



# METODERAPPORT

## SKUP 2009



# De skjulte karakterene

John Bones og Frank Ertesvåg

## Et prosjekt basert på dataverktøy og innsynsbegjæring

Regjeringen vil skjule norske skolekarakterer for offentlig innsyn.

Opposisjonen vil skjule skattelister for offentlig innsyn.

VG koblet informasjon fra begge de omstridte registrene, kombinerte dette med informasjon fra Kostra og fikk gjort statistikkfaglig analyse av funnene. Dermed kunne VG avdekke de skjulte sammenhengene.

1. **Journalister som har gjort jobben:**

John Bones og Frank Ertesvåg, VG

2. **Tittel:**

De skjulte karakterene.

3. **Publisering:**

29. november, 30. november, 1. desember, 11. desember, 16. desember, 17. desember og 29. desember i VG, samt VG Nett 29. november.

4. **Redaksjon og redaksjonsadresse:**

Verdens Gang, Postboks 1185 Sentrum, 0107 Oslo, telefon 22 000 000.

5. **Journalistenes adresse og telefonnummer:**

John Bones  
Nils Lauritssøns vei 33  
0854 Oslo  
Tlf. 9133 5101

Frank Ertesvåg  
Ovenstadlia 118

3420 Lierskogen

Tlf. 9160 0215

## Redegjørelse for arbeidet:

### **a) Når og hvordan kom arbeidet i gang, hva var ideen som startet det hele.**

Arbeidet startet 22. oktober, da Utdanningsdirektoratet la ut karakterene for grunnskolen på denne nettsiden: [www.udir.no/skoleporten](http://www.udir.no/skoleporten).

Som tidligere år etter regjeringsskiftet i 2005, måtte hver enkelt skole, fylke eller kommune søkes opp manuelt. Etter at den rødgrønne regjeringen overtok styringen av Kunnskapsdepartementet, ble det tatt grep for å hindre innsyn og sammenstillinger av karakterer og resultat på nasjonale prøver. Motivet var å ikke tilrettelegge for rangering av skoler og "uthenging" av skoler med lave resultater i mediene.

Sammenstilte data var ikke tilgjengelig, hverken i databaseformat eller som nedlastbare filer, for eksempel i Excel-format.

Ideen som startet det hele var denne:

Dette må vi greie å få ut. Karakterer har offentlig interesse.

Avgangskarakterene er de viktigste enkeltdata for å vise elevenes skoleprestasjoner.

Metodebruken skal frem til reportasjeleggingen være databasert, og reportasjene skal ta utgangspunkt i de funnene de databaserte undersøkelsene har frembrakt. Vi ønsket også å presentere noe mer enn rene karakter-rangeringer, som i seg selv forteller lite om læringskvalitet.

## **b) Hva var den sentrale problemstillingen ved starten av prosjektet?**

Da prosjektet startet, var vi i første omgang kun orientert mot å få lastet ned karakterene. Vi ønsket også å sammenligne karakterene for 2008/09 med 2005/06. Hvilken utvikling har skjedd? Er det noe mønster i utviklingen som vi kan avdekke?

Våre ambisjoner stanset i utgangspunktet der, men da vi først var kommet i gang, fikk vi stadig nye ideer til å utvikle stoffet videre. Det ble et prosjekt i stedet for en enkeltstående og grei nyhetssak.

## **c) Redegjør kort for hva som er genuint nytt i saken (sett mot evt. tidligere offentliggjort informasjon).**

Det er to nyheter som er genuint nye.

\* Vi fant ut at det ikke er noen sammenheng mellom penger som er bevilget til undervisning pr. elev i hver kommune og karakterene som oppnås.

\* Vi fant ut at det er et stort sprik mellom standpunktkarakterer og eksamenskarakterer. Ved de fleste skoler er standpunktkarakterene høyest, ved andre skoler eksamensresultatene. Elever som kommer fra skoler med «snille» standpunktkarakterer har et konkurransefortrinn i konkurransen om plass ved de videregående skolene. Standpunktkarakterene utgjør ca. 80 prosent av poengsummen som ligger til grunn for videre søkning og opptak.

## d) Kort beskrivelse av organiseringen av arbeidet, metodebruk, kildebruk, problem underveis mv. Redegjør spesielt for kildebruk og kildekritikk.

For å laste ned informasjonen, var vi i første omgang avhengig av å bruke webscraping. Derfor søkte vi bistand hos Nils Mulvad i Danmark, ettersom han er kjent for å være en dyktig bruker av «robotprogrammet» OpenKapow.

***Metoden for innstilling av roboten fremgår av vedlegget "Robot-innstillingen"***

Han sendte oss den nedlastede informasjonen i Access-format. Vi opprettet vår egen database og gjorde spørringer i denne for å kunne bryte ned informasjonen på de enkelte fagene. Vi tok fag etter fag, og til slutt regnet vi gjennomsnitt. For at gjennomsnittet skulle være riktig, måtte vi vekte for antall elever som hadde fått karakter.

***I databasen "Grunnskolen standpunkt etter webscraping fremgår metodene som er brukt. Koden "G" står for Grunnskole, koden GK står for Grunnskole kommune. På denne måten skilte vi mellom skoler og kommuner.***

***Metoden for å vekte antall elever fremgår av formelen i kolonne "V" i Excel-filens ark "Utregn karakter 8 fag".***

Det var da dette var klart, og vi for så vidt hadde gjort det vi trengte for å komme i mål med den opprinnelige ambisjonen, at vi fikk lyst til å utvikle dette til å bli et større prosjekt. Kunne vi greie å finne noen sammenhenger mellom karakterene og andre relevante parametre fra kommunene?

