt om hva som ligger i denne ligningen.

Figur 2.1. Sammenhengen mellom intern og ekstern sensur



Når vi er interessert i karakterpraksis på den enkelte skole, estimerer vi skole-spesifikke konstantledd (μ_j) og helningskoeffisienter (β_j), dvs. et konstantledd og en helningskoeffisient for hver skole, indekset med j .²

$$(2) \quad I = \alpha + \beta_j E + \gamma F + \mu_j + \varepsilon$$

Her vil det skolespesifikke konstantleddet μ_j fange opp alle uobserverbare elev og skoleforhold som varierer systematisk mellom skoler og som samtidig påvirker intern evaluering (I), med andre ord et betinget skolespesifikt karakternivå på skole j . Helningskoeffisienten, β_j kan tolkes som en betinget samvariasjon (korrelasjon). Dvs. samvariasjonen mellom intern og ekstern evaluering av elevene på den enkelte skolen, betinget på individkjenntegn samt eventuelle nivåforskjeller i standpunkt-karakterer mellom skole j og andre skoler angitt ved det skolespesifikke leddet.

Såfremt vi har et tilstrekkelig antall elever på hver skole er det enkelt å estimere en ligning som (2). Det er i imidlertid mer komplisert å tolke resultatene.³ En utfordring kan være at intern og ekstern sensur ikke fullt og helt måler samme underliggende ferdighet. For eksempel, standpunkt-karakterer baserer seg i noen grad på arbeid gjennom hele året og, sannsynligvis også for de skriftlige fagene, på deltagelse i timen. Der har elevene mulighet til å formidle sitt ferdighetsnivå muntlig, gjennom dialog med lærer og klasse, og med mulighet for oppfølgingsspørsmål, presiseringer med mer. Prøver er i hovedsak av kortere varighet, én til to timer. En skriftlig eksamen derimot er en veldig forskjellig situasjon, der eleven over en periode på flere timer besvarer en oppgave skriftlig, med begrenset mulighet til for eksempel å be om presiseringer. For en gitt ferdighet i faget, er det tydelig at disse situasjonene vil kunne belønne egenskaper som konsentrasjon og skriftlig og muntlig framstillingsevne forskjellig. Dette vil bidra til å svekke sammenhengen mellom standpunkt- og eksamens-karakter, og kan gi avvik i gjennomsnittlig karakter på ekstern og intern sensur. Disse forskjellene trenger ikke å være tilsvarende for alle (grupper av) elever. Ettersom forskjellige elevgrupper fordeler seg ulikt på skoler kan dette også bidra til systematiske forskjeller mellom intern og ekstern sensur på skolenivå.

² I praksis gjør vi dette ved å først estimere ligning (1), og predikere residualer, $\hat{\varepsilon}$ fra denne. Disse vil da være et mål på avviket til en elevs interne sensur fra vi skulle forvente, basert på den generelle sammenhengen mellom intern sensur, ekstern sensur og andre bakgrunns-karakteristika. Deretter estimerer vi skolespesifikke konstantledd og helningskoeffisienter fra disse residualene.

³ Dette er i noen grad presisert i vedlegg A.

Også avvik mellom ferdighet og karakter på elevnivå (dette gjelder både intern og ekstern sensur) kan bidra til systematisk avvik mellom intern og ekstern karaktersetting på skolenivå. Når det gjelder intern karakter kan vi tenke oss situasjoner hvor enkelte elever ikke får demonstrert sitt egentlige ferdighetsnivå gjennom de arbeidsformene som klassen/læreren har. Avvikene på elevnivå vil kunne være større for ekstern enn intern sensur. Ved skriftlig eksamen har sensor bare eksamensbesvarelsen å forholde seg til. Dette er et mye mindre informasjonsgrunnlag til å bedømme elevens ferdighetsnivå enn hva elevens faglærer har. Eksamensbesvarelsen trenger heller ikke å være et godt uttrykk for elevens underliggende ferdighetsnivå. En elev kan ha gode ferdigheter i et fag, men likevel prestere dårlig på eksamen for eksempel på grunn av dårlig dagsform, eller på grunn av et uheldig utvalg av oppgaver som tilfeldigvis omhandlet deler av faget som eleven er svakest på. I hvilken grad avvik på individnivå (både i intern og ekstern sensor) jevnes ut på skolenivå, avhenger blant annet av skolestørrelse, hvordan elevene er sortert på tvers av skoler, etc.

Videre vil målefeil i ekstern sensur bidra til å underestimere β_j . Så lenge denne målefeilen er likt fordelt på tvers av skoler er ikke dette problematisk. Vi kan forøvrig tenke oss situasjoner der noen skoler, særlig de med få elever, tilfeldigvis får en ekstern sensur som oppsummerer elevenes ferdigheter bedre (eller dårligere) enn hva som er vanlig. Dette vil bidra til et høyere (eller lavere) samsvar mellom ekstern og intern sensur på de aktuelle skolene.

For å kunne ta hensyn til alle de potensielle problemene nevnt ovenfor er det viktig å kontrollere for alle relevante elevkarakteristika. Ettersom forskjellige elevgrupper ikke er tilfeldig fordelt på skoler kan dette gi opphav til forskjeller mellom skoler. Vi forsøker å ta hensyn til dette gjennom å betinge på forskjellige elevkarakteristika i analysene. Disse observerbare kjennetegnene (for eksempel kjønn og foreldres utdanning) kan også være proxier for uobserverbare elevkjennetegn (for eksempel motivasjon og innsats), men vi kan ikke se bort fra at elevene også sorterer seg på tvers av skoler etter uobserverbare kjennetegn vi ikke klarer å ta hensyn til. Det er viktig å være klar over dette når man tolker resultatene.

For å oppsummere: Når vi estimerer ligning (2) vil β_j og μ_j fange opp forskjeller i karakterpraksis. Men vi er ikke i stand til å fullt ut isolere betydningen av karakterpraksis fra andre forhold som også vil fanges opp i disse parametrene. Dette innebærer at resultatene må tolkes med en viss varsomhet. Vi kan konkludere med hvorvidt det er forskjeller i forventet standpunkt-karakter, gitt eksamen-karakter, men strengt tatt ikke med hvorvidt dette skyldes en utrettferdig forskjell i karakterpraksis eller andre mer legitime forhold.

Den gjennomsnittlige helningskoeffisienten og konstantleddet fra (1), dvs. α og β , vil være sammenligningsgrunnlaget vårt når vi vurderer de skolespesifikke konstantleddene og helningskoeffisientene. Vi vil med andre ord uttrykke β_j og μ_j fra (2) som skole j 's avvik fra de gjennomsnittlige verdiene α og β fra (1). En skole med et positivt (negativt) avvik i β_j vil da kunne antas å overdrive (undervurdere) den relative forskjellen mellom sterke og svake elever sammenliknet med gjennomsnittet. Med andre ord, en skole med lav samvariasjon mellom intern og ekstern sensur treffer dårlig i dens relative vurdering av svake og sterke elever. En skole med et positivt (negativt) avvik i μ_j vil kunne sies å overvurdere (undervurdere) nivået til elevene på den skolen, dvs, ha en snillere (strengere) karakterpraksis enn gjennomsnittet. Et sentralt spørsmål i denne sammenhengen vil være i hvilken grad disse også/faktisk avviker, i statistisk forstand.

I resten av artikkelen kommer vi til å benytte begrepene ”skolespesifikt konstantledd og karakternivå” og ”skolespesifikk helningskoeffisient og samvariasjon” om hverandre.

Nivået for analysen er karakterpraksis på den enkelte skole. Årlige svingninger eller praksis hos enkeltlærere er det ikke mulig å kartlegge betydningen av. Likeledes vil det være vanskelig å skille mellom to (hypotetiske) skoler der den ene har en svært snill lærer som underviser i basisfagene og en skole der den gjennomgående praksis blant alle lærerne er litt for snill. Uten av det framgår eksplisitt av modellen over, benytter vi observasjoner av flere kull og teste hvorvidt karakterpraksis varierer mellom fag for den enkelte skole. Det er hensiktsmessig å bruke data for en så lang tidsperiode som mulig for bedre å kunne skille persistente forskjeller mellom skoler fra tilfeldige svingninger. Å se på gjennomsnitt over flere år samt samvariasjon i karakterpraksis på tvers av fag gjør oss i noen grad i stand til å redusere enkeltlæreres innflytelse på estimeringsresultatene.

For å ta hensyn til variasjoner over tid i intern karakterpraksis i skolene, estimeres ligning (2) med årsummier. Slik variasjon over tid i intern karakterpraksis i skolene kan for eksempel skyldes lærerplanendringer og nasjonale reformer.

Gitt de data som er tilgjengelige pr. i dag, er det i hovedsak to muligheter som peker seg ut til å analysere karakterpraksis i grunnskolen ved hjelp av noe som kan tolkes som ekstern sensur, skriftlig eksamen i grunnskolen og resultater fra videregående skole.

2.1. Sammenligning mellom standpunkt- og eksamenskarakterer i grunnskolen

Skriftlig eksamenskarakter er en åpenbar kandidat som ekstern sensur. Eksamen og standpunkt måler sannsynligvis langt på vei samme ferdighet, selv om dette neppe gjelder fullt ut. Videre er det grunn til å tro at eksamenskarakteren i stor grad er uavhengig av karakterpraksis ved elevens skole. Dette er fordi eksamen rettes av tilfeldige andre lærere, som ikke kjenner elevene de retter, og til vanlig underviser andre elever, potensielt på ett annet ferdighetsnivå. Ved muntlig eksamen er det ingen tilsvarende uavhengighet. Elevene er ikke lenger anonyme for sensor, men det viktigste argumentet mot uavhengighet av skolens karakterpraksis er at faglærer er tilstede som eksaminator, og vil kunne påvirke vurderingen. Muntlig eksamenskarakter har dermed lite potensial som ekstern sensur, selv om den sannsynligvis måler ferdigheter som inngår i grunnlaget for standpunktkarakter, og som ikke nødvendigvis måles ved skriftlig eksamen (muntlig framstilling og mer dialogpregede besvarelser innen faget).

Det er likevel noen forhold som kan bidra til systematiske forskjeller også i skriftlig eksamenskarakter. Dersom en sensor får tildelt en bunke besvarelser fra samme skole kan det forekomme noen grad av relativ karaktersetting innad i denne. Hvis en sensor får tildelt en bunke som inneholder påfallende mange gode besvarelser kan denne tenke at dette i hvert fall dels er et uttrykk for at eksamen var for enkel, og at karakternivået bør justeres deretter. Dette vil i så fall gi systematisk dårligere eksamenskarakter til elever på skoler med høyt nivå, tilsvarende ved relativ karaktersetting i standpunktkarakter. Det er også mulig at en sensor vil praktisere noen grad av relativ karaktersetting sammenlignet med det nivået denne er vant til. I så fall vil sensorer som til vanlig har sterke elever sette en lavere karakter for en gitt ferdighet. Så lenge alle skoler får rettet besvarelser av et stort antall tilfeldige sensorer vil dette jevne seg ut. Men dersom det er systematikk i hvilke sensorer som retter for hvilke skoler kan det bli systematiske skjevheter. For små skoler, i denne sammenhengen betyr det skoler som i sum har et begrenset antall eksamensbesvarelser, kan dette være et problem også med tilfeldig tilordning av sensorer. For alle små skoler vil dette igjen jevne seg ut, men så lenge vi ser på enkeltskoler med få eksamensbesvarelser vil en sensor som setter spesielt høye eller lave karakter kunne ha betydning for estimert karakterpraksis.

Når vi bruker eksamenskarakter som ekstern sensur vil vi benytte en statistisk modell som svarer helt til (2):

$$(3) \quad I_{ijf} = \alpha + \beta_{jf}E_{ijf} + \gamma F_{ijf} + \mu_{jf} + \varepsilon_{ijf}$$

Her betegner I den interne fastsatt karakter, standpunkt karakteren i 10. klasse, og E den eksterne fastsatte eksamenskarakteren, mens F er et sett (en vektor) med variable som beskriver elevens sosioøkonomiske bakgrunn. Fotskrift i betegner individ, fotskrift j betegner den grunnskolen hvor den interne karakteren er satt og fotskrift f betegner fag. μ_{jf} er et skolespesifikt konstantledd (skole j 's betingende karakternivå i fag f) som er felles for alle som har fått sin interne karakter i fag f fastsatt ved grunnskole j . β_{jf} er skole j 's helningskoeffisient i fag f , eller betinget samvariasjon. ε_{ijf} er et restledd som inneholder alle uobserverbare forhold som kan tenkes å påvirke fastsettelsen av standpunktkarakteren. Data for kombinasjoner av standpunkt og eksamen i 10. klasse gjør det mulig å estimere modellen på elevnivå.

Karakterpraksis i fag f ved grunnskolen fanges opp av både parametrene β_{jf} og μ_{jf} som begge kan estimeres ved såkalt standard regresjonsanalyse. Vi kan da teste hvorvidt:

1. β_{jf} er forskjellig mellom skoler. Vi estimerer disse parametrene og sjekker hvilke β_{jf} som er forskjellige fra gjennomsnittet
2. μ_{jf} er forskjellig fra null (positiv betyr "snill" mens negativ innbærer "streng" intern sensur)

2.2. Sammenlikning mellom standpunktkarakterer fra grunnskolen og standpunktkarakterer fra videregående skole

I tillegg til sammenlikning av standpunkt og eksamen i 10 klasse gjør skolestrukturen og opptakskriteriene på videregående skoler i Oslo og andre større byer i Norge det mulig å benytte en annen metode for vurdering av karakterpraksis. Grunnlaget er at elever fra hver enkelt ungdomsskole sprer seg på flere videregående skoler, hvor de møter elever fra andre skoler. Anta for enkelthets skyld at vi har tre grunnskoler (A, B, C) og tre videregående skoler (1, 2, 3). På alle de tre videregående skolene finnes det elever fra alle de tre grunnskolene, og disse elevene får karakterer i samme fag. De videregående skolene har sin karakterpraksis som ikke nødvendigvis trenger å være "riktig", men som vi antar ikke systematisk favoriserer elever fra bestemte grunnskoler. Hvis karakterpraksisen ved grunnskolene er lik, vil det for elever med en gitt karakter fra grunnskolen være slik at forventet karakter på videregående ikke avhenger systematisk av hvilken grunnskole de kom fra. Vi ønsker altså å teste hvorvidt det er slik at for eksempel en firer fra grunnskole A er mer verdt enn en firer fra grunnskole B eller C. Dette kan i prinsippet testes på alle de videregående skolene 1, 2 og 3.

For analysen kan man tenke seg en modell hvor den eksterne karakteren er elevens karakter på videregående. Intern sensur er fremdeles standpunktkarakter på grunnskolen. Intern sensur avhenger av ferdighet og karakterpraksis, samt evt. andre elevkarakteristika og tilfældigheter. Ettersom standpunktkarakter på videregående avhenger ikke bare av ferdighet fra grunnskolen, men også av endring av ferdighet gjennom videregående og karakterpraksis på de videregående kontrollerer vi for den videregående skolen:

$$(4) \quad I_{ijkf} = \alpha + \beta_{jf}E_{ijkf} + \gamma F_{ijkf} + \mu_{jf} + \eta_k + \varepsilon_{ijkf}$$

Dette er en enkel utvidelse av modell (2) og (3), hvor den eneste forskjellen er at vi nå har med et nytt element (η_k) som avspeiler eventuell tilleggseffekt for elever

som har fått sin eksterne karakter fastsatt ved videregående skole k . Forskjeller i karakterpraksis og/eller skolens bidrag til elevenes læring på tvers av videregående skoler, som også påvirker elevenes karakternivå, vil fanges opp av parameterne η_k .

Igjen vil eventuelle forskjeller i karakterpraksis på tvers av grunnskoler slå ut i parameterne μ_{ij} og β_{ij} . Data for kombinasjoner av standpunkt i 10. kl og karakterer fra videregående gjør det mulig å estimere modellen på elevnivå.

Denne delen av analysen vil konsentrere seg om skoler i de fire største byene. Dette skyldes at det her er en mer veldefinert og oversiktlig overgang mellom grunnskoler og videregående, hvor elever fra samme grunnskole går til ulike videregående skoler (med samme studieretning) i større grad enn for landet som helhet, noe som er en forutsetning for å kunne identifisere parameterne μ_j og β_j i modellen over. Vi vil i stor grad fokusere på basisfagene i GK/Vg1, både på studieforberedende utdanningsprogrammer og fellesfagene på yrkesfag, fordi det er her man har det største antallet av felles fag for store grupper av elever. For at standpunkt-karakterer på videregående skal være uavhengige av karakterpraksis på grunnskolen vil vi ha en viss omfordeling av elever. Hvis for eksempel relativ karaktersetning er viktig ønsker vi ikke at elever fra én ungdomsskole skal dominere en videregående skole. Vi er også avhengige av en viss variasjon i ekstern sensur, samt at vi ønsker at de som setter karakterene på videregående skal ha et sammenligningsgrunnlag av en viss størrelse, og vil derfor avgrense datamaterialet på grunnlag av størrelse på de videregående skolene, mer om dette i avsnitt 3.3.

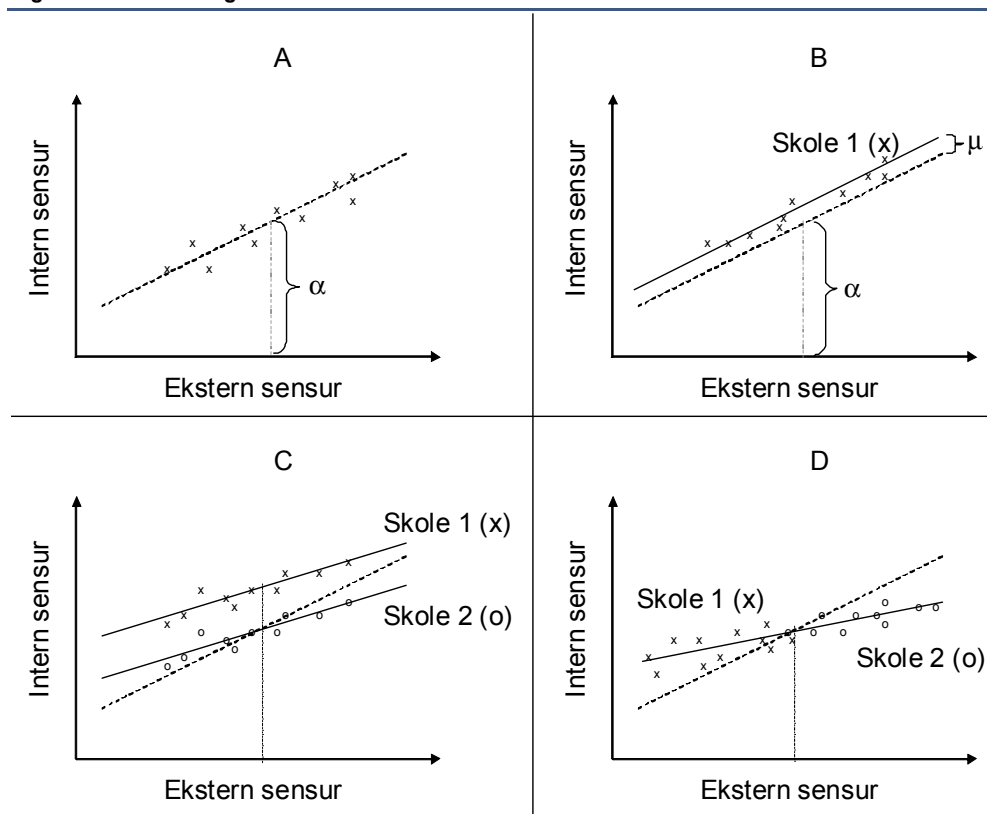
En utfordring med denne tilnærmingen er at vi som beskrevet bare kan bruke en del av datasettet. En ytterligere utfordring er at elevene fordeler seg på studieretninger og fag i videregående. I grunnskolen har alle elever samme fag, og faget for skriftlig eksamen er tilfeldig valgt. I videregående er det elevene selv som velger, slik at elevsammensetningen på forskjellige studieretninger og på forskjellige fag innen hver studieretning blir forskjellig. Vi vil derfor estimere forskjellige sammenhenger mellom standpunkt-karakter fra grunnskole og videregående skole, der forskjellene både kommer fra at vi sammenligner med forskjellige fag fra videregående, og at det er forskjellige (grupper av) elever som tar fagene.

2.3. Mer om tolkning av resultatene

Figur 2.2 skisserer noen forskjellige eksempler som er nyttige for å illustrere tolkningen av koeffisientene som estimeres i denne analysen. Del A av Figur 2.2 viser et eksempel på hvordan forholdet mellom den interne og eksterne sensuren kan se ut for en enkeltskole, eller for hele datamaterialet, uten å ta hensyn til enkeltskoler. Figuren viser enkeltelever som kryss, og en heltrukken regresjonslinje som oppsummerer sammenhengen mellom intern og ekstern sensur. Hvis vi tolker figuren som sammenhengen for hele datamaterialet, uten hensyn til enkeltskoler, angir i vår spesifikkasjon koeffisienten $\hat{\alpha}$ den forventete interne karakteren til elever som har en gjennomsnittlig prestasjon mhp den eksterne sensuren. I figuren er dette avstanden opp til regresjonslinjen, målt langs den vertikale stiplede linjen. Denne avstanden er dermed et mål på hvordan den interne elevvurderingen forholder seg til den eksterne elevvurdering for en gjennomsnittselev i Norge. Helning av linjen, β i vår notasjon, forteller oss noe om sammenhengen mellom intern og ekstern sensur når vi avviker fra gjennomsnittet.

Del B av Figur 2.2 gir et eksempel på hvordan en skole kan avvike i nivå fra den gjennomsnittlige sammenhengen mellom ekstern og intern sensur som gjelder for hele landet. Den gjennomsnittlige sammenhengen for hele landet er tegnet inn som en stiplet linje. For gitt ekstern sensur har de enkelte observasjonene fra Skole 1 gjennomgående høyere intern sensur enn den stiplede linjen skulle tilsi. Derfor ville det skole-spesifikke konstantleddet μ være positiv for Skole 1, og størrelsen av μ vil fortelle noe om hvor mye den skolen avviker i vurdering av elevenes nivå. I del B av figuren ”bommer” skolen like mye (i størrelse μ) uansett om en flink eller en dårlig elev vurderes.

Figur 2.2. Tolkning av resultatene



Del C av Figur 2.2 gir to eksempler på hvordan en skole kan avvike fra den gjennomsnittlige sammenhengen mellom ekstern og intern sensur, som igjen er tegnet som en stiplet linje. Skole 2, hvis elever angis med åpne sirkler, ligger på et generelt nivå som ligner en del på det som gjelder i gjennomsnitt for landet, men det ser ut som at den relative vurderingen av svake vs. flinke elever er noe annerledes på den skolen enn for landet i gjennomsnitt. Den forskjellen ville gjenspeiles ved at den skolespesifikke helningskoeffisienten til Skole 2 avviker fra den generelle (gjennomsnittlige) helningskoeffisienten for landet i sin helhet. I eksemplet gitt ved Skole 2 i del C av Figur 2.2 ser det ut til at skolen er dårlig til å skille mellom flinke og svake elever, slik at flinke elever undervurderes med den interne sensuren mens svake elever overvurderes. Derfor kan vi også si at Skole 2 treffer dårlig i dens relative vurdering av flinke og svake elever, hvor ”dårlig” her betyr at den relative vurderingen er annerledes enn for landet i gjennomsnitt. Skole 1 i del C av Figur 2.2 treffer også dårlig når det gjelder den relative vurderingen av flinke og svake elever, men i det eksemplet blir alle elevene (gitt ved x’ene i figurene) overvurdert i forhold til den gjennomsnittlige sammenhengen mellom intern og ekstern sensur for hele landet. Grunnen for det er at det skolespesifikke konstantleddet til Skole 1 er både positivt og stort. Forekomsten av under- og overvurdering av elever ved en spesifikk skole derfor avhenger av både nivåforskjeller (gitt ved μ), forskjeller i den relative vurderingen av svake og sterke elever (gitt ved β), og forholdet mellom disse.

Del D av Figur 2.2 viser et siste spesielt eksempel av hvordan vi faktisk kan finne at to skoler ha de samme skolespesifikke konstantleddene og helningsparametrene, men hvor den ene skolen undervurderer sine elever og den andre overvurderer sine elever. Denne situasjonen oppstår fordi elevene ved Skole 1 (angitt med x’ene) er alle ganske svake, men alle elevene ved Skole 2 er ganske flinke.

Til slutt er det verdt å merke seg at det først og fremst er i konstantleddene vi noenlunde presist estimerer variasjon av betydning. Det er dermed del B av figur 2.2 som er mest relevant for tolkingen av resultatene vi presenterer i neste kapittel, mens kompliserende forholdene i del C og D viser seg å mindre betydning.

3. Data

I dette kapitlet gir vi en nærmere beskrivelse av de ulike datakildene vi benytter i analysen.

3.1. Elevprestasjoner

Elevenes læringsutbytte måles ved avgangskarakterer fra ungdomsskolen samt karakterer fra videregående. Vi har eksamens- og standpunkt karakterer fra grunnskolen for skoleårene 2001/02 til 2007/08 (7 skoleår). For videregående har vi karakterdata for skoleårene 2003/04 til 2007/08. Disse dataene er samlet inn av Utdanningsdirektoratet. Kunnskapsløftet (KL) ble innført i grunnskolen fra skoleåret 2007/08 og i videregående fra skoleåret 2006/07 (på vg1). Dette har medført noen endringer i fag og organisering av eksamen, som vi omtaler under.

Karakterer fra grunnskolen

De fleste elevene vil ved avgang fra grunnskolen få standpunktkarakter i alle fag, og i tillegg bli trukket ut til en skriftlig eksamen og en muntlig eksamen i tilfeldige fag. Fagene der det avholdes skriftlig eksamen er engelsk skriftlig, matematikk og norsk. Før innføringen av KL, under Læreplan 97 (L97), skulle elever som trekkes ut til skriftlig eksamen i norsk normalt ha eksamen i både hovedmål og sidemål. Omtrent halvparten så mange elever ble trukket ut i norsk som i engelsk og matematikk. For avgangskullet 2008, etter innføring av KL, var det felles eksamen i norsk hovedmål og sidemål, og omtrent like mange elever ble trukket ut i norsk, engelsk og matematikk. Innholdet i fagene er lite endret ved innføringen av KL. Steffensen og Ziade (2009) finner at gjennomsnittlige standpunktkarakterer i tilsvarende fag under KL og L97 er stabile. Bildet er mer sammensatt for skriftlig eksamen, her er det et forholdsvis markert fall i norsk fra 2007 til 2008, mens det er en mer beskjeden økning i både engelsk og matematikk. I prinsippet kan det være en utfordring at vi bruker data både før og etter innføringen av KL. Alle elever vil imidlertid ha standpunkt og skriftlig eksamenskarakter fra enten KL eller L97, vi er videre interessert i systematiske forskjeller på skolenivå i avvikene mellom disse karakterene. Disse kan tenkes å være påvirket av innføringen av KL, men vi prøver å finne persistente, "dypere" egenskaper ved skolene. I den grad disse egenskapene svekkes eller forsvinner ved innføringen av KL blir de mindre tydelige estimert, men også mindre relevante. Vi tar også hensyn til nivåforskjeller mellom år gjennom spesifikasjonen vi bruker.

Karakterer fra videregående skole

Karakterdata i videregående registreres i alle fag, på alle trinn. Kunnskapsløftet ble innført på vg1 (første trinn) fra skoleåret 2006/07. Utfordringen med innføringen av KL, som erstattet læreplanen fra Reform 94 (R94), er større når vi bruker data fra både videregående og grunnskolen. Dette er fordi det ikke lenger nødvendigvis er slik at en elev har karakterer fra enten KL eller L97/R94: Elever som gikk ut av grunnskolen i 2005 og 2006 vil begge ha standpunktkarakterer fra L97, men, om vi antar at begge går rett over i videregående utdanning, resultater fra videregående under hhv R94 og KL. Det er ikke klart at de samme sammenhengene gjelder for disse. Vi vil fokusere på fag fra R94, ettersom vi har flere observasjoner av elever med disse fagene.

Videregående skole er, i motsetning til grunnskolen, preget av et stort mangfold av fag. Elevene fordeler seg først på studieretninger (-programmer) og har innen hver studieretning også en betydelig frihet til å velge fag. Som beskrevet i avsnitt 2.2 er vi avhengige av et visst antall elever, fordelt på et visst antall videregående skoler. Videre, innen hver videregående skole må vi ha elever fra et visst antall grunnskoler for å kunne bruke videregående standpunktkarakterer som ekstern sensur i vårt estimeringsopplegg. Dette utelukker alle fag utenom de store fellesfagene, stort sett fra GK. Dette betyr at vi vil fokusere på basisfagene, dvs. norsk, engelsk og matematikk, som er de samme fagene vi studerer vha. skriftlig eksamenskarakter fra grunnskolen. I prinsippet kunne et hvilket som helst fag fra videregående bli brukt som ekstern sensur, såfremt det fanger deler av elevens ferdighet som den interne sensuren ønsker å måle, og såfremt det ikke er påvirket av

karakterpraksisen på grunnskolen. For at fagene i størst mulig grad skal avhenge av samme ferdigheter vil vi bruke korresponderende fag for videregående og grunnskole, dvs., vi vil analysere karakterpraksisen i matematikk i grunnskolen ved å bruke de forskjellige fellesfagene i matematikk fra videregående som ekstern sensur. Tabell 3.1 viser fagene fra videregående vi vil benytte.

Tabell 3.1. Fag i videregående, men korresponderende fag fra grunnskolen

Fag	Studieretning	Grunnskolefag
VF1000	Allmennfag	Norsk
VF1010	Yrkesfag	Norsk
VG4000	Allmennfag	Norsk
VG1200	Allmennfag	Engelsk
VF1210	Yrkesfag	Engelsk
VG1330	Allmennfag	Matematikk
VG1331	Allmennfag	Matematikk
VF1320	Yrkesfag	Matematikk

Alle fagene er GK-fag, med unntak av VG4000, som er avsluttende norsk hovedmål skriftlig på VK2. Ettersom det første avgangskullet med grunnskolekarakter er våren 2002 og data for videregående finnes fra skoleåret 2003/04, vil vi for fagene på GK bruke resultater fra 2004 til 2006 (før innføringen av KL), mens vi for VG4000 bruker karakterer fra 2004 til 2008. Generelt sett kan vi ikke anta at sammenhengene mellom karakter i norsk fra grunnskolen og for eksempel VF1000 og VF1010 er de samme. Derfor gjør vi analysene separat for hvert fag fra videregående.

3.2. Data for sosioøkonomisk bakgrunn

Skoler har til dels svært forskjellig elevsammensetning, målt langs en rekke ulike dimensjoner. Som vist blant annet i Hægeland mfl. (2005a) er det til dels sterk samvariasjon mellom mange familiebakgrunnsvariable og elevers skoleprestasjoner. Som diskutert tidligere kan det være forskjeller i hvordan forskjellige elevgrupper presterer på ekstern og intern sensur, gitt ferdighetsnivå. Når elevgruppene fordeler seg forskjellig på skolene kan dette føre til systematiske forskjeller på skolenivå i sammenhengen mellom intern og ekstern sensur. Dette er en viktig motivasjon for å kontrollere for sosioøkonomisk bakgrunn.

Når man skal beskrive skolene innenfor og utenfor programmet med hensyn til elevsammensetning, er det derfor mange variable man kan bruke. Ettersom vi i analysene av intern sensur kontrollerer for tilsvarende ferdighet målt ved ekstern sensur, reduseres imidlertid betydningen av detaljert kontroll for elevbakgrunn, se Hægeland, Kirkebøen og Raaum (2010) for et tilsvarende eksempel. Vi bruker derfor en enklere spesifikasjon av elevbakgrunn enn det som for eksempel er brukt til beregning av skolebidragsindikatorer (Hægeland mfl., 2005b). Nedenfor gir vi en kort oversikt over hva slags kjennetegn vi inkluderer i vår analyse:

- Elevens kjønn
- Hvorvidt foreldre er gift/samboende
- Mors og fars utdanningsnivå.
- Innvandret og norskfødte med innvandrerforeldre
- Økonomiske ressurser: Vi lager et mål på familieinntekt basert på summen av mors og fars pensjonsgivende inntekt de siste ti årene, som reflekterer perioden eleven har gått på skolen. I analysene bruker vi logaritmen av denne inntekten.

Alle variable er beregnet for tidspunktet eleven fullførte grunnskolen, også der vi bruker data fra videregående skole. Samlet sett gir disse variablene en karakterisering av elevenes familiebakgrunn, slik den "fremkommer" i administrative datakilder. I en del andre undersøkelser av betydningen av hjemmebakgrunn for skoleprestasjoner o.l., har man ofte med variable som er ment mer direkte å reflektere hjemmemiljøet eller familiens "kulturelle kapital." Slike variable kan gi nyttig tilleggsinformasjon, men de finnes ikke i administrative registre, og lar seg vanskelig samle inn for hele populasjonen. I tillegg vil slike variable samvariere sterkt med dem vi kan hente fra registerinformasjonen og vil således bidra lite til å forklare ytterligere variasjon i karakterer.

3.3. Tilrettelegging av datamaterialet

Alle data er hentet fra administrative registre, der data er lagret med identifika-sjonsnumre som gjør det mulig å koble forskjellige datakilder. Alle datakildene er i prinsippet fullregistreringer, i praksis er det et beskjedent mindretall vi ikke har alle variable for. For analyse med eksamenskarakterer fra grunnskolen er det drøye 5 prosent av elevene som ikke har noen skriftlig eksamen, og ytterligere et fåtall som bare er registrert med skriftlig eksamen i et fag utenom norsk (hovedmål), engelsk eller matematikk. Disse utgår fra datamaterialet. Det gjør også omtrent 10 prosent av elevene, der vi mangler en eller flere kontrollvariable, i all hovedsak fordi vi ikke er i stand til å koble med data for en eller begge foreldre.

I videregående bruker vi også bare elever som har registrert standpunktkarakter fra grunnskole og videregående, samt alle kontrollvariable. De aller fleste elevene som finnes i grunnskoledataene gjenfinnes på videregående, enten med allmennfaglige eller yrkesfaglige fag.⁴ Mens det i grunnskolen er et svært stort flertall av elevene som fullfører ved normert alder (et visst mindretall ett år for tidlig eller sent), er det stor aldersspredning i videregående. I analysene bruker vi bare elever som er 20 år eller yngre.⁵ Som beskrevet i avsnitt 2.2 kreves det en viss blanding av elever fra forskjellige grunnskoler til videregående skoler for at vi skal kunne bruke standpunkteksamen i videregående som ekstern sensur. Vi gjør derfor en del avgrensinger i datamaterialet. For det første bruker vi bare skoler i de fire største byene (Oslo, Bergen, Trondheim og Stavanger/Sandnes).⁶ Elever i alle disse kommunene fordeler seg i all hovedsak med minst ti elever på minst ti videregående skoler både med allmennfag og yrkesfag.⁷ Elever i alle andre kommuner fordeler seg på færre skoler. Dermed bruker vi bare observasjoner fra videregående skoler med minst 30 elever, og der ingen grunnskole bidrar med mer enn 25 prosent av elevene. Vi gjør ingen avgrensning etter grunnskolestørrelse eller beliggenhet. Dette betyr at vi for eksempel får med elever som flytter fra en annen kommune mellom grunnskole og videregående, og vi får dermed med et betydelig antall grunnskoler som er representert med bare noen få elever. Det er naturlig å ha med disse, da de utgjør en del av elevgrunlaget på de videregående skolene vi studerer. Vi presenterer imidlertid bare resultater for grunnskoler med minst 20 elever, noe som utelukker disse skolene. Ingen grunnskoler som bidrar med færre enn tre elever vil heller ha noen innflytelse på de presenterte resultatene, ettersom vi estimerer to skolespesifikke koeffisienter. Utover dette vil vektingen avhenge av antall elever. Det er få skoler som bidrar med mer enn to og mindre enn 20 elever, og disses beskjedne antall elever medfører også at disse får liten vekt.