Da bestemte vi oss for å laste inn informasjon fra fire andre datakilder:

\* Skattelistedene. Vi lastet skattelistedene inn i Access og beregnet gjennomsnittlig inntekt i alle kommuner i hele Norge. Denne spørringen importerte vi til vår database med karakterer. Her og nå legger vi ikke ved skattedatabasen, kun resultatet av spørringen som gjelder gjennomsnittslønn i de enkelte kommuner. Spørringen ble importert som egen tabell og har navnet "**snitt inntekt komm 08**"

\* KOSTRA. Vi lastet ned to tabeller. Den ene hadde informasjon om undervisningsutgiftene til grunnskolen i alle kommuner i landet, den andre hadde informasjon om antall elever i grunnskolen i alle kommuner. Da vi

# De skjulte karakterene

koblet disse to tabellene, kunne vi beregne hvor mye hver kommune brukte til undervisning pr. elev. Vi søkte råd hos SSB om hvilke tabeller som var mest relevant å bruke, ettersom vi ville unngå å blande inn feilkilder som skoleskyss, vedlikehold av bygningsmasse etc.

## **Metoden fremgår av de første fem arkene i Excel-filen.**

\* Ordførerparti. Tidligere har vi laget en fil med navn, kjønn og partitilknytning til ordførerne i alle landets kommuner. Denne ble importert til karakterdatabasen som egen tabell under navnet "**ordførerparti**".

\* Størrelse på avgangskullet. Med utgangspunkt i basen med karakterer kunne vi regne ut hvor stort avgangskullet var ved de enkelte skoler. Vi forhørte oss med Utdanningsdirektoratet om å få tilgang til en fil som inneholdt antall elever ved alle landets skoler. Det fikk vi, men filen hadde den svakheten at den ikke var utstyrt med den enkelte skoles ID-nummer. Dermed hadde den så mange feilkilder at den var ubrukelig i et prosjekt som baserte seg på databasert journalistikk.

Da vi til slutt hadde all informasjon i en og samme database, kunne vi koble data opp mot hverandre. Nøkkelen for koblingene var dels kommunenummer, dels skolenes ID-nummer.

Vi velger ikke å redegjøre for hver enkelt spørring i databasen, men grunnprinsippet har vært dette: Først spør vi etter hvert enkelt fag, deretter lager vi en ny spørring for å samle alle fag, og til slutt kombinerer dette med de andre parametrene. Underveis har vi brukt Excel som lagringssted og hjelpeverktøy.

## **Metoden for å samkjøre data finnes i spørringen "G kom knr parti utgift fylke karakter".**

### STATISTIKKFAGLIG BISTAND

Nå hadde vi de faktaopplysningene vi trengte, og den totale mengden av datainformasjon var så massiv at vi måtte få en statistikkfaglig gjennomgang av det vi hadde. Vi visste at forskningssjef Geert Laier Christensen i den danske tenketanken Cepos hadde utført en slik analyse for danske skoler, og derfor tok vi kontakt med ham og spurte om han kunne bruke våre data til å foreta en tilsvarende regresjonsanalyse av våre data. Det kunne han.

Det vi ba ham gjøre, var å finne ut om det var noen sammenheng mellom karakterer, som avhengig variabel, og de uavhengige variablene gjennomsnittsinntekt i kommunene, utgifter pr. elev i kommunene, ordførerfarge, ordførerkjønn og antall elever i avgangskullet.

Christensen fikk disse to Excel-arkene oversendt: "**All info kommune**" og "**8 fag karakt vs størrelse**", og hans analyse er altså basert på den informasjonen som ligger der.

# De skjulte karakterene

Christensens metode og avrapportering fremgår av vedlegg ”*Regresjonsanalysen*”, men her nevner vi de viktigste konklusjonene.

- \* Det er ingen signifikant forskjell mellom utgifter pr. elev og karakterer.
- \* Det er en signifikant forskjell mellom gjennomsnittsinntekt i kommunen og karakterer.
- \* Det er en signifikant forskjell mellom antall elever i avgangskullet og karakterene.
- \* Det er ingen signifikant forskjell mellom ordførerparti og karakterer, eller mellom ordførerkjønn og karakterer. Når det gjelder politisk farge, hadde vi delt i tre grupper: Røde, som vi definerte som de tre rødgrønne regjeringspartiene pluss Rødt, de blå, som vi definerte som de H, Frp, KrF og Venstre, og andre, som er de lokale listene uten klar politisk farge. Vi ser argumentene mot denne inndelingen, men vi valgte å skille mellom regjeringspartiene og de andre, for å se om det var noen forskjell.

Det mest oppsiktsvekkende i vårt materiale var at det ikke var noen sammenheng mellom ”penger inn og karakterer ut”. Da vi skulle reportasjelegge stoffet, lette vi derfor etter to skoler i hver sin ende av skalaen: En kommune med stort pengebruk og lave karakterer, og en skole med lavt pengebruk og høye karakterer.

Vi endte opp med tre case-reportasjer fra tre ulike ungdomsskoler i tre ulike kommuner i Norge.

\* I Oppegård fant vi en skole med liten ressursbruk pr. elev og høye karakterer.

\* I Søndre Land fant vi en skole med høy ressursbruk pr. elev og lave karakterer.

\* I Vang fant vi en skole med høy ressursbruk pr. elev og gode karakterer. Vi oppsøkte disse tre skolene, intervjuet skolesjef, rektor, lærere og elever. Hele tiden med mest vekt på våre funn - karakterer, sosial bakgrunn og ressursbruk for å levendegjøre og teste våre funn i form av reportasjer.

Reportasjene ble publisert over seks sider pluss hovedoppslag i VG 29. november. VG Nett hadde utviklet en interaktiv grafikk, der leserne, basert på våre data, kunne søke opp den informasjonen de ønsket.