Tabell 3.2. Elever i grunnskolen: Deskriptiv statistikk etter skriftlig eksamenskarakter

	Alle grunnskole- elever		Elever med skriftlig eksamenskarakter i:					
			Engelsk		Matematikk		Norsk	
	Snitt	Std.avvik	Snitt	Std.avvik	Snitt	Std.avvik	Snitt	Std.avvik
Skr eksamen			3,63	1,07	3,24	1,10	3,64	0,97
Stp engelsk	3,78	1,07	3,78	1,06	3,79	1,06	3,79	1,07
Stp matematikk	3,51	1,13	3,51	1,13	3,52	1,13	3,51	1,13
Stp norsk	3,90	0,97	3,90	0,96	3,91	0,96	3,89	0,97
Jente	0,49		0,49		0,49		0,49	
Mors utdanning (år)	12,99	2,98	12,97	2,95	13,01	2,96	12,98	3,03
Fars utdanning (år)	13,03	3,09	13,02	3,07	13,05	3,07	13,02	3,12
Innvandrere	0,02		0,02		0,02		0,02	
Norskfødt av innv.foreldre	0,02		0,02		0,02		0,02	
Log(inntekt)	8,42	0,60	8,42	0,59	8,42	0,59	8,44	0,59
Foreldre gift/samboende ..	0,68		0,69		0,69		0,68	
Antall	359	498	131	162	132	814	86	140

⁴ Se omtale av Tabell 3.3 og 3.4 og fotnote 8.

⁵ Ettersom vi krever at elevene skal ha registrert resultater fra grunnskolen, har ikke dette kravet noen stor praktisk betydning.

⁶ I praksis vil dette si skoler der minst 50 prosent av elevene kommer fra nevnte kommuner, denne måten å avgrense på har neppe noen betydning.

⁷ Elevene i Trondheim fordeler seg med minst ti elever på ni skoler på yrkesfag.

I den venstre delen i Tabell 3.2 gir vi en deskriptiv oversikt over karakterer og bakgrunnskarakteristika for alle grunnskoleelever. Gjennomsnittlig standpunkt karakter er høyest i norsk og lavest i matematikk. Som forventet er ca 50 prosent av elevene jenter. Fars gjennomsnittlige utdanning er litt høyere enn mors gjennomsnittlige utdanning. Litt i underkant av 70 prosent av elevene har gifte eller samboende foreldre. Videre ser vi at det er ingen tegn til systematisk sortering av elever på tvers av eksamen (se høyre del av Tabell 3.2), disse tallene er svært like for elever med hver av de skriftlige eksamenene.

Tabell 3.3. Elever i grunnskole og videregående: Deskriptiv statistikk etter fag fra videregående, allmennfag

	Alle elever fra samme grunnskoler		Med skriftlig standpunkt karakter i:									
			VF1000		VG1200		VG1330		VG1331		VG4000	
	Snitt	Std. avvik	Snitt	Std. avvik	Snitt	Std. avvik	Snitt	Std. avvik	Snitt	Std. avvik	Snitt	Std. avvik
Stp. vgs			3,99	0,85	4,05	1,00	4,13	1,12	2,88	1,07	3,86	0,99
Stp engelsk	3,89	1,04	4,33	0,85	4,15	0,92	4,51	0,81	4,05	0,85	4,24	0,87
Stp matematikk	3,60	1,14	4,13	0,96	3,89	1,03	4,56	0,80	3,47	0,82	4,01	0,99
Stp norsk	3,96	0,95	4,38	0,78	4,22	0,82	4,54	0,73	4,14	0,78	4,31	0,79
Jente	0,49		0,52		0,56		0,48		0,60		0,57	
Mors utdanning (år)	13,22	3,57	14,20	3,27	13,70	3,48	14,53	3,39	13,47	3,44	13,96	3,33
Fars utdanning (år)	13,47	3,67	14,52	3,41	13,99	3,58	14,99	3,52	13,65	3,41	14,26	3,49
Innvandrere	0,03		0,02		0,03		0,03		0,03		0,02	
Norskfødt av innv. foreldre	0,04		0,03		0,04		0,04		0,04		0,02	
Log(inntekt)	8,51	0,69	8,67	0,58	8,61	0,65	8,73	0,60	8,55	0,65	8,75	0,56
Foreldre gift/samboende	0,68		0,73		0,70		0,76		0,67		0,70	
Antall	74 110		14 821		22 034		9 810		6 215		23 261	

Tabell 3.4. Elever i grunnskole og videregående: Deskriptiv statistikk etter fag fra videregående, yrkesfag

	Alle avgangselever fra samme grunnskoler		Med skriftlig standpunkt karakter i:					
			VF1010		VF1210		VF1320	
	Snitt	Std. avvik	Snitt	Std. avvik	Snitt	Std. avvik	Snitt	Std. avvik
Stp vgs			3,59	0,94	3,54	1,03	3,21	1,27
Stp engelsk	3,89	1,04	3,36	0,95	3,34	0,95	3,33	0,95
Stp matematikk	3,60	1,14	3,00	0,96	3,00	0,97	2,99	0,96
Stp norsk	3,96	0,95	3,48	0,85	3,46	0,85	3,47	0,85
Jente	0,49		0,45		0,44		0,45	
Mors utdanning (år)	13,22	3,57	12,36	2,92	12,28	3,09	12,22	3,13
Fars utdanning (år)	13,47	3,67	12,53	2,94	12,49	3,02	12,44	3,05
Innvandrere	0,03		0,02		0,03		0,03	
Norskfødt av innv. foreldre	0,04		0,02		0,03		0,03	
Log(inntekt)	8,51	0,69	8,38	0,62	8,36	0,67	8,35	0,68
Foreldre gift/samboende	0,68		0,60		0,60		0,60	
Antall	74 110		13 336		12 675		13 454	

En deskriptiv oversikt over elevene som inngår i de analysene hvor standpunkt karakteren fra videregående brukes som et mål på eksternt sensur er gitt i Tabell 3.3 (allmennfag) og 3.4 (yrkesfag). De første kolonnene i disse tabellene baserer seg på alle elever ved de grunnskolene som inngår i denne analysen, merk at venstre del i Tabell 3.3 og 3.4 er like. Dette omfatter omtrent en femtedel av alle elevene, og gjennomsnittlig standpunkt karakterer fra grunnskolen er litt høyere for denne gruppen av elever sammenliknet med den forrige tabellen hvor alle grunnskoleelever inngikk. Det samme er også mors og fars utdanning. Forskjellene er imidlertid beskjedne. Kolonnene til høyre viser deskriptiv statistikk for elevene med de aktuelle standpunkt karakterene fra videregående. Dette antall elever er vesentlig mindre. Det skyldes dels at elevene fordeler seg på studieretninger og i noen grad fag, og dels at det er flere årskull som ligger til grunn når vi ser på grunnskolerresultater.⁸ Videre er det en klar sortering av elever til forskjellige studieretninger. De elevene som fortsetter på allmennfag har høyere standpunkt karakterer fra grunnskolen, høyere utdannede foreldre og høyere andel foreldre som er gift/samboende enn de elevene som fortsetter med yrkesfag. En tilsvarende, men mindre forskjell finnes mellom elever med de to matematikkvariantene VG1330 og VG1331.

⁸ Vi gjenfinner omtrent 85 prosent av elevene som er registrert med grunskolerresultater i hver av årgangene 2002-2005 på med resultater fra GK året etter, og ytterligere noen få prosent i senere år. Omtrent halvparten av de som er registrert i grunnskolen gjenfinnes med VG4000 tre år senere.

4. Resultater

I den første delen av dette kapittelet presenterer vi resultater som sier noe om sammenlikningen mellom standpunkt- og eksamenskarakterer i grunnskolen. Resultater relatert til sammenlikningen mellom standpunkt karakterer i grunnskolen og standpunkt karakterene i videregående er presentert i den andre delen.

4.1. Skriftlig eksamen som ekstern sensur

Resultatene fra estimering av likning (1) for alle fag med og uten kontrollvariable er rapportert i Tabell 4.1. Vi finner en sterk samvariasjon mellom eksamen og standpunkt, spesielt for matematikk hvor koeffisienten (denne svarer omtrent til den partielle korrelasjonen) er lik 0,82 i den spesifikasjonen som inkluderer kontrollvariable (kolonne (4)). Samvariasjonen mellom eksamen og standpunkt er svakest for norsk. For alle fag ser vi at korrelasjonen mellom standpunkt og eksamen er litt større i spesifikasjonen uten kontrollvariable (kolonne (1), (3) og (5)). Dette er ikke uventet ettersom familiebakgrunn er korrelert med prestasjonene på skolen, og disse korrelasjonene fanger opp noe av korrelasjonen mellom standpunkt og eksamen. På den annen side reduseres ikke korrelasjonene mellom standpunkt og eksamenskarakter markant etter inkluderingen av kontrollvariable, noe som skulle tilsi at *avviket* mellom standpunkt karakter og eksamensresultatet ikke er høyt korrelert med kontrollvariablene. Alle de seks korrelasjonene er statistisk signifikante til 5 prosents nivå.

Konstantleddet, α , kan som nevnt tolkes som den forventede standpunkt karakteren dersom alle forklaringsvariablene i likning (1) er lik gjennomsnittet.⁹ Av de øvrige kontrollvariablene ser vi at for gitt eksamenskarakter får jenter signifikant bedre standpunkt karakterer enn gutter i alle fag. Standpunkt karakterene (gitt eksamenskarakter) øker også med foreldres utdanning og inntekt. Elever med foreldre som er gift/samboere får også høyere standpunkt karakterer gitt deres eksamenskarakter enn elever med foreldre som ikke bor sammen, det får også innvandrere og norskfødte barn av innvandrere.

Tabell 4.1. Resultater fra å estimere likning (1). Avhengig variabel er standpunkt karakterene på grunnskolen i henholdsvis engelsk, matematikk og norsk (hovedmål).

	(1)	(2)	(3)	(4)	(5)	(6)
	ENS	ENS	MAT	MAT	NOH	NOH
Eks/ β	0,769** (0,002)	0,713** (0,002)	0,853** (0,001)	0,825** (0,002)	0,679** (0,002)	0,587** (0,003)
Jente		0,178** (0,004)		0,084** (0,003)		0,268** (0,005)
Klassetrinn,mors høyeste utd .		0,022** (0,001)		0,013** (0,001)		0,031** (0,001)
Klassetrinn,fars høyeste utd ...		0,016** (0,001)		0,010** (0,001)		0,020** (0,001)
Innvandrer		0,110** (0,014)		0,075** (0,013)		0,080** (0,019)
Norskfødt av innv.foreldre		0,115** (0,015)		0,117** (0,014)		0,142** (0,018)
Log(inntekt)		0,067** (0,004)		0,040** (0,003)		0,076** (0,005)
Foreldre gift/samboende		0,087** (0,004)		0,102** (0,004)		0,132** (0,005)
Konstantledd / α	3,756** (0,002)	3,780** (0,002)	3,470** (0,002)	3,506** (0,002)	3,835** (0,002)	3,870** (0,002)
Årsdummier inkl.	Nei	Ja	Nei	Ja	Nei	Ja
Antall observasjoner	146 808	131 536	150 443	134 642	97 940	87 522
R ²	0,60	0,61	0,69	0,70	0,47	0,51

Merk: Standardfeil i parentes.

** betyr at estimatene er statistisk signifikante til 5/1 prosents nivå.

⁹ Eksamenskarakter og alle øvrige kontrollvariable er målt som avvik fra gjennomsnittet, slik at alle variablene i Tabell 4.1 har gjennomsnitt lik 0.

Generell beskrivelse av skolespesifikke konstantleddene (μ_{jf}) og de skolespesifikke helningskoeffisientene (β_{jf})

Som allerede nevnt vil vi uttrykke de skolespesifikke konstantleddene (μ_{jf}) og de skolespesifikke helningskoeffisientene (β_{jf}) som skole j 's avvik fra de globale/gjennomsnittlige verdiene α og β fra ligning (1). I det følgende ser vi på i hvilken grad μ_{jf} og β_{jf} varierer på tvers av skoler. Er det slik at noen skoler avviker mye fra gjennomsnittet, mens at andre avviker mindre?

Tabell 4.2 gir en oversikt over gjennomsnittsverdien, standardavvikene, minimums- og maksimumsverdiene til *alle* de estimerte skolespesifikke konstantleddene, og *kun* de estimerte skolespesifikke konstantleddene som er statistisk signifikante. Med signifikant menes at vi er rimelig sikre på at forskjellene er systematisk og ikke skyldes tilfeldig variasjon. Dette gjelder 40-50 prosent av skolene, noe færre for norsk enn de andre fagene. En formell test om *noen* av skolene avviker signifikant sier også at dette er tilfelle.¹⁰ For disse estimatene skiller vi mellom skoler som undervurderer (streng skoler) og skoler som overvurderer (snille skoler) karakternivået på sin skole. De skolefaste konstantleddene er målt som avvik fra konstantleddet (α) i tabell 4.1, som kan tolkes som en form for gjennomsnittlig effekt for alle skolene i landet. Gjennomsnittet av de skolespesifikke konstantleddene er dermed nær null når vi ser på alle skoler under ett. Standardavviket (som gir uttrykk for variasjonen i karakterpraksis) til de skolespesifikke konstantleddene, er derimot ca 0,16-0,17 karakterpoeng i både matematikk, engelsk og norsk. Standardavviket (på elevnivå) til både standpunkt- og eksamenskarakter er til sammenligning omtrent ett karakterpoeng. Dette tilsier at det er nokså stor variasjon i karakterpraksis på tvers av skoler. Dersom vi kun betrakter gruppen av skoler med signifikante effekter øker standardavviket til ca 0,21-0,24. Omtrent halvparten av skolene har estimerte skolespesifikke konstantledd som er signifikante. Videre ser vi at standardavviket er høyere i den gruppen av skoler med positive signifikante skolespesifikke konstantledd (snille skoler) enn i gruppen av skoler med negative signifikante skolespesifikke effekter (streng skoler). Det er altså større spredning blant skolene som overvurderer elevenes karakternivå enn blant de som undervurderer.

Tabell 4.2. Deskriptiv oversikt over skolefaste konstantledd, μ_{jf} (målt som avvik fra globalt gjennomsnitt)

	Matematikk			Engelsk			Norsk					
	Alle	Signifikante		Alle	Signifikante		Alle	Signifikante				
		Alle	Pos		Neg	Alle		Pos	Neg	Alle	Pos	Neg
Gj.snitt	0,002	0,003	0,215	-0,171	0,005	0,009	0,203	-0,175	0,004	0,015	0,222	-0,201
St.avvik	0,165	0,212	0,130	0,077	0,162	0,213	0,112	0,084	0,166	0,235	0,109	0,095
Min	-0,863	-0,863	0,067	-0,863	-0,771	-0,771	0,060	-0,771	-0,575	-0,575	0,076	-0,575
Max	1,103	1,103	1,103	-0,057	0,923	0,923	-0,068	0,718	0,718	0,718	0,718	-0,081
Ant skoler ..	962	472	257	215	949	461	255	206	816	339	186	153

Merk: Alle verdier er vektet med antall elever. Grunnskoler med færre en 20 elever i løpet av den tidsperioden vi ser på er utelatt. Signifikant betyr statistisk signifikant til 5 prosents nivå.

Tabell 4.3 gir en oversikt over gjennomsnittsverdien, standardavviket, minimums- og maksimumsverdien til *alle* de estimerte β_{jf} 'ene (skolespesifikke helningskoeffisientene) og *kun* de estimerte β_{jf} 'ene som er signifikante. Vi skiller også her mellom signifikant positive β_{jf} 'er og signifikant negative β_{jf} 'er. Sammenliknet med μ_{jf} er standardavviket til β_{jf} mye lavere (mellom 0,07 og 0,09), men øker til 0,12-0,17 i den gruppen av skoler med signifikant β_{jf} . Når vi betrakter positive og negative β 'er hver for seg, ser vi at standardavviket er nokså likt. Merk også at blant de skolene med signifikant β_{jf} har flertallet en positiv β_{jf} , ved disse skolene er samvariasjonen mellom standpunkt og eksamenskarakter sterkere enn gjennomsnittet. Estimert β_{jf} er signifikant for ca 15 prosent av skolene. Dersom det bare var tilfeldig variasjon, og ingen systematiske forskjeller, ville vi likevel vente å finne at 5 prosent av skolene (feilaktig) fremsto som signifikant avvikende. Andelen skoler

¹⁰ En F-test av om alle de skolespesifikke konstantleddene er lik null gir testestimatorene 6,83, 8,09 og 4,94 for hhv. engelsk, matematikk og norsk. Med omtrent 1200 skolespesifikke koeffisienter og et stort data-materiale tilsier dette at vi på alle vanlige signifikansnivåer kan forkaste hypotesen om ingen forskjeller.

som faktisk avviker kan dermed være lavere enn 15 prosent. En formell test tilsier imidlertid at det er forskjeller mellom skoler i helningskoeffisientene.¹¹

Tabell 4.3. Deskriptiv oversikt over helningskoeffisienten, β_{jf} (målt som avvik fra globalt gjennomsnitt)

	Matematikk				Engelsk				Norsk			
	Alle		Signifikante		Alle		Signifikante		Alle		Signifikante	
	Alle	Pos	Neg		Alle	Pos	Neg		Alle	Pos	Neg	
Gj.snitt	0,015	0,042	0,118	-0,100	0,010	0,016	0,128	-0,124	0,010	0,007	0,185	-0,139
St.avvik ...	0,067	0,115	0,048	0,057	0,078	0,134	0,054	0,041	0,093	0,173	0,074	0,052
Min	-0,640	-0,640	0,062	-0,640	-0,295	-0,295	0,061	-0,295	-0,450	-0,450	0,093	-0,450
Max	0,450	0,450	0,450	-0,053	0,447	0,447	0,447	-0,065	0,540	0,540	0,540	-0,087
Ant skoler	962	146	106	42	949	172	115	57	816	121	72	49

Merk: Verdiene er vektet med antall elever. Grunnskoler med færre en 20 elever i løpet av den tidsperioden vi ser på er utelatt.

Et annet interessant spørsmål som dukker opp er om det er noen sammenheng mellom β_{jf} og μ_{jf} . For eksempel, er det slik at skoler som er snille/strengere er bedre/dårligere til å skille mellom flinke og mindre flinke elever? I Tabell 4.4 – 4.6 rapporterer vi korrelasjonskoeffisientene (vektet med antall elever) for hvert fag. Vi ser at korrelasjonen mellom β_{jf} og μ_{jf} er beskjeden i matematikk, og svært nær null i engelsk og norsk. Det er derfor ingenting som tyder på at skoler som overvurderer i nivå samtidig avviker i hvilken grad de skiller mellom sterke og svake elever.