## INNSYNSBEGJÆRING FOR Å FINNE AVVIK MELLOM STANDPUNKT OG EKSAMEN

Etter at dette var publisert, fikk vi lyst til å utfordre myndighetenes mangel på åpenhet når det gjelder karakterer. Etersom data ligger på [www.skoleporten.no](http://www.skoleporten.no), er de offentlige, og da plikter også myndighetene, i kraft av den nye paragraf 9 i den reviderte Offentlegglova, å gi oss innsyn i hele databasen: ”Rett til å krevje innsyn i ei samanstilling frå databasa”. Uten



# De skjulte karakterene

innsyn i databasen var vi avhengig av å laste ned informasjonen manuelt, noe vi vil anslå til et sted mellom 30 000 og 40 000 søk, eller vi kunne brukt webscraping enda en gang.

Etter seks dager fikk vi innsyn i det vi ba om for grunnskolen vedkommende, men Utdanningsdirektoratet var i tvil om de kunne gi oss det vi ba om for de videregående skolene. Vi kunne hente en minnebrikke med karakterdata for grunnskolen hos Utdanningsdirektoratet.

Korrespondansen mellom VG og Utdanningsdirektoratet fremgår av vedlegget "*Innsynsbegjæringen*".

Vi behandlet data, hovedsakelig i Access, men også i Excel, og vårt mål var å finne ut eventuelle avvik mellom standpunktkarakterer, som vi hadde fra før, og eksamenskarakter, som vi hadde fått innsyn i etter vår begjæring.

Det fant vi!

Avvikene var store på enkeltfag, men også på gjennomsnitt. På enkeltfag var det sterkeste avviket mellom standpunkt og eksamen innen norsk hovedmål 1,6 karakterenhet, innen matematikk 1,3 og innen engelsk skriftlig 1,4. Største avvik på snitt var 0,72.

Det var også betydelige avvik den andre veien, altså at eksamenskarakteren var bedre enn standpunktkarakteren. På snitt var avviket 0,4, innen norsk hovedmål 0,7, innen matematikk 1,2 og innen engelsk skriftlig 1,1.

***Metodene for vår jobbing i Access fremgår av databasen som heter "Eksamen grunnskolen". For å kunne sammenligne med standpunktkarakterene, importerte vi standpunktkarakterene som egne tabeller. Koden "X" betyr eksamen, koden "s" (i spørringene) og "G" (i tabellene) betyr standpunktkarakter.***

***Metoden for gjennomsnittsberegninger fremgår av Excel-dokumentet "Metode snitt vs. Eksamen", og forklaring på utregningene ligger i ark "3 fag utregn", rad 1, samt i kolonne W.***

## VIDEREGÅENDE SKOLER

Da arbeidet med grunnskolekarakterene var ferdig, fikk vi utlevert eksamenskarakterene uten noen videre korrespondanse mellom VG og Utdanningsdirektoratet. Mens vi hadde problematisert grunnskolekarakterene i forhold til inntekt, penger pr. elev, størrelse på avgangskull, ordførerfarge og ordførerparti, laget vi ett registrerende og ett problematiserende oppslag om de videregående skolene.

Det registrerende oppslaget baserte seg på webscraping og var vinklet på den skolen som hadde den beste gjennomsnittskarakteren.

Hovedvinkelen på den problematiserende saken var å se om det var noen forskjeller mellom de private og offentlige skolene, og ellers laget vi noen enkle rangeringer. Tilgjengelige data i denne saken var basert på

# De skjulte karakterene

innsynsbegjæringen.

Vi fant at de private videregående skolene så vidt har bedre karakterer enn de offentlige videregående skolene både innen standpunkt og eksamen. Vi laget en reportasje i tilknytning til dette fra Akademiet videregående skole i Oslo.

***Metoden fremgår av databasene "Eksamen videregående off vs privat" og "Eksamen videregående", samt Excel-filen "Metode videregående offentlig og privat". Data har vært frem og tilbake mellom Access og Excel, uten at vi går nærmere inn på detaljene her. Koden "P" står for privat, koden "O" for offentlig.***

Helt til slutt laget vi en kuriøs liten sak, som et punktum for satsingen: Frisørene som utdannes i Strømmen har vesentlig bedre karakterer enn frisørene som utdannes i Kirkenes: 4,8 mot 2,9. Vi har ikke undersøkt i hvilken grad dette gjenspeiles i frisyrene på Romerike og i Finnmark.

## Spesielle erfaringer du vil nevne.

Publisering på to plattformer den første dagen skapte litt usikkerhet blant dem som grundig hadde lest begge versjonene. På VG Nett brukte vi de tre karakterene norsk hovedmål, engelsk skriftlig og matematikk fordi dette ga oss mulighet til å sammenligne med resultatene tre år tidligere.

I papirutgaven gikk vi bredere ut og tok med, i tillegg til disse tre fagene, også engelsk muntlig, naturfag, samfunnsfag, norsk muntlig og norsk sidemål i utregningene av gjennomsnitt. De to plattformene opererer derfor med litt forskjellig snittkarakterer.

Oslo, 7. januar 2010

John Bones

Frank Ertesvåg

## Vedlagte filer:

Vedlegg 1: Regresjonsanalyse

Vedlegg 2: Robot-innstilling

Vedlegg 3: Innsynsbegjæring

## Vedlagt på CD:

### Access database:

Eksamen grunnskolen, inneholder 7 tabeller og 16 spørringer.

Eksamen videregående skoler, inneholder 8 tabeller og 24 spørringer.

Grunnskolen standpunkt etter webscraping, inneholder 21 tabeller og 31 spørringer.

### Excel-filer:

Metode grunnskolen, inneholder 16 ark.

Metode snitt vs eksamen, inneholder 13 ark.

Metode videregående offentlig og privat, inneholder 6 ark.

### Word-fil:

link til VG Nett

# De skjulte karakterene VEDLEGG

## Vedlegg 1: Regresjonsanalysen

### Formålet

Analysens formål er at undersøge sammenhengen mellom en række faktorer og elevernes kundskabsniveau målt ved hjelp af et karaktergennemsnit.