Tabell 4.4. Korrelasjon mellom β_{jf} og μ_{jf} i matematikk

	μ_{jf}	β_{jf}	Nivåeffekt	Snitt eksamen
μ_{jf}	1			
β_{jf}	0,1284	1		
Nivåeffekt	0,9916	0,0909	1	
Snitt eksamen	-0,4148	-0,2650	-0,3945	1

Merk: Antall skoler = 1265. Verdiene er vektet med antall elever. Grunnskoler med færre en 20 elever i løpet av den tidsperioden vi ser på er utelatt. "Nivåeffekt" er et estimat av den totale nivåeffekten på den lokale gjennomsnittselev.

Tabell 4.5. Korrelasjon mellom β_{jf} og μ_{jf} i engelsk

	μ_{jf}	β_{jf}	Nivåeffekt	Snitt eksamen
μ_{jf}	1			
β_{jf}	0,0176	1		
Nivåeffekt	0,99111	0,0118	1	
Snitt eksamen	-0,3104	-0,3813	-0,3001	1

Merk: Antall obs = 948. Verdiene er vektet med antall elever. Grunnskoler med færre en 20 elever i løpet av den tidsperioden vi ser på er utelatt. "Nivåeffekt" er et estimat av den totale nivåeffekten på den lokale gjennomsnittselev.

Tabell 4.6. Korrelasjon mellom β_{jf} og μ_{jf} i norsk

	μ_{jf}	β_{jf}	Nivåeffekt	Snitt eksamen
μ_{jf}	1			
β_{jf}	-0,0082	1		
Nivåeffekt	0,9869	-0,0114	1	
Snitt eksamen	-0,2558	-0,2956	-0,2526	1

Merk: Antall obs = 815. Verdiene er vektet med antall elever. Grunnskoler med færre en 20 elever i løpet av den tidsperioden vi ser på er utelatt. "Nivåeffekt" er et estimat av den totale nivåeffekten på den lokale gjennomsnittselev.

Vi fortsetter med å se på hvordan μ_{jf} og β_{jf} korrelerer med: 1) estimatet av den totale nivåeffekten på skolens gjennomsnittselev og 2) med skolens snittkarakter på eksamen.

Den skolefaste effekten uttrykker forventet forskjell mellom intern og ekstern sensur for en gjennomsnittselev (denne gjennomsnittseleven er den samme for alle skoler). Ettersom vi også lar skolene variere i hvordan de skiller mellom sterke og svake elever vil imidlertid forskjellen mellom standpunkt og eksamen ikke være den samme for alle elever. Hvis en skole i tillegg har en elevmasse som avviker fra den gjennomsnittlige er det mulig at denne skolen systematisk gir sine elever for gode eller dårlige

¹¹ En F-test av om alle de skolespesifikke helningskoeffisientene er lik null gir testestimatene 1,90, 1,74 og 1,63 for hhv. engelsk, matematikk og norsk. Med omtrent 1200 skolespesifikke koeffisienter og et stort datamateriale tilsier dette at vi på alle vanlige signifikansnivåer kan forkaste hypotesen om ingen forskjeller. Vi ser imidlertid også at testestimatene er vesentlig lavere enn de tilsvarende for de skolespesifikke konstantleddene, som er et uttrykk for at helningskoeffisientene forklarer mindre av den variasjonen som finnes i dataene.

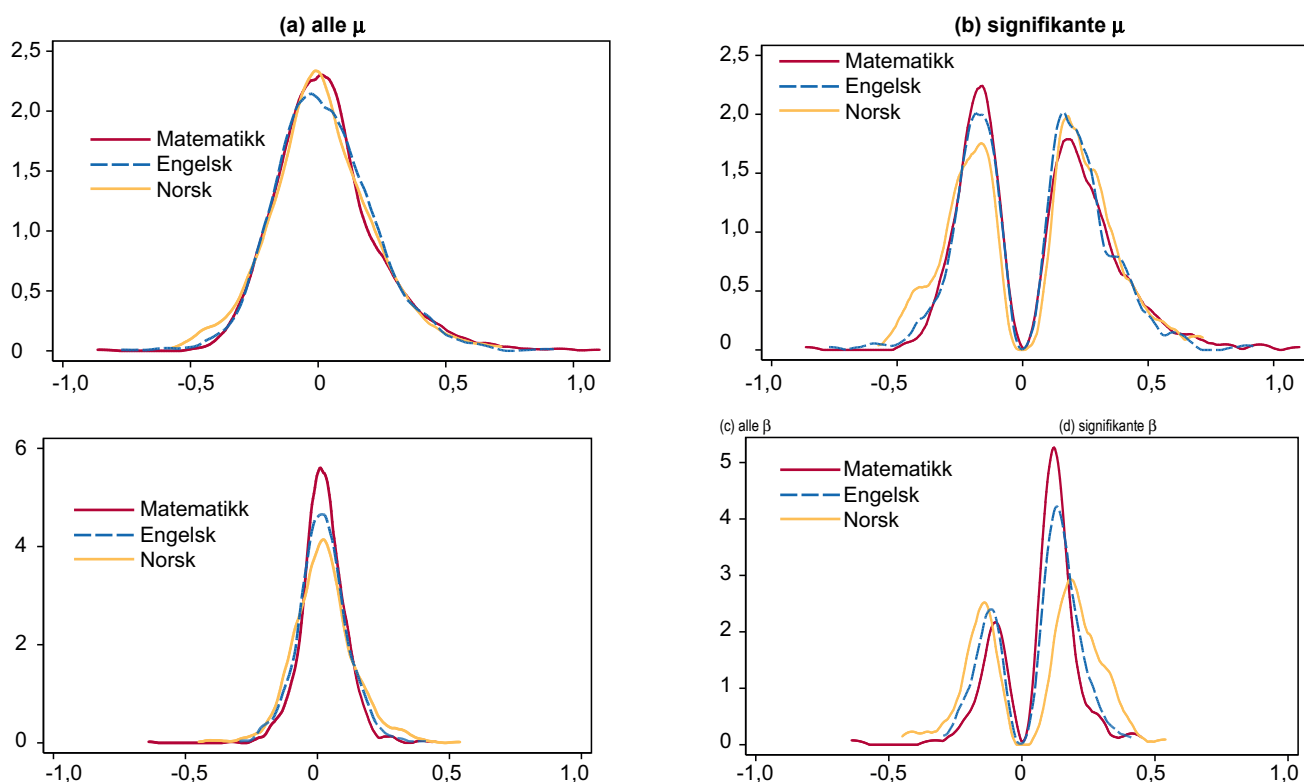
standpunkt karakterer, selv om den gir riktig karakter til en hypotetisk gjennomsnittselev (jf. diskusjonen rundt figur 2.2 D i avsnitt 2.3). Den totale nivåeffekten på skolens gjennomsnittselev gir summen av den skolefaste effekten og betydningen av et eventuelt avvik i den skolespesifikke samvariasjonen mellom intern og ekstern sensur (β_{if}) for en elev som er typisk for skolen. Fra tabell 4.4-4.6 ser vi at det er en svært høy korrelasjon mellom skolefast effekt og total nivåeffekt for snittelev, over 0,97. Korrelasjonen mellom total nivåeffekt og β_{if} er derimot nær null. Dette bekrefter inntrykket fra tabell 4.2 og 4.3 at det er de skolefaste konstantleddene, og ikke helningskoeffisientene β_{if} , som er kvantitativt viktige for forskjeller i karakterpraksis.

Til slutt ser vi at skolens snittkarakter på eksamen er negativt korrelert med de øvrige målene. Det vil si at skoler med elever som får gode eksamenskarakterer har en tendens til å både undervurdere disse i nivå, og også, om enn mindre kvantitativt viktig, i mindre grad skille mellom sterke og svake elever. Dette kan tyde på relativ karaktersetting, og gir en motivasjon for å studere forskjellene i karakterpraksis mellom skoler med sterkere og svakere elever nærmere.

Selv om standardavvikene i Tabell 4.2 og 4.3 gir en pekepinn på variasjonen i μ_{if} og β_{if} på tvers av skoler, er det også ønskelig å gi et mer helhetlig bilde av fordelingen. I det følgende ser vi nærmere på fordelingen til μ_{if} og β_{if} .

Figur 4.1 viser fordelingene av de estimerte skolespesifikke konstantleddene og helningskoeffisientene for hvert av fagene matematikk, norsk og engelsk i grunnskolen. Del (a) and del (c) av figurene viser hele fordelingen av de estimerte parametrene, mens del (b) og del (d) viser fordelingen av kun de *signifikante* estimerte parametrene, dvs. parametrene til de skolene hvor vi kan være rimelig sikre på at forskjellene er systematiske og ikke bare pga. tilfeldig variasjon i elevresultatene. Resultatene for de skolespesifikke konstantleddene er nokså like mellom fagene, mens det er noen forskjeller mellom fagene når det gjelder de skolespesifikke helningskoeffisientene. For disse det ser ut som at skoler som treffer dårlig ved den relative vurderingen av svake og sterke elever tenderer til å *overvurdere* den relative forskjellen mellom elever i matematikk og engelsk.

Figur 4.1. Fordeling av de estimerte skolespesifikke konstantleddene (μ_{if}) og helningskoeffisientene (β_{if}) i matematikk, norsk og engelsk

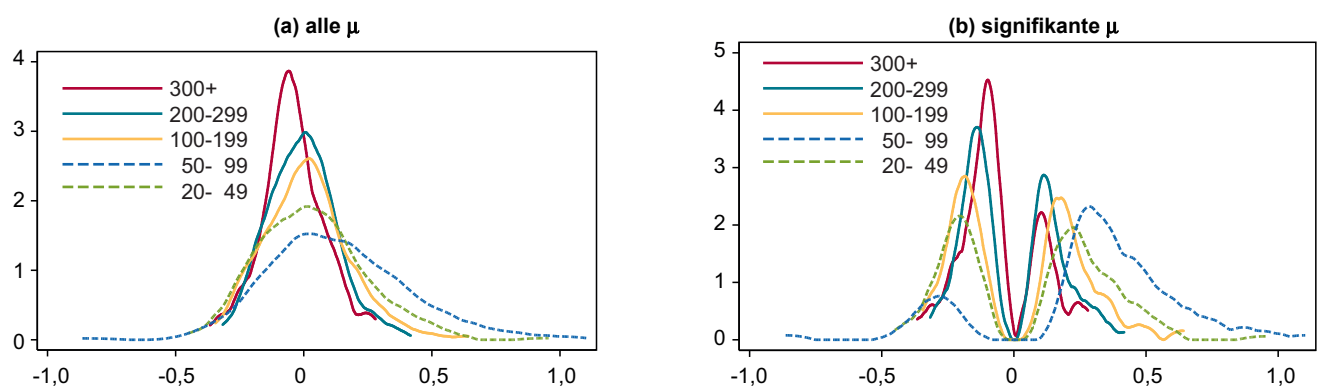


Fordelingen av de skolespesifikke konstantleddene etter skolestørrelse

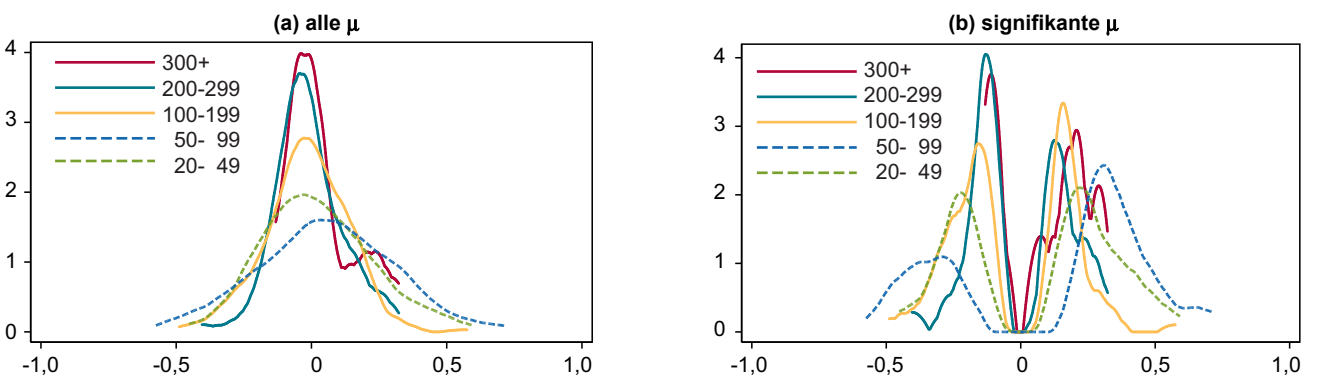
Figurene 4.2-4.4 viser hele fordelingen av de estimerte skolespesifikke konstantleddene (μ_{ij}) etter skolestørrelse for hvert av fagene matematikk, norsk hovedmål og engelsk (skriftlig). Det ser ut til å være noe mindre spredning i de estimerte skolespesifikke konstantleddene for norsk sammenlignet med de andre fagene, noe som ikke framkommer fra standardavvikene i tabell 4.1. Tilfeldig variasjon bidrar til større spredning i resultatene til de små skolene. Siden tilfeldig variasjon gjør det spesielt vanskelig å avdekke små signifikante forskjeller for små skoler, ser vi imidlertid at de minste *signifikante* avvikene for de små skolene er større sammenlignet med de større skolene i alle fag.

Skolestørrelsen ser ut til å ha betydning for elevvurderingene ved standpunkt-karakter sammenlignet med eksamenskarakter. Spesielt de små skolene (med bare 20-49 elever med eksamenskarakterene i fagene) skiller seg ut med en skjev fordeling hvor forholdsvis mange av de skolespesifikke konstantleddene er positive. Med andre ord, resultatene tyder på at små skoler har mer av en tendens til å overvurdere nivået til deres elevers prestasjoner ved standpunktkarakterene (sammenlignet med eksamenskarakterene). Det motsatte ser ut til å være tilfelle for de største skolene (med flere enn 300 elever) i matematikk og engelsk; resultatene tyder på at slike skoler heller *undervurderer* nivået til deres elever i de to fagene. For middels store skoler (med cirka 100-299 elever) ser det ut til at omtrent like mange skoler systematisk undervurderer som overvurderer nivået til deres elever. Med andre ord, nivåforskjeller i vurdering av elever ser ut til å variere systematisk med skolestørrelse for spesielt store og små skoler.

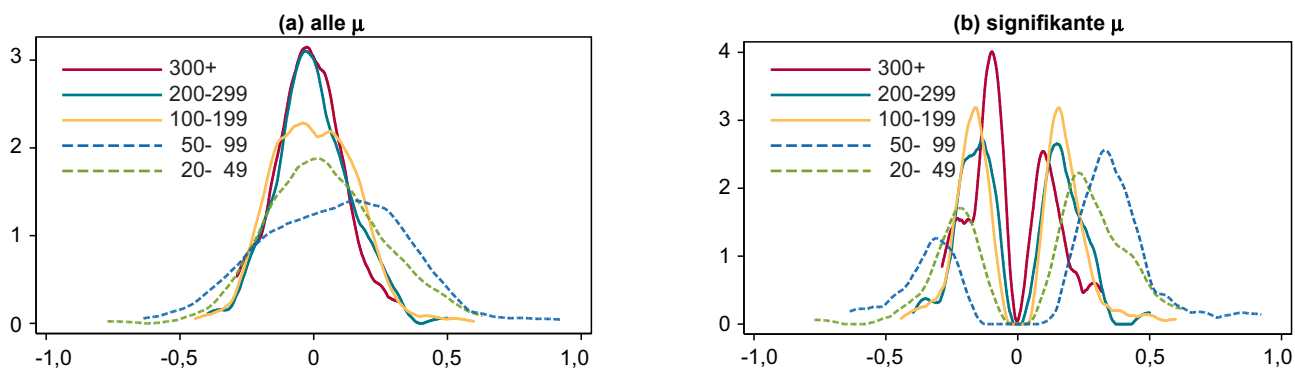
Figur 4.2. Fordeling av de estimerte skolespesifikke konstantleddene (μ_{ij}) i matematikk etter skolestørrelse



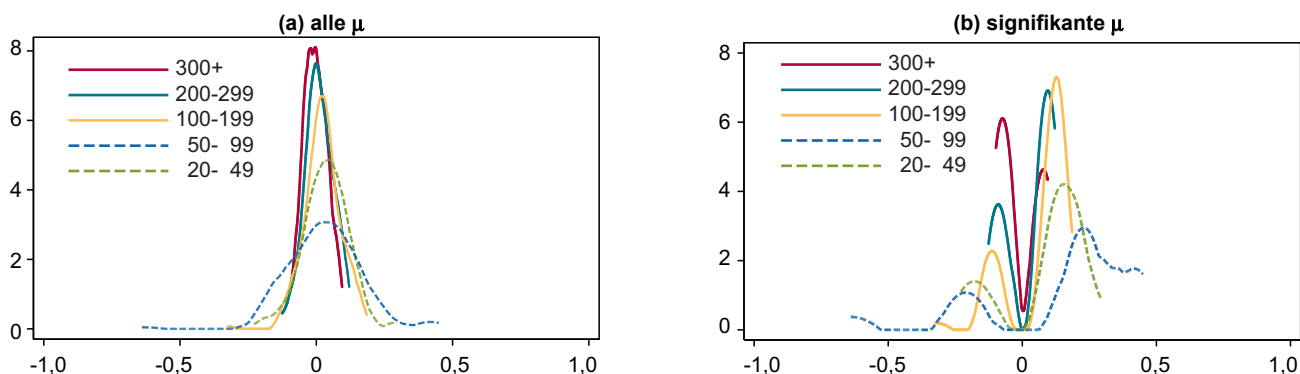
Figur 4.3. Fordeling av de estimerte skolespesifikke konstantleddene (μ_{ij}) i norsk etter skolestørrelse



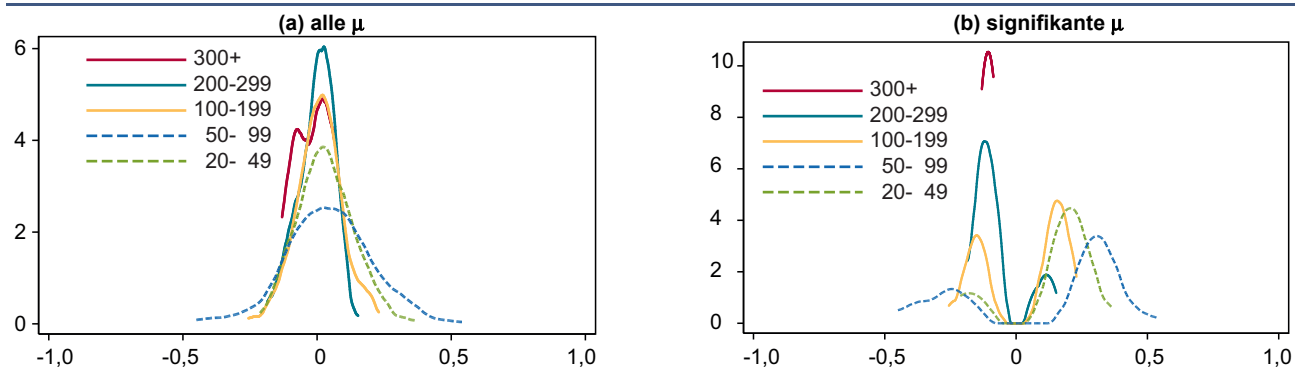
Figur 4.4. Fordeling av de estimerte skolespesifikke konstantleddene (μ_{jt}) i engelsk etter skolestørrelse



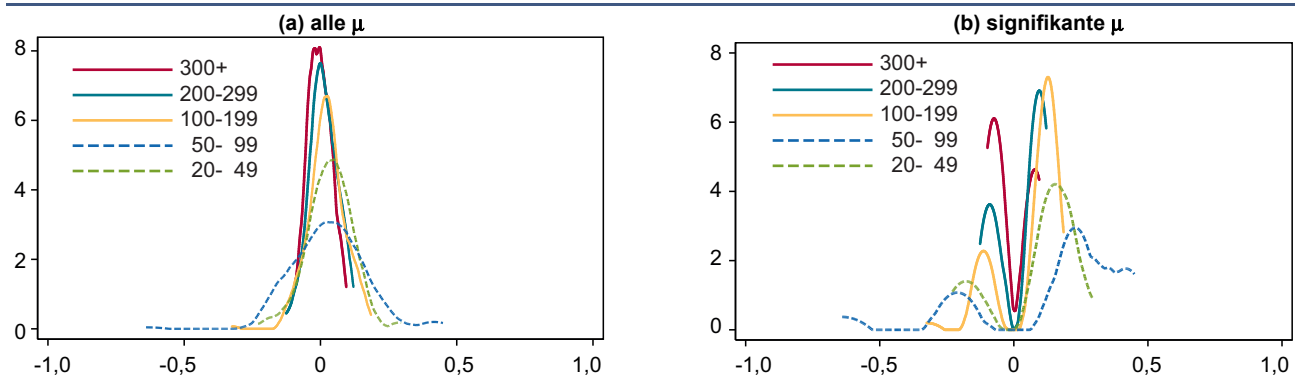
Figur 4.5. Fordeling av de estimerte skolespesifikke helningskoeffisientene (β_{jt}) i matematikk etter skolestørrelse



Figur 4.6. Fordeling av de estimerte skolespesifikke helningskoeffisientene (β_{jt}) i norsk etter skolestørrelse



Figur 4.7. Fordeling av de estimerte skolespesifikke helningskoeffisientene (β_{jt}) i engelsk etter skolestørrelse

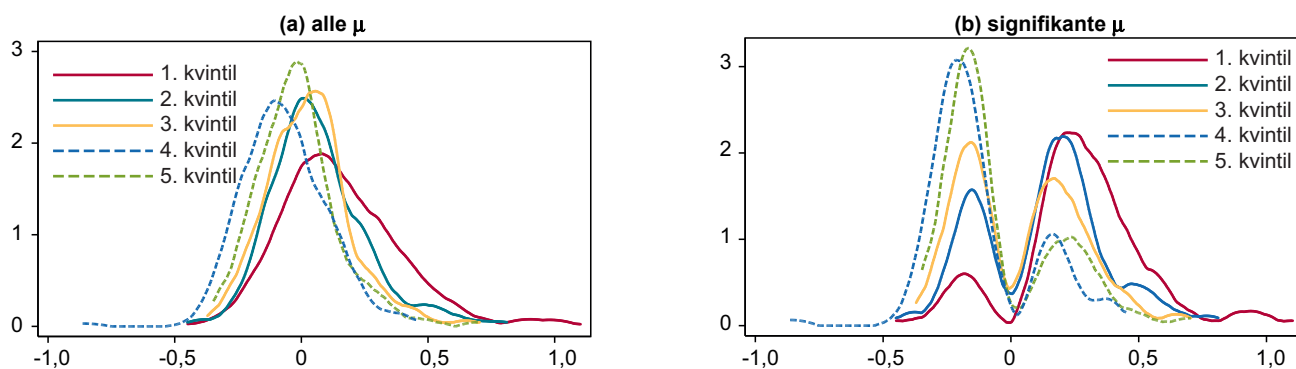


Figurene 4.5-4.7 viser hele fordeling etter skolestørrelse av de estimerte skolespesifikke helningskoeffisientene (β_{jf}) for hvert av fagene matematikk, norsk hovedmål og engelsk (skriftlig). Igjen viser del (a) av figurene hele fordeling av de estimerte parametrene mens del (b) viser fordeling av kun de *signifikante* estimerte parametrene. Spredning av de skolespesifikke helningskoeffisientene i matematikk ser ut til å være noe større enn i de andre fagene. Når vi ser på fordelingen av de signifikante helningskoeffisientene (del (b) av figurene) er det noe tegn til at de små skolene overvurderer den relative forskjellen mellom svake og sterke elever i alle tre fag, mens de største skolene undervurderer den relative forskjellen mellom svake og sterke elever i engelsk. I den grad mellomstore skoler avviker ved deres skolespesifikke helningskoeffisientene er det omtrent like mange som har positive som negative β_{jf} .

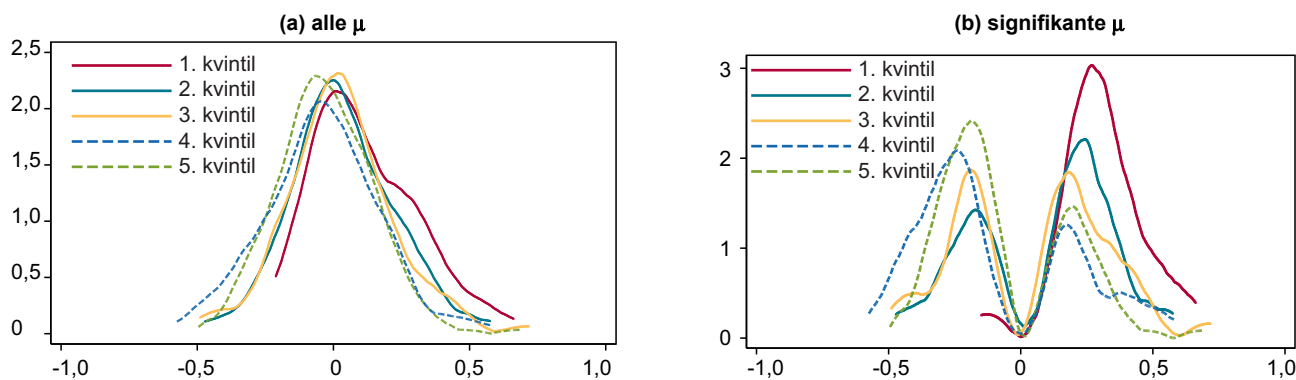
Fordelingen av de skolespesifikke konstantleddene etter skolens gjennomsnittlig eksamenskarakter

Figurene 4.8-4.10, som viser skolespesifikke konstantleddene etter skolens gjennomsnittlig eksamenskarakter, kan tyde på et innslag av relativ karaktersetting for standpunktvurderingene i ungdomsskolen. Skolene med laveste gjennomsnittlige eksamenskarakterer ser ut til å i større grad overvurderer deres elever i nivå ved fastsetting av standpunkt karakterer, mens skolene med høyeste gjennomsnittlige eksamenskarakterer ser heller ut til å undervurdere prestasjonsnivået til deres elever. Blant den femdelen av skoler som befinner seg midt på treet i eksamenspresentasjoner (3. kvintil) er det omtrent like mange som undervurderer som overvurderer deres elever i nivå ved fastsetting av standpunkt karakterer.

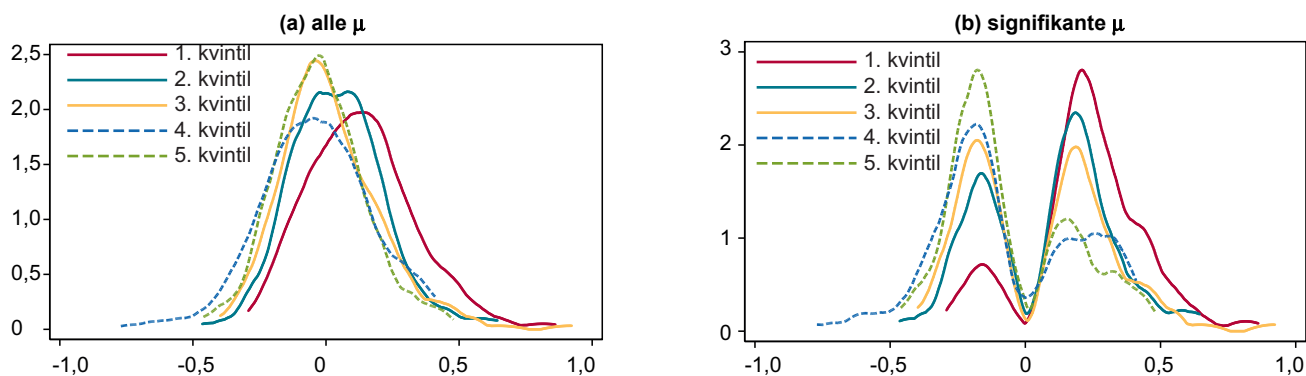
Figur 4.8. Fordeling av de estimerte skolespesifikke konstantleddene (μ_{jf}) i matematikk etter skolens gjennomsnittlig eksamensprestasjon (i kvintiler)



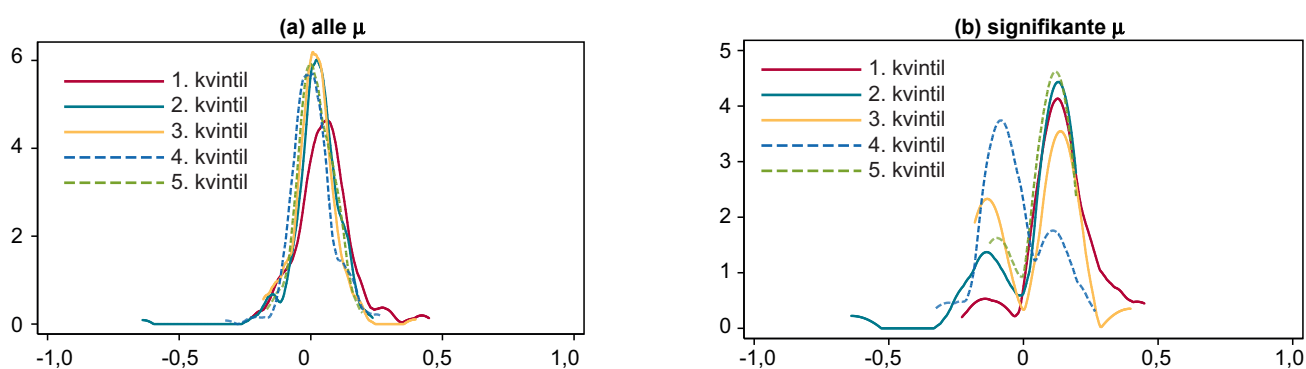
Figur 4.9. Fordeling av de estimerte skolespesifikke konstantleddene (μ_{jf}) i norsk etter skolens gjennomsnittlig eksamensprestasjon (i kvintiler)



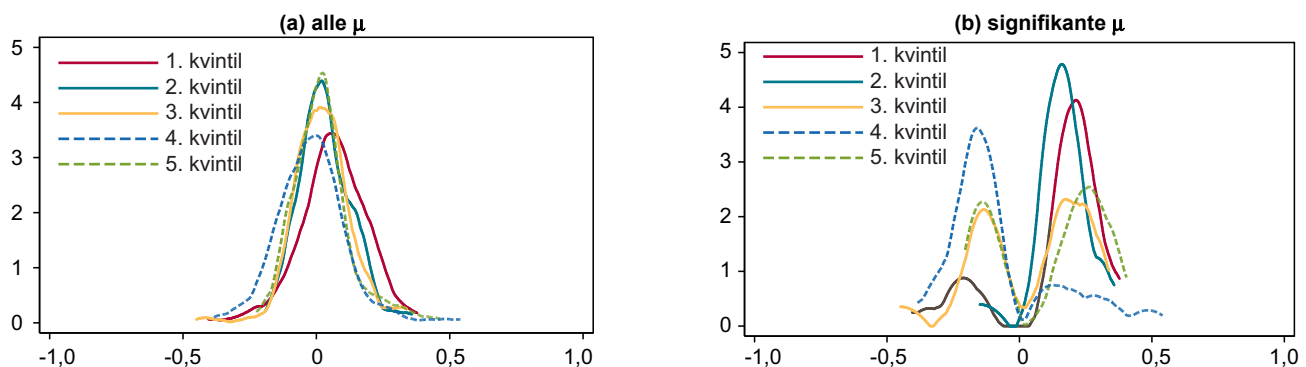
Figur 4.10. Fordeling av de estimerte skolespesifikke konstantleddene (μ_{jt}) i engelsk etter skolens gjennomsnittlig eksamensprestasjon (i kvintiler)



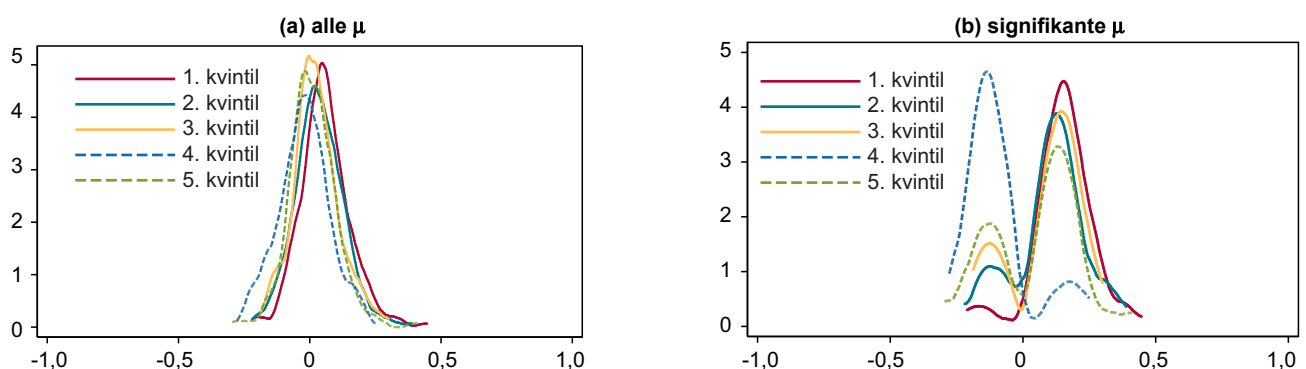
Figur 4.11. Fordeling av de estimerte skolespesifikke helningskoeffisientene (β_{jt}) i matematikk etter skolens gjennomsnittlig eksamensprestasjon (i kvintiler)



Figur 4.12. Fordeling av de estimerte skolespesifikke helningskoeffisientene (β_{jt}) i norsk etter skolens gjennomsnittlig eksamensprestasjon (i kvintiler)



Figur 4.13. Fordeling av de estimerte skolespesifikke helningskoeffisientene (β_{jt}) i engelsk etter skolens gjennomsnittlig eksamensprestasjon (i kvintiler)



Et klart mønster kan også leses når vi ser på fordelingen av de skolespesifikke helningskoeffisientene i Figurene 4.11-4.13. I de få tilfellene hvor skolen treffer dårlig i dens relative vurdering av svake er sterke elever, ser det ut som at skolene med lavere gjennomsnittskarakterer i større grad overvurderer den relative forskjellen mellom svakere og sterke elever mens skolene med høye gjennomsnittskarakterer undervurderer forskjellen. I denne sammenhengen er det viktig å huske at det er forholdsvis få skoler med signifikante estimerte β_{jf} . Forholdet som beskrives i Figur 4.11-4.13 gjelder derfor for et svært begrenset antall skoler.

Samvariasjon mellom μ_{jf} og β_{jf} på tvers av fag

I tillegg til å se på fordelingen av μ_{jf} og β_{jf} på tvers av skoler, er det også interessant å se på samvariasjonen mellom μ_{jf} og β_{jf} på tvers av fag. I Tabell 4.7 – 4.9 presenterer vi korrelasjoner i μ_{jf} på tvers av fag, mens at vi i Tabell 4.10 – 4.12 presenterer korrelasjoner i β_{jf} på tvers av fag.

De skolespesifikke konstantleddene er høyt korrelert på tvers av fag. Dette betyr at en skole som er snill i sin karaktersetning i matematikk også oftere er snill i sin karaktersetning i engelsk og norsk (se Tabell 4.7). Forskjellige fag betyr her også forskjellige elever, oftest forskjellige faglærere og ofte forskjellige årganger. Dette tyder igjen på at avvikene i de enkelte fagene i noen grad skyldes underliggende egenskaper ved skolene, og ikke bare for eksempel enkeltlæreres karaktersetning. Disse korrelasjonene endres minimalt når vi kun ser på de skolene hvis skolefaste effekt i både matematikk, engelsk og norsk er signifikant til 5 prosents nivå (Tabell 4.8), eller de skolene hvis skolefaste effekt i både matematikk, engelsk og norsk er signifikant til 10 prosents nivå (Tabell 4.9).

Tabell 4.7. Korrelasjon mellom fag i skolefaste effekter, μ_{jf} (alle)

	Matematikk	Engelsk	Norsk
Matematikk	1		
Engelsk	0,5341	1	
Norsk	0,5012	0,5978	1

Merk: Ant skoler = 747. Grunnskoler med færre en 20 elever i løpet av den tidsperioden vi ser på er utelatt.

Tabell 4.8. Korrelasjoner mellom fag i skolefaste effekter, μ_{jf} (statistisk signifikant til 5 % nivå)

	Matematikk	Engelsk	Norsk
Matematikk	1		
Engelsk	0,6641	1	
Norsk	0,6242	0,6127	1

Merk: Ant skoler = 89. Grunnskoler med færre en 20 elever i løpet av den tidsperioden vi ser på er utelatt.