### Variabler i modellen

I den fulde model indgår følgende faktorer:

#### Den afhængige variabel.

De gennemsnitlige karakterer (Snitt8fag)

Her kunne også have været anvendt snittet på 3 fag i skolen, men det gør ikke nogen reel forskel, da der er en meget høj korrelation mellem disse to mål for karaktergennemsnit (Persons r på 0,944).

#### Uafhængige (forklarende) variabler

##### **1.a Gennemsnitlig indtægt i kommunen (Snittinntekt\_kom)**

1.b DUMMY\_INDT

1.c interaktion DUMMY\_INDT\* Snittinntekt\_kom

##### **2. Udgifter pr elev i kommunen (Utgiftprelev)**

##### **3. Ordførerparti opdelt i de 3 kategorier rød, blå og andre (part\_kat)**

##### **4. Kønnen på ordføreren (Kjønnordfører)**

##### **5.a Antal elever i skolen med karakter (Antallelevermedkarakter)**

5.b elev\_dum3k

5.c interaktion elev\_dum3k\* Antallelevermedkarakter

# De skjulte karakterene VEDLEGG

Når der anvendes flere variabler for indtægt og antal på skolen er det fordi, der er tale om en kurvelineær sammenhæng. En visuel inspektion viser, at der er tale om u-formede sammenhænge, hvorfor der medtages dummy variabler, så modellen tillader forskellige sammenhænge mellem den uafhængige og den afhængige for forskellige niveauer af den afhængige variabel.

## Metode:

Sammenhængen mellem de uafhængige variabler og den afhængige variabel (gennemsnitskarakter) modelleres ved hjælp af en multipel regressionsmodel. Analysen er gennemført i statistik programmet STATA. Da analyseenheden er skole, men da der også anvendes data på kommuneniveau, bliver optionen robust cluster anvendt, for at tage højde for den inflation i standardfejlen, der ellers ville være tale om. Formuleret anderledes: i udgangspunktet ville modellen forvente at hver analyseenhed (dvs. hver skole) er uafhængige observationer, men det har vi ikke, da skolerne er samlet i kommuner. Derfor skal der tages højde for dette.

I analysen anvendes 397 kommuner og 864 skoler.

## Resultater:

I tabellen nedenfor er vist resultaterne af analysen. Der vises dels en fuld model og dels en reduceret model, som kun indeholder de variabler som har signifikante sammenhænge med den afhængige variabel karaktergennemsnit på 8 fag. Den reducerede model indeholder udelukkende gennemsnitlig indkomst i kommunen og antal elever på skolen.

# De skjulte karakterene VEDLEGG

Tabel 1: resultater for regressionsmodel med 8 fags karaktergennemsnit på skole

Indtægt	Dummy under 210.000 kr gns indtægt	0,8350715	0.000	0,893653	0.000
	Gennemsnitlig indtægt i kommunen	2,55e-06	0.000	2.45e-06	0.000
	Interaktion: dummy intg og indtægt	-4.03e-06	0.000	-4.27e-06	0.000
Udgift pr. elev	Udgift per elev	1,60e-06	0.112		
Parti	Højreorienterede (ref)	0			
	Venstreorienterede	0,0129235	0.515		
	Andre	-0,0646829	0.238		
ordfører køn	Mand (ref)	0			
	Kvinde	-0,0310719	0.089		
Antal elever	Under 18 (ref)	0		0	
	18 - 100	-0,1655955	0.199	-0,10783	0.000
	100 -	-0,3299177	0.042	-0,05884	0.089
	Antal elever med karakterer	-0,006012	0.518		
	Interaktion dummy 18-100 og antal elever	0,0058617	0.527		
	Interaktion dummy 100 - og antal elever	0,0076207	0.415		
	Konstant	3,317	0.000	3,3533	0.000
	<b>R2</b>	<b>0,1098</b>		<b>0.0922</b>	

Man kan ud fra den reducerede model konkludere følgende:

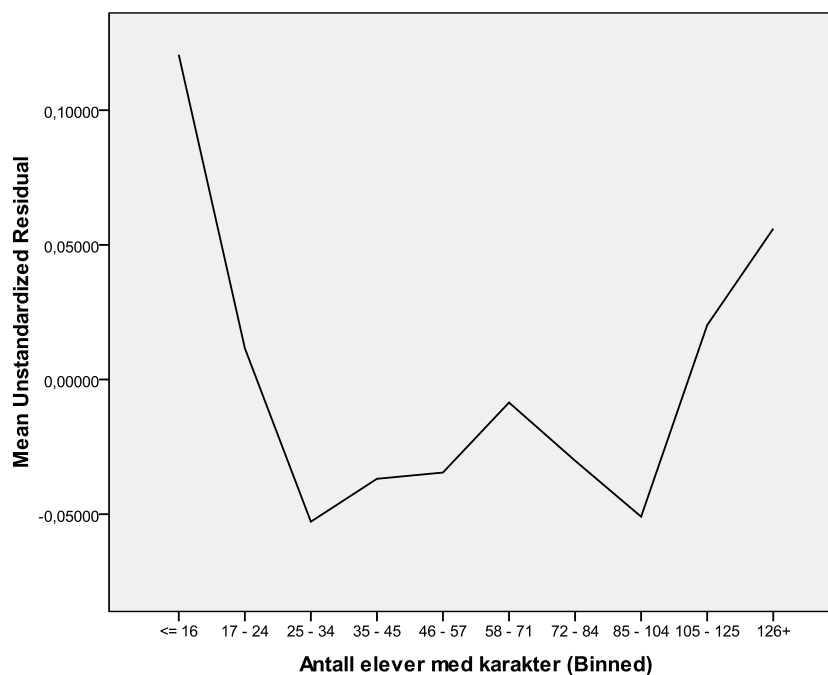
1. Der er en statistisk signifikant sammenhæng mellem gennemsnitlig indkomst i kommunen og karaktererne.
2. Der er en statistisk signifikant sammenhæng mellem antal elever med karakterer på skolen og karaktergennemsnittet.
3. Der er *ikke* nogen signifikant sammenhæng mellem ordførerparti og karaktergennemsnittet.
4. Der er *ikke* nogen signifikant sammenhæng mellem udgifter pr. elev og karaktergennemsnittet.
5. Der er *ikke* nogen signifikant sammenhæng mellem ordførers køn og karakterer.