Tabell 4.9. Korrelasjoner mellom fag i skolefaste effekter, μ_{jf} (statistisk signifikant til 10 % nivå)

	Matematikk	Engelsk	Norsk
Matematikk	1		
Engelsk	0,5733	1	
Norsk	0,4895	0,5497	1

Ant skoler = 133. Grunnskoler med færre en 20 elever i løpet av den tidsperioden vi ser på er utelatt.

Det er liten grad av samvariasjon mellom fag i β_{jf} (se Tabell 4.10). Det vil si, at en skole har sterk samvariasjon mellom standpunkt og eksamen i matematikk betyr ikke at den har like sterk samvariasjon mellom standpunkt og eksamen i norsk og engelsk.

Når vi kun betrakter de skolene som har signifikant β_{jf} (Tabell 4.11 og 4.12) i både matematikk, engelsk og norsk øker korrelasjonene, men på grunn av få observasjoner må vi være forsiktig med å tillegge disse funnene for stor vekt. I tillegg er det også nokså stor variasjon i korrelasjonene mellom Tabell 4.8 (β_{jf} er statistisk signifikant til 5 prosents nivå) og Tabell 4.9 (β_{jf} er statistisk signifikant til 10 prosents nivå), et ytterligere tegn på at disse funnene må tolkes med varsomhet.

Tabell 4.10. Korrelasjon mellom fag i β_{jf} (alle)

	Matematikk	Engelsk	Norsk
Matematikk	1		
Engelsk	0,0983	1	
Norsk	0,0608	0,1426	1

Merk: Ant skoler=746. Grunnskoler med færre en 20 elever i løpet av den tidsperioden vi ser på er utelatt.

Tabell 4.11. Korrelasjon mellom fag i β_{jf} (signifikante til 5 prosents nivå)

	Matematikk	Engelsk	Norsk
Matematikk	1		
Engelsk	0,5850	1	
Norsk	0,3211	0,0489	1

Merk: Ant skoler = 9. Grunnskoler med færre en 20 elever i løpet av den tidsperioden vi ser på er utelatt.

Tabell 4.12. Korrelasjon mellom fag i β_{jf} (signifikante til 10 prosents nivå)

	Matematikk	Engelsk	Norsk
Matematikk	1		
Engelsk	0,4541	1	
Norsk	0,3277	0,5483	1

Merk: Ant skoler = 18. Grunnskoler med færre en 20 elever i løpet av den tidsperioden vi ser på er utelatt.

Robusthetsanalyser

For å undersøke og bekrefte robustheten av hovedfunnene ble modellen også estimert med flere alternative spesifikasjoner, hvor hovedspesifikasjonen (baseline) er den som ble brukt for å estimere de hittil omtalte resultatene. Tabell 4.13 formidler noen hovedinnsikter fra de alternative spesifikasjonene ved å sammenligne resultatene vedrørende de skolespesifikke konstantleddene i matematikk. De skolespesifikke helningskoeffisientene undersøkes ikke nærmere siden få av disse var betydningsfulle i hovedspesifikasjonen.

Tabell 4.13. Robusthet: beskrivelse av fordeling av de skolespesifikke konstantleddene ved estimering med alternative spesifikasjoner i faget matematikk.

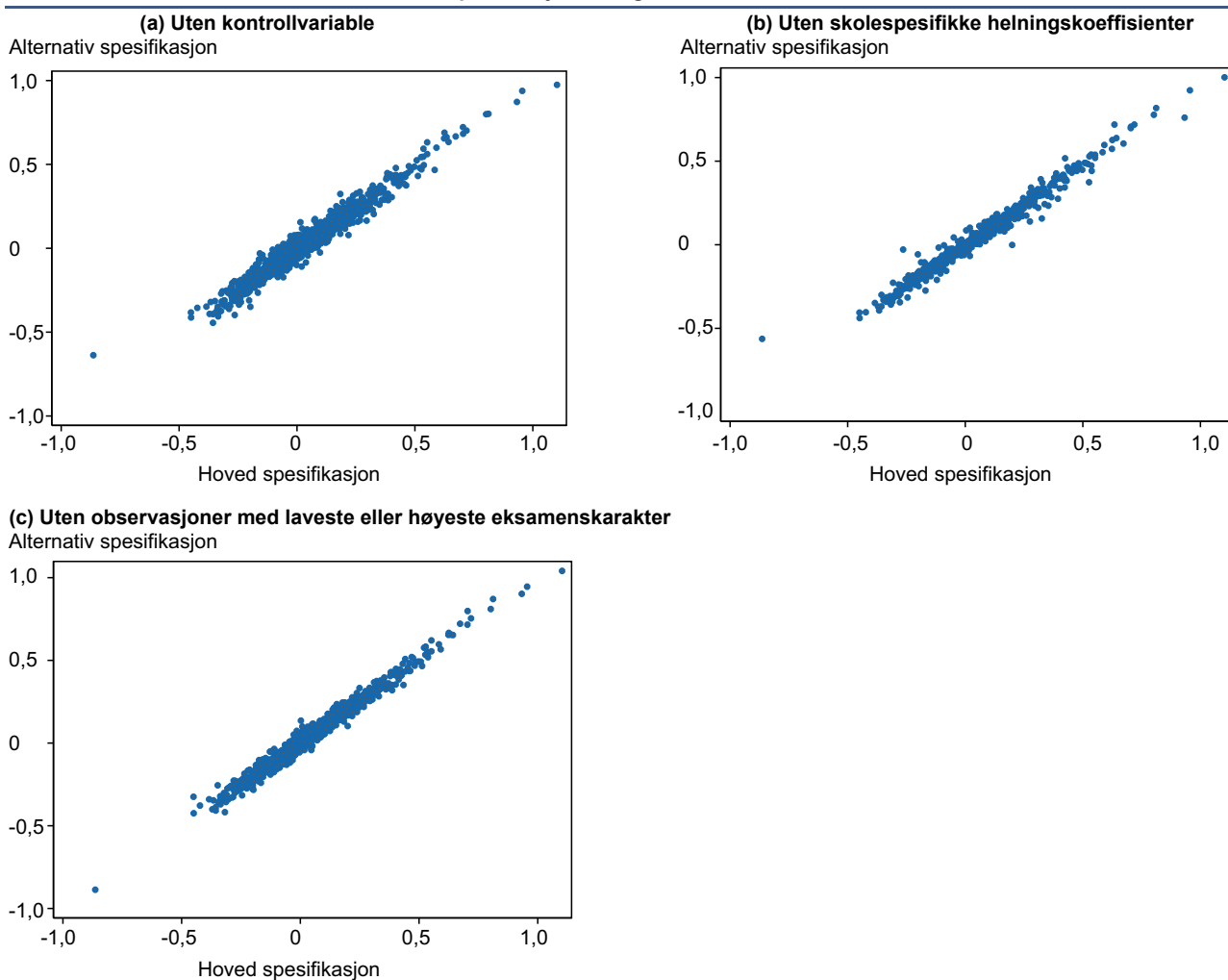
	Spesifikasjon				
	Baseline	(1)	(2)	(3)	(4)
Standardavvik μ_j	0,2060	0,2119	0,1989	0,2071	0,2060
Andel signifikante μ_j	0,4912	0,4934	0,4995	0,4932	0,4880
Korrelasjon til baseline μ_j		0,9207	0,9878	0,9974	0,9998
Minste differanse til baseline μ_j		-1,4931	-0,2468	-0,0925	-0,0152
Største differanse til baseline μ_j		1,7020	0,3001	0,1010	0,0166
Med kontrollvariable	x		x	x	x
Med skolespesifikke β	x	x		x	x
Med årsspesifikke β					x
Uten karakterene 1 og 6				x	

Den første alternative spesifikasjonen er en spesifikasjon uten kontrollvariable. Den er oppført i kolonnen (1) i Tabell 4.13 og presenteres for å understreke viktigheten av å inkludere kontrollvariablene knyttet til elevenes sosioøkonomiske bakgrunn. Korrelasjonen mellom de skolespesifikke konstantleddene estimert ved den spesifikasjonen og de skolespesifikke konstantleddene estimert ved hovedspesifikasjonen er 0,92. Den er høy, men betyr allikevel at noen av de skolespesifikke konstantleddene avviker nokså mye fra det som ble estimert med hovedspesifikasjonen. Dette viser at det er viktig å ta hensyn til kjennetegn som beskriver elevene sosialbakgrunn for å få et mer riktig bilde av hvordan standpunkt- og eksamenskarakterene avviker på enkelte skoler.

Den andre spesifikasjonen er estimert uten skolespesifikke helningskoeffisientene, dvs. er en noe mer restriktiv modell enn hovedspesifikasjonen. Funnene angående de skolespesifikke konstantleddene avviker lite mellom den spesifikasjonen og hovedspesifikasjonen.

Den tredje spesifikasjonen estimeres uten observasjoner med den laveste (1) eller høyeste (6) eksamenskarakteren for å sjekke om noen av hovedresultatene er følsomme for topp- eller bunnkoding. Funnene angående de skolespesifikke konstantleddene avviker lite mellom den spesifikasjonen og hovedspesifikasjonen.

Figur 4.14. Sammenligning av de skolespesifikke konstantleddene estimert ved hovedspesifikasjon med de skolespesifikke konstantleddene med alternativspesifikasjoner i faget matematikk



Den siste spesifikasjonen er mindre restriktiv enn hovedspesifikasjonen ved at det tillater både årsspesifikke konstantledd og årsspesifikke helningskoeffisienter. Resultatene mht. de skolespesifikke konstantleddene avviker lite mellom den spesifikasjonen og hovedspesifikasjonen.

Oppsummert finner vi at det er en høy korrelasjon mellom baseline-spesifikasjonene og alle de alternative spesifikasjonene. Figur 4.14 viser forholdet mellom de estimerte skolespesifikke konstantleddene i hovedspesifikasjonen og alternativspesifikasjonene for hver enkel skole. Få av punktene avviker mye fra 45-graders linje og viser dermed at resultatene er nokså sammenfallende for de aller fleste skoler.

Tabell 4.14 og Tabell 4.15 sammen med Figur B.1 og Figur B.2 i vedlegg B presenterer de samme funnene for henholdsvis norsk og engelsk, og tegner stort sett samme bilde som robusthetsanalysene for matematikk.

Tabell 4.14. Robusthet: beskrivelse av fordeling av de skolespesifikke konstantleddene ved estimering med alternative spesifikasjoner i faget norsk

	Spesifikasjon				
	Baseline	(1)	(2)	(3)	(4)
Standavvik μ_j	0,2004	0,2160	0,1972	0,2015	0,2003
Andel signifikante μ_j	0,4160	0,4291	0,4233	0,4069	0,4160
Korrelasjon til baseline μ_j		0,8702	0,9866	0,9964	0,9998
Minste differanse til baseline μ_j		-1,6728	-0,2385	-0,0771	-0,0201
Største differanse til baseline μ_j		0,6941	0,1915	0,1370	0,0174
Med kontrollvariable	x		x	x	x
Med skolespesifikke β	x	x		x	x
Med årsspesifikke β					x
Uten karakterene 1 og 6				x	

Tabell 4.15. Robusthet: beskrivelse av fordeling av de skolespesifikke konstantleddene ved estimering med alternative spesifikasjoner i faget engelsk

	Spesifikasjon				
	Baseline	(1)	(2)	(3)	(4)
Standavvik μ_j	0,1992	0,1925	0,2021	0,1991	0,2015
Andel signifikante μ_j	0,4863	0,4757	0,4835	0,4810	0,4598
Korrelasjon til baseline μ_j		0,9881	0,9691	9,9999	0,9820
Minste differanse til baseline μ_j		-0,3204	-0,4689	-0,0077	-0,1605
Største differanse til baseline μ_j		0,1320	0,1989	0,0073	0,2741
Med kontrollvariable	x		x	x	x
Med skolespesifikke β	x	x		x	x
Med årsspesifikke β					x
Uten karakterene 1 og 6				x	

4.2. Standpunkt karakter fra videregående som ekstern sensur

Resultatene fra å estimere ligning (1) er presentert i Tabell 4.16 (norsk hovedmål skriftlig), Tabell 4.17 (engelsk skriftlig) og Tabell 4.18 (matematikk). Vi estimerer likning (1) med og uten kontrollvariable for alle fag. Totalt er det 16 spesifikasjoner, tilsvarende hva vi gjorde med skriftlig eksamen fra grunnskolen som ekstern sensur i forrige avsnitt. Som omtalt i kapittel 3 er det begrensninger i data som gjør at bare noen fag fra videregående er aktuelle som ekstern sensur. Videre betegner A1 allmenn fag 1. klasse, Y1 yrkesfag 1. klasse og A3 allmenn fag 3. klasse. Merk også at det er to matematikk linjer for allmenn fag 1. klasse, nemlig VG1330 og VG1331, der VG1330 er mer teoretisk og ment for elever som tenker å fortsette med matematikk og VG1331.

Vi finner at det er en sterk sammenheng mellom standpunkt karakterene på videregående og standpunkt karakterene på grunnskolen. Sammenhengen er sterkest for VF1320 (matematikk, 1. klasse yrkesfag) og svakest for VF1000 (norsk, 1. klasse allmenn). For norsk og engelsk er sammenhengene vesentlig svakere enn tilsvarende sammenhenger mellom standpunkt og eksamens karakter i grunnskolen i avsnitt 4.1.

Når det gjelder de andre kontrollvariablene ser vi at betinget på standpunkt karakteren fra videregående, får jenter bedre standpunkt karakter i ungdomsskolen enn gutter, gitt eksamens karakteren. Standpunkt karakterene øker også stort sett med foreldres utdanning og inntekt (spesielt for engelsk), mens innvandringsbakgrunn ser ut til å være av mindre betydning (spesielt for enkelte fag). Kort oppsummert, hvor viktig foreldrenes bakgrunn er ser ut til å variere på tvers av fag og fagretning.

Tabell 4.16. Resultater fra å estimere ligning (1). Avhengig variabel er standpunktkarakteren i norsk hovedmål fra grunnskolen

	(1)	(2)	(3)	(4)	(5)	(6)
	VF1000	VF1000	VF1010	VF1010	VG4000	VG4000
	A1	A1	Y1	Y1	A3	A3
Standpkt vgs/ β	0,492** (0,018)	0,424** (0,020)	0,526** (0,019)	0,473** (0,022)	0,574** (0,016)	0,496** (0,018)
Jente		0,273** (0,035)		0,124** (0,037)		0,155** (0,031)
Klassetrinn, mors høyeste utd ...		0,014* (0,006)		0,015* (0,007)		0,037** (0,005)
Klassetrinn, fars høyeste utd		0,015** (0,006)		0,006 (0,007)		0,036** (0,005)
Innvandrer		-0,266* (0,104)		-0,007 (0,122)		0,012 (0,112)
Norskfødt av innv.foreldre		-0,004 (0,085)		0,163 (0,110)		0,189* (0,088)
Log(inntekt)		0,071* (0,033)		0,092** (0,032)		0,011 (0,032)
Foreldre gift/samboende		0,032 (0,038)		0,120**		0,098** (0,032)
Konstantledd / α	3,968** (0,016)	4,024** (0,017)	3,527** (0,016)	3,526** (0,017)	3,878** (0,014)	3,915** (0,015)
Årsdummier inkl.	Nei	Ja	Nei	Ja	Nei	Ja
Antall observasjoner	2 176	1 809	2 675	2 276	3 805	3 311
R ²	0,25	0,30	0,22	0,23	0,25	0,29

Merk: Standardfeil i parentes. */** betyr at estimatene er statistisk signifikante til 5/1 prosents nivå.

Tabell 4.17. Resultater fra å estimere ligning (1). Avhengig variabel er standpunktkarakter i engelsk skriftlig fra grunnskolen

	(7)	(8)	(9)	(10)
	VG1200	VG1200	VF1210	VF1210
	A1	A1	Y1	Y1
Standpkt vgs/ β	0,684** (0,009)	0,616** (0,011)	0,655** (0,012)	0,613** (0,014)
Jente		0,045* (0,019)		-0,006 (0,026)
Klassetrinn, mors høyeste utd		0,022** (0,003)		0,012* (0,005)
Klassetrinn, fars høyeste utd		0,021** (0,003)		0,012* (0,005)
Innvandrer		0,163** (0,056)		-0,114 (0,080)
Norskfødt av innv.foreldre		0,006 (0,056)		-0,013 (0,079)
Log(inntekt)		0,071** (0,017)		0,068** (0,022)
Foreldre gift/samboende		0,038 (0,021)		0,098** (0,027)
Konstantledd / α	4,018** (0,009)	4,062** (0,009)	3,496**	3,530** (0,013)
Årsdummier inkl.	Nei	Ja	Nei	Ja
Antall observasjoner	7 597	6 555	4 853	4 112
R ²	0,42	0,43	0,38	0,38

Merk: Standardfeil i parentes. */** betyr at estimatene er statistisk signifikante til 5/1 prosents nivå.

Tabell 4.18. Resultater fra å estimere ligning (1). Avhengig variabel er standpunktkarakteren i matematikk fra grunnskolen

	(11) VG1330 A1	(12) VG1330 A1	(13) VG1331 A1	(14) VG1331 A1	(15) VF1320 Y1	(16) VF1320 Y1
Standpkt vgs/ β	0,862** (0,017)	0,859** (0,020)	0,791** (0,022)	0,788** (0,026)	0,933** (0,013)	0,908** (0,014)
Jente		0,126** (0,032)		0,217** (0,044)		0,015 (0,027)
Klassetrinn, mors høyeste utd ...		0,004 (0,006)		0,010 (0,008)		0,001 (0,005)
Klassetrinn, fars høyeste utd		0,020** (0,005)		0,010 (0,007)		0,010 (0,005)
Innvandrere		0,275** (0,102)		0,086 (0,127)		-0,176* (0,079)
Norskfødt av innv.førelde		0,184* (0,082)		0,220 (0,119)		0,020 (0,079)
Log(inntekt)		0,040 (0,031)		0,045 (0,039)		0,023 (0,024)
Foreldre gift/samboende		0,084* (0,037)		0,181** (0,046)		0,113** (0,028)
Konstantledd / α	4,102** (0,015)	4,143** (0,016)	2,867** (0,020)	2,932** (0,021)	3,175** (0,012)	3,240** (0,013)
Årsdummier inkl.	Nei	Ja	Nei	Ja	Nei	Ja
Antall observasjoner	3038	2667	1686	1383	5125	4285
R ²	0,45	0,46	0,44	0,47	0,52	0,52

Merk: Standardfeil i parentes. */** betyr at estimatene er statistisk signifikante til 5/1 prosents nivå.