For at illustrere de signifikante sammenhænge vises nedenfor sammenhængen mellem de signifikante uafhængige variable og den afhængige variabel, når der er kontrolleret for den anden signifikante variabel. I praksis er dette gjort ved at opdele den afhængige variabel i kategorier og derefter plote gennemsnittet på residualen fra regressionen uden den pågældende variabel for hver af kategorierne.

Som det fremgår af figurene er der begge steder en uformet sammenhæng. De skoler der klarer sig bedst er således de helt små skoler, samt de store skoler. Skolerne i midten (mellem 18 og 100 elever) ser ud til at klare sig dårligst.

# De skjulte karakterene VEDLEGG

Figur 1. Sammenhæng mellem karaktergennemsnit og antal elever på skolen kontrolleret for indkomst i kommunen.

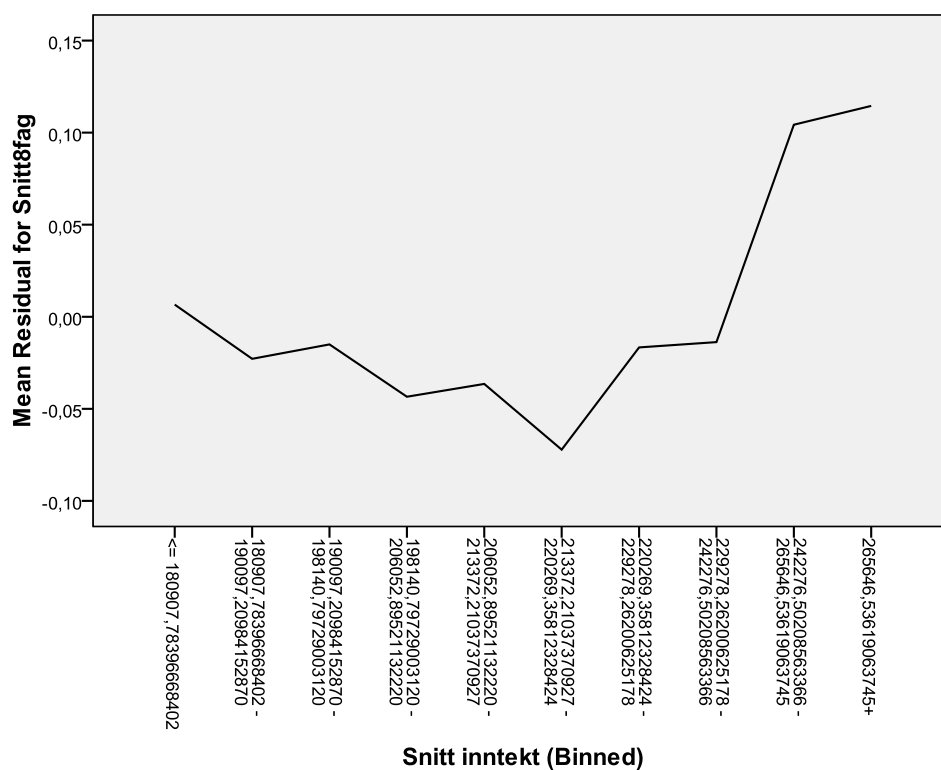


Med hensyn til indtægt er der også en kurvelineær sammenhæng. De laveste indtægter klarer sig relativt bedre end kommunerne med mellemindtægterne. Allerbedst går det for eleverne i kommunerne med de allerhøjeste indtægter. I denne sammenhæng skal man være opmærksom på, at der arbejdes med gennemsnit for en lang række individer (dels på skoleniveau og dels på kommuneniveau), og resultaterne skal derfor heller ikke tolkes som var de på individniveau.



# De skjulte karakterene VEDLEGG

Figur 2: Sammenhæng mellem gennemsnitlig indtægt i kommune og gennemsnitlig karakter kontrolleret for størrelse af skole