Generell beskrivelse av skolespesifikke konstantleddene (μ_{if}) og de skolespesifikke helningskoeffisientene (β_{if})

En deskriptiv oversikt over de estimerte skolefaste effektene (målt som avvik α estimert uten skoleeffekter, som omtrent svarer til gjennomsnittlig skolefast effekt) fra å estimere likning (4) er presentert i Tabell 4.19 (norsk), Tabell 4.20 (engelsk) og Tabell 4.21 (matematikk). Også her er gjennomsnittlig skolespesifikke konstantledd nær null dersom alle effektene betraktes. Derimot er gjennomsnittet av de signifikante skolespesifikke effekter langt fra null, noe som skyldes svært få observasjoner i den gruppen. Standardavviket til μ_{if} er veldig høyt, noe som indikerer at det er stor variasjon i μ_{if} på tvers av skoler. Det svært lave antallet signifikante estimater betyr imidlertid at resultatene er så usikre at det ikke kan legges noen vesentlig vekt på dem.

Tabell 4.19. Deskriptiv oversikt over skolefaste konstantledd (μ_{if}) i norsk (målt som avvik fra gjennomsnitt)

	VF1000		VF1010		VG4000	
	A1 Alle	A1 Sign	Y1 Alle	Y1 Sign	A3 Alle	A3 Sign
Gjennomsnitt	0,0201	-0,9935	-0,0155		0,0113	0,7838
St.avvik	0,5493	0,0796	0,6016		0,2857	0
Min	-1,1292	-1,1292	-0,9849		-0,6563	
Max	1,4109	-0,9006	1,3173		0,7838	
Ant skoler	41	4	47	0	66	1

Merk: Verdiene er vektet med antall elever. Grunnskoler med færre en 20 elever i løpet av den tidsperioden vi ser på er utelatt. Signifikant betyr statistisk signifikant til 5 prosents nivå.

Tabell 4.20. Deskriptiv oversikt over skolefaste konstantledd (μ_{if}) i engelsk (målt som avvik fra gjennomsnitt)

	VG1200		VF1210	
	A1 Alle	A1 Sign	Y1 Alle	Y1 Sign
Gjennomsnitt	0,0035	-0,8477	0,0069	-0,1146
St.avvik	0,2005	0	0,8743	1,1646
Min	-0,8477		-1,5914	-1,5914
Max	0,5046		1,2624	1,2624
Ant skoler	92	1	74	28

Merk: Verdiene er vektet med antall elever. Grunnskoler med færre en 20 elever i løpet av den tidsperioden vi ser på er utelatt. Signifikant betyr statistisk signifikant til 5 prosents nivå.

Tabell 4.21. Deskriptiv oversikt over skolefaste konstantledd (μ_{jt}) i matematikk (målt som avvik fra global gjennomsnitt)

	VG1330		VG1331		VF1320	
	A1_0 Alle	A1_0 Sign	A1_1 Alle	A1_1 Sign	Y1 Alle	Y1 Sign
Gjennomsnitt	-0,0260		0,0243		-0,0575	-0,3469
St.avvik	0,8369		0,6392		0,6297	0,9301
Min	-1,6468		-1,078		-1,1548	-1,1548
Max	0,8320		0,7814		0,9742	0,9742
Ant skoler	46	0	25	0	83	12

Merk: Verdiene er vektet med antall elever. Grunnskoler med færre en 20 elever i løpet av den tidsperioden vi ser på er utelatt. Signifikant betyr statistisk signifikant til 5 prosents nivå.

Tabell 4.22. Deskriptiv oversikt over skolespesifikke helningskoeffisienter (β_{jt}) i norsk (målt som avvik fra global gjennomsnitt)

	VF1000		VF1010		VG4000	
	A1 Alle	A1 Sign	Y1 Alle	Y1 Sign	A3 Alle	A3 Sign
Gjennomsnitt	-0,0318	0,0173	-0,0437	0,2313	-0,0437	-0,2313
St.avvik	0,1831	0,4489	0,1944	0,3512	0,1944	0,3512
Min	-0,4949	-0,4949	-0,4668	0,4668	-0,4668	-0,4668
Max	0,4937	0,4937	0,5364	0,5364	0,5364	-0,5364
Ant skoler	41	4	47	4	66	4

Merk: Verdiene er vektet med antall elever. Grunnskoler med færre en 30 elever i løpet av den tidsperioden vi ser på er utelatt. Signifikant betyr statistisk signifikant til 5 prosents nivå.

Tabell 4.23. Deskriptiv oversikt over skolespesifikke helningskoeffisienter (β_{jt}) i engelsk (målt som avvik fra global gjennomsnitt)

	VG1200		VF1210	
	A1 Alle	A1 Sign	Y1 Alle	Y1 Sign
Gjennomsnitt	-0,0515	-0,1847	-0,0407	-0,1443
St.avvik	0,1278	0,1407	0,1517	0,2710
Min	-0,3436	-0,3436	-0,4731	-0,4731
Max	0,3558	0,3558	0,2927	0,2927
Ant obs	92	13	74	10

Merk: Verdiene er vektet med antall observasjoner. Grunnskoler med færre en 20 elever i løpet av den tidsperioden vi ser på er utelatt. Signifikant betyr statistisk signifikant til 5 prosents nivå.

Tabell 4.24. Deskriptiv oversikt over skolespesifikke helningskoeffisienter (β_{jt}) i matematikk (målt som avvik fra global gjennomsnitt)

	VG1330		VG1331		VF1320	
	A1_0 Alle	A1_0 Sign	A1_1 Alle	A1_1 Sign	Y1 Alle	Y1 Sign
Gjennomsnit	-0,0138	0,1536	-0,0397	-0,4717	0,0280	0,0190
St.avvik	0,1574	0,5190	0,1913	0	0,1476	0,3039
Min	-0,3151	-0,3151	-0,4876		-0,3688	-0,3688
Max	0,7193	0,7193	0,2636		0,3525	0,3525
Ant skoler ...	46	2	25	1	83	5

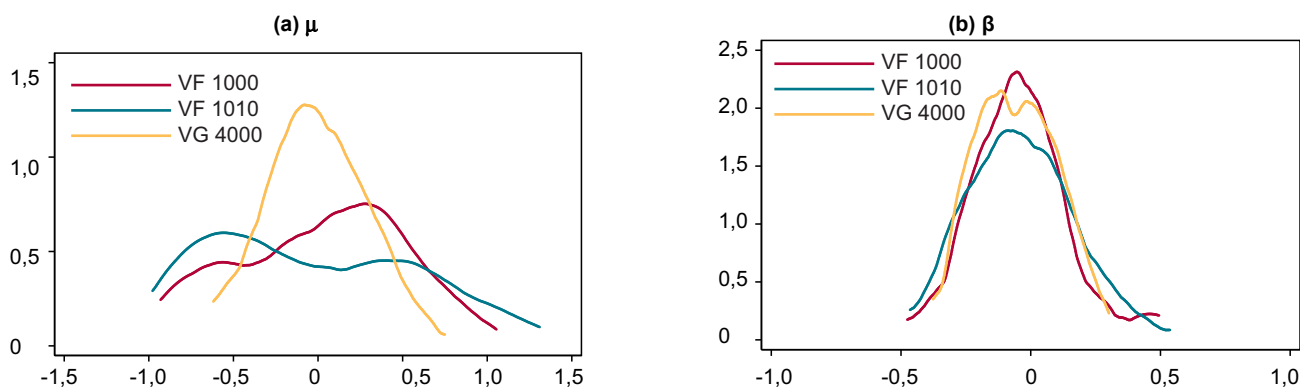
Merk: Verdiene er vektet med antall elever. Grunnskoler med færre en 20 elever i løpet av den tidsperioden vi ser på er utelatt. Signifikant betyr statistisk signifikant til 5 prosents nivå.

En deskriptiv oversikt over de estimerte β_{jt} 'ene (målt som avvik fra β estimert uten skolespesifikke koeffisienter) fra å estimere likning (2) er presentert i Tabell 4.22 (norsk), Tabell 4.23 (engelsk) og Tabell 4.24 (matematikk). Som før er gjennomsnittlig β nødvendigvis nær null. Standardavvikene er noe mindre sammenliknet med μ og svært få skoler har verdier på β som er statistisk signifikante.

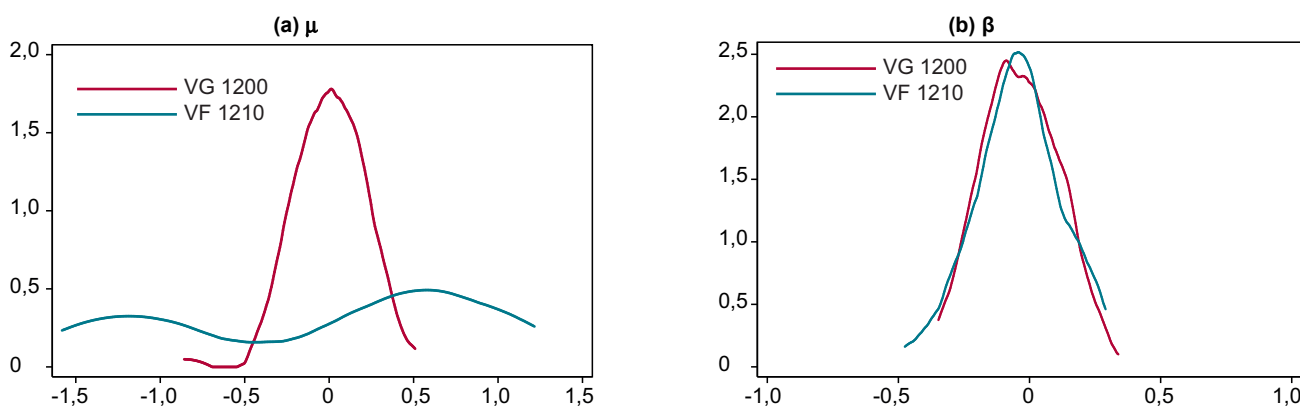
Figur 4.15-4.17 viser fordelingen av de skolespesifikke effektene i hhv. norsk, engelsk og matematikk etter hvilket fag som brukes som ekstern sensur. Figurene bekrefter inntrykket fra tabellene av at det er betydelig spredning i estimatene, men disse er altså for usikre til å kunne vektlegges.

Det lar seg ikke gjøre å lage en korrelasjonstabell som viser persistens i μ på tvers av fag. Grunnen til det er at når vi betinger på at antall observasjoner skal være større enn 20 i alle fag sitter vi kun igjen med 7 observasjoner. Dette gjelder også når vi ser på allmennfag og yrkesfag hver for seg.

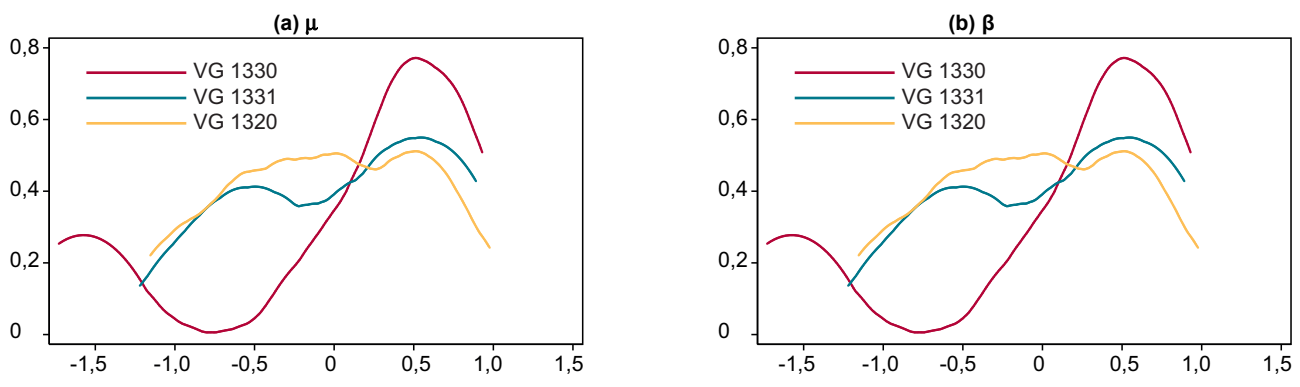
Figur 4.15. Fordeling av de estimerte skolespesifikke konstantleddene i norsk etter fag på videregående



Figur 4.16. Fordeling av de estimerte skolespesifikke konstantleddene i engelsk etter fag på videregående



Figur 4.17. Fordeling av de estimerte skolespesifikke konstantleddene i matematikk etter fag på videregående



5. Konklusjon

Når vi sammenligner standpunktkarakterer med skriftlige eksamenskarakterer, finner vi en betydelig variasjon mellom grunnskoler i nivået på karaktersettingen (de skolespesifikke konstantleddene, μ_{jf}), selv om vi kontrollerer for elevkjenntegn. Vi finner også at det er høy persistens på tvers av fag. Med det mener vi at skoler som gir høye/lave standpunktkarakterer (relativt til eksamenskarakterer) i for eksempel engelsk også gir høye/lave standpunktkarakterer i matematikk og norsk. Dette betyr videre at standpunktkarakterer satt av forskjellige lærere på forskjellige elever i forskjellige år gir tilsvarende resultater. Avvikene fra forventet standpunktkarakter, gitt eksamenskarakter, ser der ut til å drives av persistente skolekjenntegn, heller enn for eksempel enkeltlærere. Noen skoler - særlig skoler med få elever - ser ut til å overvurdere nivået til elevene på sin skole, mens andre - særlig store skoler - ser ut til å undervurdere. Vi finner også klare indikasjoner på at skoler med lave gjennomsnittlige eksamenskarakterer har en tendens til å overvurdere elevenes nivå ved fastsetting av standpunktkarakterer, mens skoler med høye gjennomsnittlige eksamenskarakterer heller tenderer mot å undervurdere prestasjonsnivået til elevene på sin skole.

Når det gjelder samvariasjon mellom standpunkt og eksamen (β_{jf}) er det både mindre variasjon på tvers av skoler og lavere persistens på tvers av fag. I sum er forskjeller i denne samvariasjonen mye mindre viktig for forskjeller i karakterpraksis enn forskjellen i nivået på intern karakter. Det er imidlertid tegn til at skoler med lave gjennomsnittskarakterer har en tendens til å overvurdere den relative forskjellen mellom svake og sterke elever, mens at skoler med høye gjennomsnittskarakterer tenderer til det motsatte. Men dette gjelder få skoler, ettersom det er få estimerte β 'er som er signifikante, og vi kan ikke konkludere med at det finnes variasjon av betydning her.

Når standpunktkarakter fra videregående benyttes som ekstern sensur blir resultatene for usikre og lite presise til at de kan gis noen meningsfull tolkning. Det er flere grunner til at vi kan vente mindre tydelige resultater fra videregående: For det første er vi henvist til å bruke en mindre del av datamaterialet. Grunnen til det er at vi trenger en tilstrekkelig omfordeling av elever, slik at karaktersetting på videregående blir uavhengig av karakterpraksis i grunnskolen. Videre fordeler forskjellige elevgrupper seg forskjellig på studieretninger og fag. Ettersom vi ikke kan anta at sammenhengene mellom resultater fra grunn- og videregående skole er de samme for alle fag og elevgrupper estimerer vi flere forskjellige sammenhenger, der hver er basert på et begrenset antall elever, som ikke er representative for alle elever ved grunnskolene de kommer fra.

Det at vi finner såpass markerte forskjeller i gjennomsnittlig standpunktkarakter for gitt eksamenskarakter kan tilsi at det finnes betydelige forskjeller i karakterpraksis. At skoler med gode eksamenskarakterer ser ut til å ha en streng karakterpraksis - og motsatt for skoler med svake eksamenskarakterer - kan tolkes som et uttrykk for relativ karaktersetting. Imidlertid, som diskutert i kapittel 2, må resultatene tolkes med varsomhet. Grunnen til det er at vi ikke kan utelukke at sammenhengene vi estimerer mellom standpunkt- og eksamenskarakter fanger opp en rekke forhold utover karakterpraksis, som for eksempel utelatte relevante kontrollvariable.

Referanser

Gravaas, B. C., T. Hægeland, L. J. Kirkebøen and K. Steffensen (2008): Skoleresultater 2007. En kartlegging av karakterer fra grunn- og videregående skoler i Norge. Notater 2008/24. Statistisk sentralbyrå

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

Hægeland, T., L. J. Kirkebøen and O. Raaum (2010): *Skolebidragsindikatorer for videregående skoler i Oslo* Rapport 2010/36. Statistisk sentralbyrå

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

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

Hægeland, T., O. Raaum and K. G. Salvanes (2004): Pupil achievement, school resources and family background. Discussion paper 397. Statistisk sentralbyrå

Steffensen og Ziade (2009): *Skoleresultater 2008. En kartlegging av karakterer fra grunnskoler og videregående skoler i Norge*. Rapport 2009/23. Statistisk sentralbyrå

Vedlegg A: En matematisk framstilling av sammenhengen mellom standpunkt og eksamenskarakter

I dette vedlegget gir vi en kortfattet matematisk formulering av betraktningene. En matematisk formulering er mer presis, og illustrerer tydeligere hvordan de betraktningene i kapittel 2 knytter an til analysene. Det er imidlertid lite nytt stoff i dette avsnittet.

Intern sensur (I) og ekstern sensur (E) antas å avhenge av hver sin ferdighet, x og y , der det er en sammenheng mellom de to ferdighetene:

$$(A1) \quad y = \rho x + \omega$$

Her måler ρ styrken i sammenhengen mellom y og x , mens ω oppsummer alt som utgjør en del av y , men som ikke inngår i x . Dette omfatter for eksempel evne til å formidle forståelse i faget muntlig. Denne sammenhengen kan variere mellom grupper av elever, for eksempel etter kjønn eller foreldrebakgrunn. I denne sammenheng antar vi at vi kan betinge tilstrekkelig på alle relevante forhold, og ser bort fra slike forskjeller.