## Forbehold:

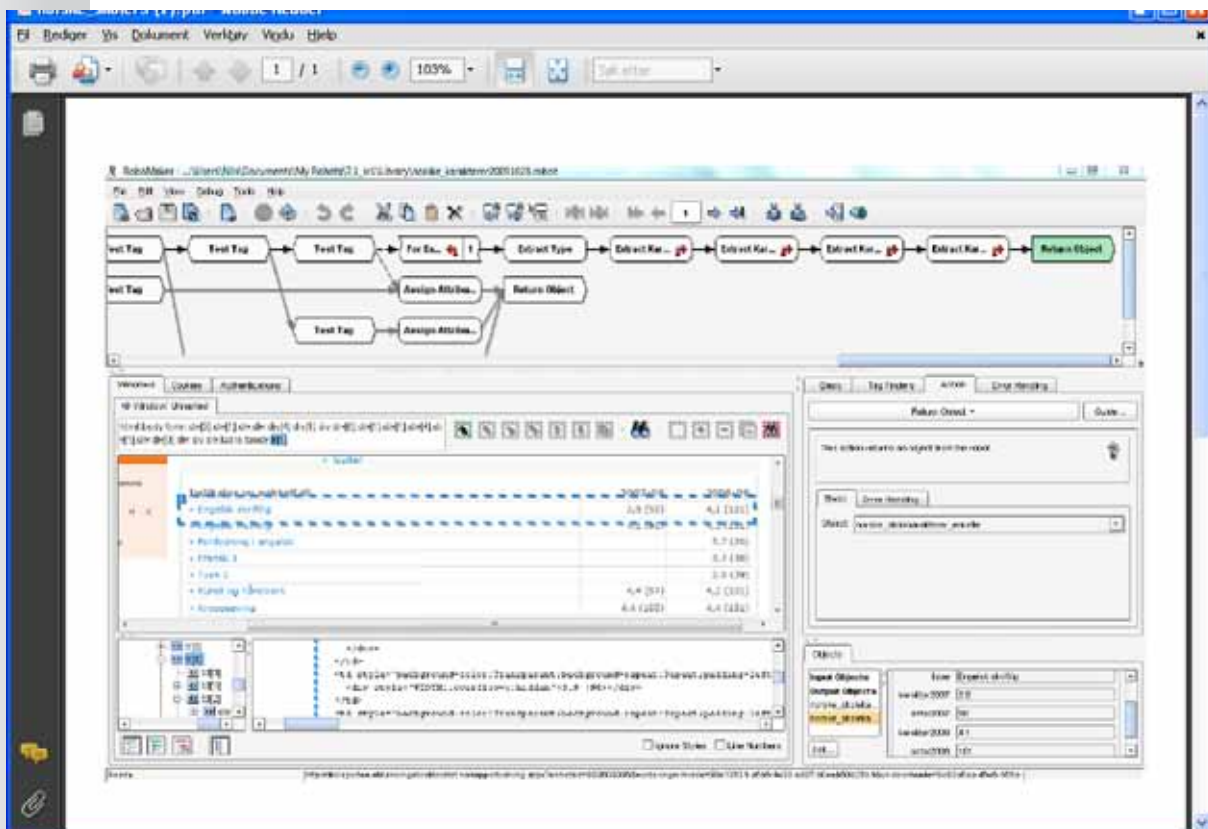
1. Da den eneste sociale baggrundsvariabel, der anvendes, er indtægt, er der givetvis en del varians i den afhængige variabel, som ikke er forklaret i modellen, men som kunne forklares ved hjælp af andre baggrunds variabler. Dette betyder eksempelvis, at såfremt man anvender flere ressourcer (dvs. højere udgift i modellen) for at kompensere for en svag social baggrund, så kan sammenhængen mellem udgifter og resultater være sløret, fordi der i modellen ikke er kontrolleret fuldt ud for social baggrund. Problemet vil være størst, hvis sammenhængen mellem indtægt og andre sociale baggrunds faktorer varierer meget mellem forskellige kommuner. Eller formuleret anderledes, hvis man i stort omfang kan forudse en anden variabel såsom uddannelse ved hjælp af indtægt er problemet ikke så stort.

2. Der arbejdes med gennemsnit på dels skoleniveau (karakterer) og dels kommuneniveau (udgift pr elev og gennemsnitlig indtægt). Det betyder, at en del af den interessante varians på individniveau ikke medtages, og sammenhængen mellem indtægt og gennemsnit bliver derfor mindre klar. Specielt skal man være meget forsigtig med på baggrund af analysen at drage konklusioner om hvorledes den enkelte skole klarer sig, fordi der i modellen kontrolleres for indtægts gennemsnit i kommunen.

# De skjulte karakterene VEDLEGG

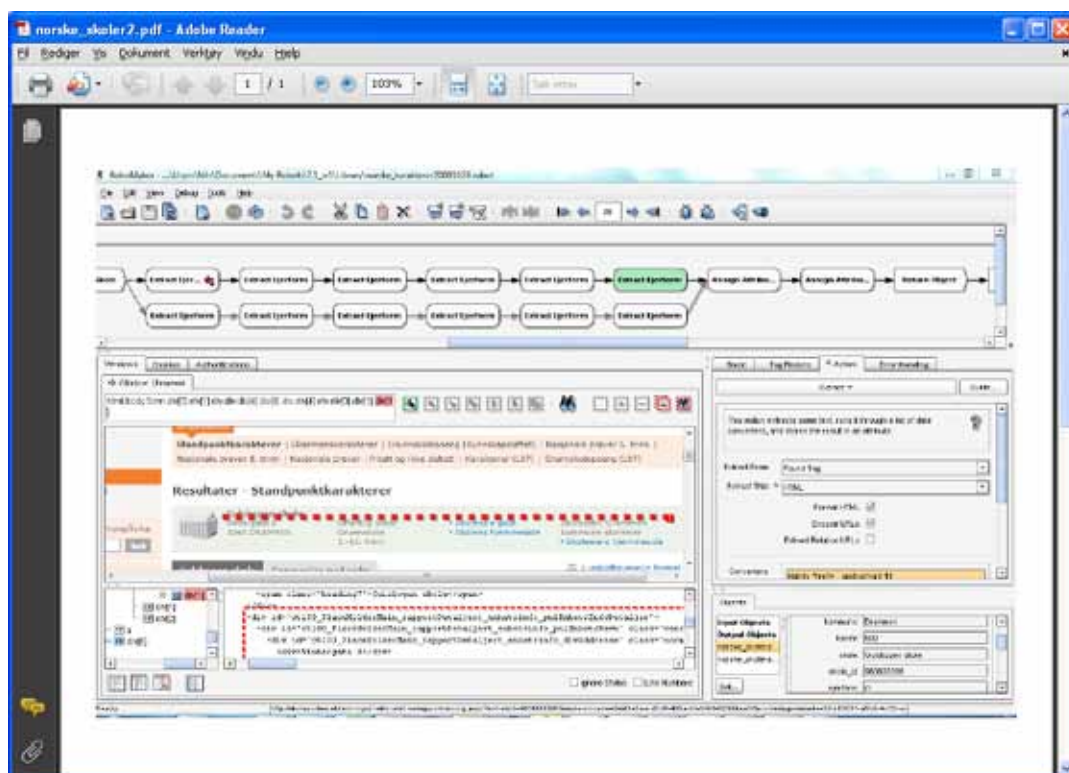
## Vedlegg 2: Robot-innstillingen

### SLIK BLE ROBOTEN STILT INN



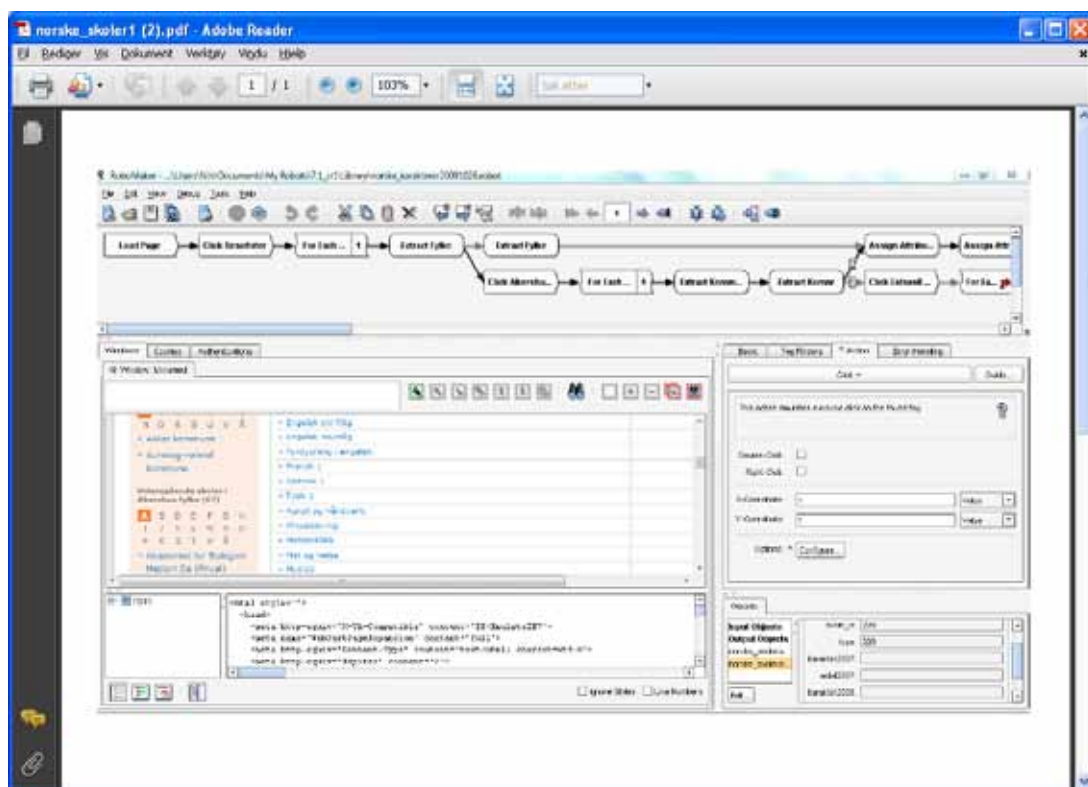
Trinn 1: Først måtte roboten stilles inn slik at den kunne velge fylke-kommune-grunnskole/videregående-offentlig/privat.

# De skjulte karakterene VEDLEGG



Trinn 2: Dernest må roboten stilles inn slik at den kan velge basisopplysninger om den enkelte skole: Id-nummer, adresse, epost, fylke-kommune, type osv.

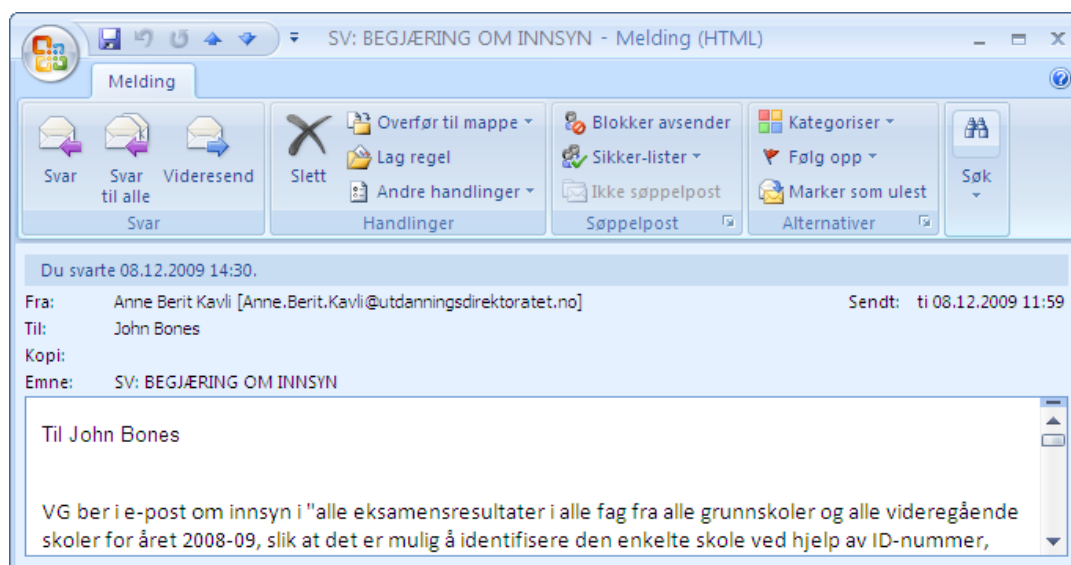
# De skjulte karakterene VEDLEGG



Trinn 3: Til slutt må roboten stilles inn for å trekke ut de detaljerte karakteropplysningene for hvert enkelt fag. Nøkkelen er id-nummeret for hver skole.