Karakteren settes av lærer/sensor, på bakgrunn av dennes observasjon av ferdigheten. I virkeligheten kan I og E bare anta heltallsverdier fra en til seks. Dette kompliserer imidlertid analysen en del, så vi vil i denne omgang anta at dette ikke er viktig og betrakte både ferdighet og karakter som kontinuerlige variable. Observasjonen er beheftet med en viss feil. For ekstern sensur kun en tilfeldig feil, v . Denne er 0 i gjennomsnitt og antas å være ukorrelert med alle øvrige variable i modellen, men har en fordeling med en viss spredning (som vi for eksempel kan uttrykke ved standardavviket). Intern sensur er også utsatt for slik tilfeldig feil, v , men kan i tillegg inneholde systematiske avvik fra ferdighetsnivå. ψ_j er det systematiske avviket i nivå på skolenivå, når denne er høy svarer det til at alle elever, uavhengig av ferdighet, får en god karakter. ϕ_j uttrykker i hvor stor grad skolen skiller mellom gode og dårlige elever:

$$(A2) \quad \begin{aligned} E &= x + v \\ I &= \psi_j + \phi_j y + v \end{aligned}$$

ψ_j og ϕ_j utgjør til sammen skolens karakterpraksis, og er det vi er interessert i å studere. Når $\psi_j=0$ og $\phi_j=1$ er det ingen form for systematiske avvik, dette kan vi tenke på som "riktig" karakterpraksis.

Ut fra disse forutsetningene kan vi uttrykke sammenhengen mellom I og E :

$$(A3) \quad I = \psi_j + \phi_j(\rho x + \omega) + v = \psi_j + \phi_j(\rho(E - v) + \omega) + v$$

Når vi estimerer en ligning $I = \alpha + \beta_j E + \mu_j + \varepsilon$ får vi følgende koeffisienter:

$$(A4) \quad p \lim \hat{\beta}_j = \frac{\text{cov}(E_j, I_j | \hat{\mu}_j)}{\text{var}(E_j)}$$

$$(A5) \quad p \lim \hat{\mu}_j = \bar{I}_j - \hat{\beta}_j \bar{E}_j$$

De estimerte koeffisientene avhenger av den empiriske samvariasjonen mellom E og I . Hvis vi har tilstrekkelig mange observasjoner, det vil vi ikke nødvendigvis ha på hver skole, vil denne svare til en form for "sann" samvariasjon. Denne samvariasjonen fanger imidlertid opp flere forhold, som vi kan se fra (A2) og (A3).

Det at ferdighetene som ligger til grunn for E og I er forskjellige bidrar til en lavere samvariasjon. Dette er felles for alle skoler, såfremt sammenhengen mellom de to ferdighetene er tilsvarende for alle skoler. Hvis det er forskjeller mellom grupper av elever, og disse fordeler seg forskjellig på skoler, trenger ikke dette å være tilfellet.

Videre vil målefeilen i E bidra til ytterligere redusert samvariasjon mellom E og I. Dette er intuitivt rimelig, når E er upresis gir den et dårligere mål på underliggende ferdighet, og dermed en svakere sammenheng med observert standpunktkarakter. Størrelsen på denne effekten avhenger av spredningen til målefeilen. Såfremt denne er lik for alle skoler vil alle skoler skaleres like mye, men med et begrenset antall elever per skole kan det være forskjeller.

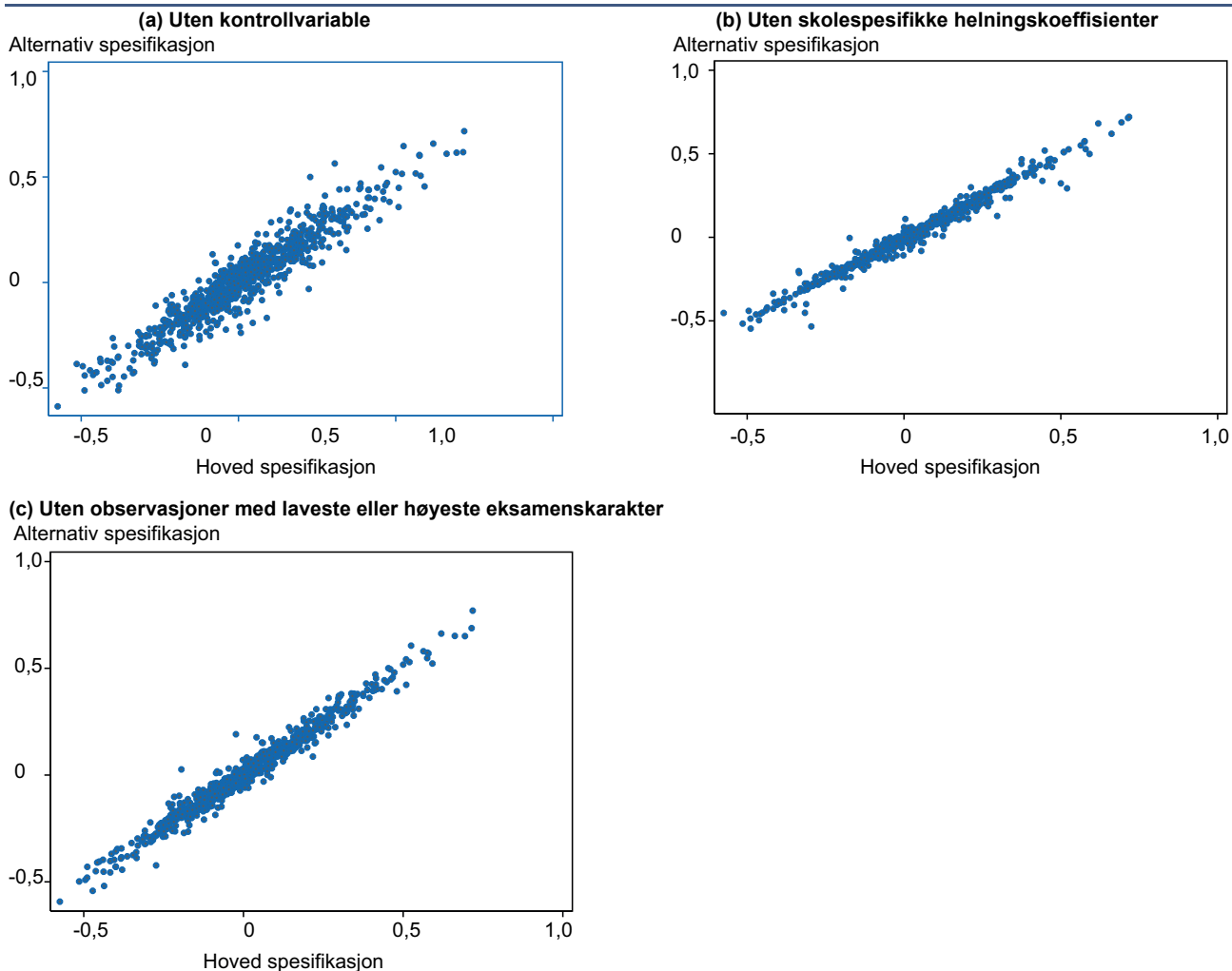
I modellen over er det ikke spesifisert at I avhenger av noe utover ferdighet og karakterpraksis, som for eksempel innsats. I den grad dette er tilfelle, og vi ikke klarer å betinge på disse øvrige forholdene, vil det kunne bidra til ytterligere skjevheter i estimatene, der vi tillegger betydningen av disse forholdene, som vi ikke observerer, til E, som vi observerer.

Til slutt, karakterpraksis, uttrykt ved ψ_j og φ_j kan tenkes å samvariere med E eller andre variable vi observerer og kontrollerer for. I så fall kan vi tillegge noe av karakterpraksisen til disse variablene.

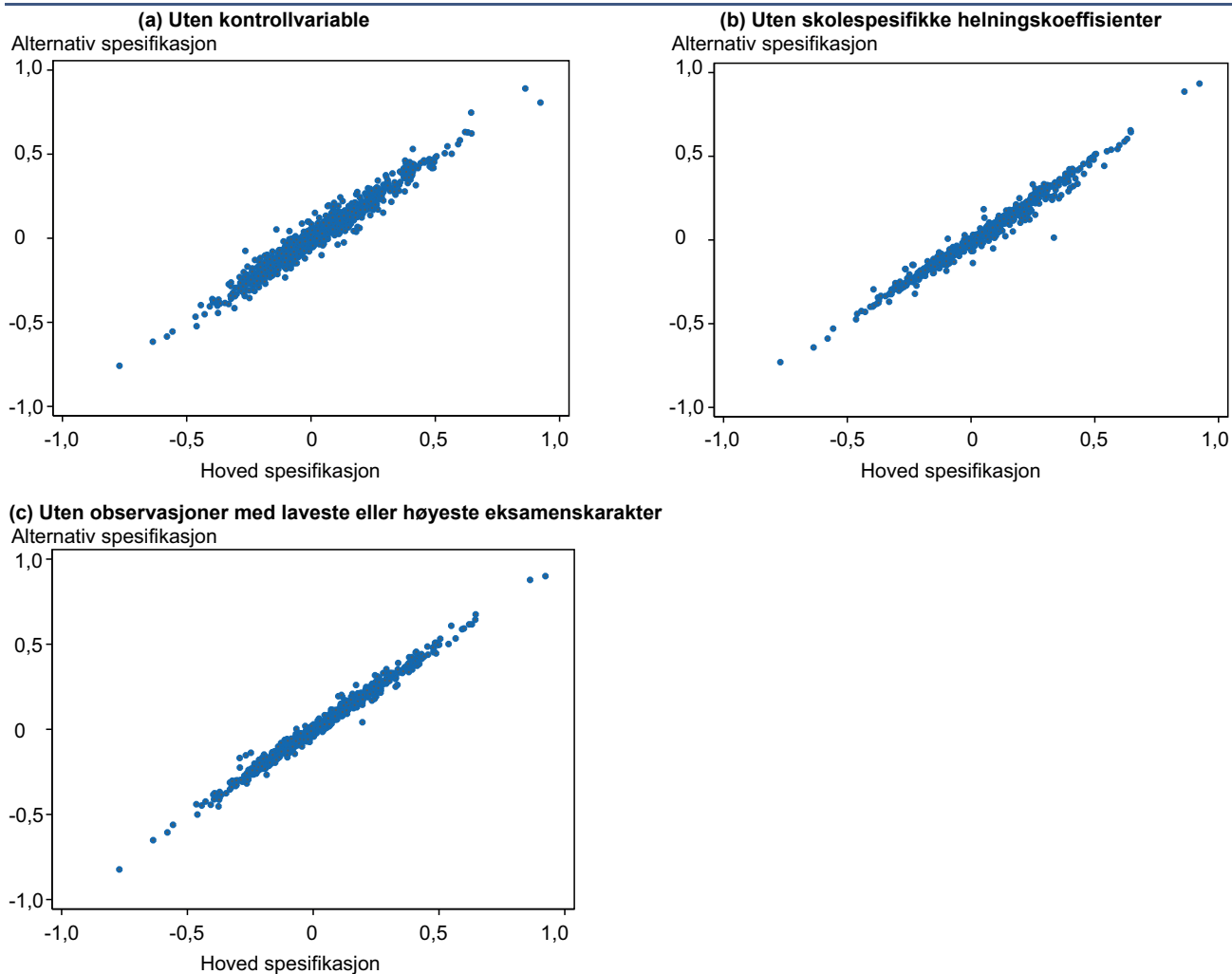
I sum gir dette en rekke utfordringer, som gjør at vi ikke kan være sikre på at de estimerte helningskoeffisienten ikke uttrykker φ_j , og de skolespesifikke konstantleddene heller ikke uttrykker ψ_j .

Vedlegg B: Robusthet: Samvariasjon mellom forskjellige estimater

Figur B.1. Sammenligning av de skolespesifikke konstantleddene estimert ved hovedspesifikasjon med de skolespesifikke konstantleddene med alternativspesifikasjoner i faget norsk



Figur B.2. Sammenligning av de skolespesifikke konstantleddene estimert ved hovedspesifikasjon med de skolespesifikke konstantleddene med alternativspesifikasjoner i faget engelsk



Figurregister

2.1. Sammenhengen mellom intern og ekstern sensur	10
2.2. Tolkning av resultatene	15
4.1. Fordeling av de estimerte skolespesifikke konstantleddene (μ_{jt}) og helningskoeffisientene (β_{jt}) i matematikk, norsk og engelsk	23
4.2. Fordeling av de estimerte skolespesifikke konstantleddene (μ_{jt}) i matematikk etter skolestørrelse	24
4.3. Fordeling av de estimerte skolespesifikke konstantleddene (μ_{jt}) i norsk etter skolestørrelse	24
4.4. Fordeling av de estimerte skolespesifikke konstantleddene (μ_{jt}) i engelsk etter skolestørrelse	25
4.5. Fordeling av de estimerte skolespesifikke helningskoeffisientene (β_{jt}) i matematikk etter skolestørrelse	25
4.6. Fordeling av de estimerte skolespesifikke helningskoeffisientene (β_{jt}) i norsk etter skolestørrelse	25
4.7. Fordeling av de estimerte skolespesifikke helningskoeffisientene (β_{jt}) i engelsk etter skolestørrelse	25
4.8. Fordeling av de estimerte skolespesifikke konstantleddene (μ_{jt}) i matematikk etter skolens gjennomsnittlig eksamensprestasjon (i kvintiler)	26
4.9. Fordeling av de estimerte skolespesifikke konstantleddene (μ_{jt}) i norsk etter skolens gjennomsnittlig eksamensprestasjon (i kvintiler)	26
4.10. Fordeling av de estimerte skolespesifikke konstantleddene (μ_{jt}) i engelsk etter skolens gjennomsnittlig eksamensprestasjon (i kvintiler)	27
4.11. Fordeling av de estimerte skolespesifikke helningskoeffisientene (β_{jt}) i matematikk etter skolens gjennomsnittlig eksamensprestasjon (i kvintiler)	27
4.12. Fordeling av de estimerte skolespesifikke helningskoeffisientene (β_{jt}) i norsk etter skolens gjennomsnittlig eksamensprestasjon (i kvintiler)	27
4.13. Fordeling av de estimerte skolespesifikke helningskoeffisientene (β_{jt}) i engelsk etter skolens gjennomsnittlig eksamensprestasjon (i kvintiler)	27
4.14. Sammenligning av de skolespesifikke konstantleddene estimert ved hovedspesifikasjon med de skolespesifikke konstantleddene med alternativspesifikasjoner i faget matematikk	30
4.15. Fordeling av de estimerte skolespesifikke konstantleddene i norsk etter fag på videregående	35
4.16. Fordeling av de estimerte skolespesifikke konstantleddene i engelsk etter fag på videregående	35
4.17. Fordeling av de estimerte skolespesifikke konstantleddene i matematikk etter fag på videregående	35
B.1. Sammenligning av de skolespesifikke konstantleddene estimert ved hovedspesifikasjon med de skolespesifikke konstantleddene med alternativspesifikasjoner i faget norsk	40
B.2. Sammenligning av de skolespesifikke konstantleddene estimert ved hovedspesifikasjon med de skolespesifikke konstantleddene med alternativspesifikasjoner i faget engelsk	41

Tabellregister

3.1. Fag i videregående, men korresponderende fag fra grunnskolen	17
3.2. Elever i grunnskolen: Deskriptiv statistikk etter skriftlig eksamenskarakter	18
3.3. Elever i grunnskole og videregående: Deskriptiv statistikk etter fag fra videregående, allmennfag	19
3.4. Elever i grunnskole og videregående: Deskriptiv statistikk etter fag fra videregående, yrkesfag	19
4.1. Resultater fra å estimere ligning (1). Avhengig variabel er standpunktkarakterene på grunnskolen i henholdsvis engelsk, matematikk og norsk (hovedmål)	20
4.2. Deskriptiv oversikt over skolefaste konstantledd, μ_{jf} (målt som avvik fra globalt gjennomsnitt)	21
4.3. Deskriptiv oversikt over helningskoeffisienten, β_{jf} (målt som avvik fra globalt gjennomsnitt)	22
4.4. Korrelasjon mellom β_{jf} og μ_{jf} i matematikk	22
4.5. Korrelasjon mellom β_{jf} og μ_{jf} i engelsk	22
4.6. Korrelasjon mellom β_{jf} og μ_{jf} i norsk	22
4.7. Korrelasjon mellom fag i skolefaste effekter, μ_{jf} (alle)	28
4.8. Korrelasjoner mellom fag i skolefaste effekter, μ_{jf} (statistisk signifikant til 5 % nivå) ...	28
4.9. Korrelasjoner mellom fag i skolefaste effekter, μ_{jf} (statistisk signifikant til 10 % nivå) .	28
4.10. Korrelasjon mellom fag i β_{jf} (alle)	29
4.11. Korrelasjon mellom fag i β_{jf} (signifikante til 5 prosents nivå)	29
4.12. Korrelasjon mellom fag i β_{jf} (signifikante til 10 prosents nivå)	29
4.13. Robusthet: beskrivelse av fordeling av de skolespesifikke konstantleddene ved estimering med alternative spesifikasjoner i faget matematikk	29
4.14. Robusthet: beskrivelse av fordeling av de skolespesifikke konstantleddene ved estimering med alternative spesifikasjoner i faget norsk	31
4.15. Robusthet: beskrivelse av fordeling av de skolespesifikke konstantleddene ved estimering med alternative spesifikasjoner i faget engelsk	31
4.16. Resultater fra å estimere ligning (1). Avhengig variabel er standpunktkarakteren i norsk hovedmål fra grunnskolen	32
4.17. Resultater fra å estimere ligning (1). Avhengig variabel er standpunktkarakter i engelsk skriftlig fra grunnskolen	32
4.18. Resultater fra å estimere ligning (1). Avhengig variabel er standpunktkarakteren i matematikk fra grunnskolen	33
4.19. Deskriptiv oversikt over skolefaste konstantledd (μ_{jf}) i norsk (målt som avvik fra gjennomsnitt)	33
4.20. Deskriptiv oversikt over skolefaste konstantledd (μ_{jf}) i engelsk (målt som avvik fra gjennomsnitt)	33
4.21. Deskriptiv oversikt over skolefaste konstantledd (μ_{jf}) i matematikk (målt som avvik fra global gjennomsnitt)	34
4.22. Deskriptiv oversikt over skolespesifikke helningskoeffisienter (β_{jf}) i norsk (målt som avvik fra global gjennomsnitt)	34
4.23. Deskriptiv oversikt over skolespesifikke helningskoeffisienter (β_{jf}) i engelsk (målt som avvik fra global gjennomsnitt)	34
4.24. Deskriptiv oversikt over skolespesifikke helningskoeffisienter (β_{jf}) i matematikk (målt som avvik fra global gjennomsnitt)	34