# De skjulte karakterene VEDLEGG

## Vedlegg 3: Innsynsbejæringen



Skjermdumpen over viser dateringen av svaret fra Utdanningsdirektoratet.

Under er selve korrespondansen.

Først vises svaret fra Utdanningsdirektoratet, under dette ligger vår innsynsbejæring, til sist ligger tilsagnet om innsyn i karakterer også fra videregående.

Til John Bones

VG ber i e-post om innsyn i "alle eksamensresultater i alle fag fra alle grunnskoler og alle videregående skoler for året 2008-09, slik at det er mulig å identifisere den enkelte skole ved hjelp av ID-nummer, kommunenummer, kommunenavn og fylke." VG ber videre om at dette blir "levert som Excel-filer eller i databaseformatet Access".

I følge offentleglova § 9 kan alle "... krevje innsyn i ei samanstilling av opplysningar som er elektronisk lagra i databasane til organet dersom samanstillinga kan gjerast med enkle framgangsmåtar."

Utdanningsdirektoratet har gjennomgått de elektronisk lagrede opplysningene det

# De skjulte karakterene VEDLEGG

bes om innsyn i, og finner at arbeidet med å sammenstille resultatene slik VG ber om, og samtidig ivareta bestemmelsene om personvern og taushetsplikt, ikke kan gjøres med enkle fremgangsmåter, jf offl. § 9. Bakgrunnen for dette er som følger:

VG ønsker ” innsyn i alle eksamensresultater i alle fag fra alle grunnskoler og alle videregående skoler for året 2008-09”. Utdanningsdirektoratet sitter imidlertid ikke med karakterer i alle fag i aggregert form (dvs på skolenivå og oppover), men kun på individnivå. Vi har som mål å aggregere opp resultatene for alle fag og publisere dem på et eget statistikkprogram direktoratet benytter seg av utenom Skoleporten (Nesstar), men vi har ikke hatt mulighet til å prioritere dette nå i høst. Det betyr at de fagene vi har ferdig aggregert (og prikket i tråd med kravene til personvern og taushetsplikt slik at opplysningene kan publiseres eller utleveres for innsyn) kun er de fagene som publiseres i Skoleporten. Det ville dermed medføre et betydelig arbeid for direktoratet å skulle aggregere og prikke alle dataene VG ber om innsyn i.

Karakterer på grunnskolen ble publisert i Skoleporten tidligere i høst. Denne statistikken inneholder eksamenskarakterer i de fagene som er publisert i Skoleporten, aggregert på skoler, kommuner, fylker og nasjonalt. Det betyr at hver skole ikke er mulig å identifisere ved hjelp av kommunenummer, kommunenavn og fylke, men kun med organisasjonsnummer og navn. Filen kan pga filens størrelse ikke leveres ut på Excel, men vi er innstilt på å levere den ut i Access. Filen vil bli på ca. 128 Mb, noe som gjør at vi bør avtale hvordan den kan leveres. For eksempel kan en minnepinne hentes etter avtale i direktoratet.

Forut for hver publisering av Utdanningsdirektoratets karakterstatistikk foregår det et meget omfattende statistisk kvalitetssikringsarbeid som blant annet også ivaretar hensynet til personvern og taushetsplikt. Dette er et arbeid som tar direktoratets spesialister flere uker. For årets statistikk har direktoratet i planleggingen av arbeidet tatt sikte på at statistikken for de videregående skolene vil være klar til publisering 15. desember, noe som er vesentlig tidligere enn publiseringer for tidligere år. Det er ikke mulig å forsere dette arbeidet utover det som er planlagt i forhold til publiseringsdatoen 15. desember. Den 15. desember vil dataene da være tilgjengelig på Skoleporten.

Vennligst si ifra dersom VG ønsker tilgang til grunnskolekarakterene før man kan få tilgang på karakterer for videregående skole (15. desember).

Mvh

Anne-Berit Kavli  
Avdelingsdirektør/Head of Department  
Utdanningsdirektoratet/Norwegian Directorate for Education and  
Training  
Tlf: +47 23301217/+47 97699069

# De skjulte karakterene VEDLEGG

**Fra:** John Bones [mailto:John.Bones@vg.no]  
**Sendt:** 2. desember 2009 16:04  
**Til:** Petter Skarheim; Helge Lund; Anne Berit Kavli  
**Emne:** BEGJÆRING OM INNSYN

Innsyn i eksamensresultater

Med henvisning til offentlighetsloven § 3, begjærer vi innsyn i alle eksamensresultater i alle fag fra alle grunnskoler og alle videregående skoler for året 2008-09, slik at det er mulig å identifisere den enkelte skole ved hjelp av ID-nummer, kommunenummer, kommunenavn og fylke. Ifølge offentlighetsloven § 3, lovens hovedregel, er alle *Saksdokument, journalar og liknande register for organet (er) opne for innsyn dersom ikkje anna følgjer av lov eller forskrift med heimel i lov.*"

Vi vil understreke at vi ikke ber om innsyn i opplysninger som direkte eller indirekte kan identifisere enkeltelever. I denne sammenheng viser vi til Utdanningsdirektoratets egne retningslinjer punkt 4.2.1 der det blant annet heter: Offentlighetsloven gir ikke hjemmel for å nekte innsyn i resultater så lenge det ikke er grunnlag for direkte eller indirekte identifisering av enkeltelever. Ved eventuell innsynsbeğjæring av skolerresultater kan det derfor bli aktuelt å gi innsyn i rapporter for skoler der identifisering av enkeltelever ikke er mulig.

Vi ønsker opplysningene levert som Excel-filer eller i databaseformatet Access og henviser i denne sammenheng til offl. § 9. *Rett til å krevje innsyn i ei samanstilling frå databasar, hvor det heter: "Alle kan krevje innsyn i ei samanstilling av opplysningar som er elektronisk lagra i databasane til organet dersom samanstillinga kan gjerast med enkle framgangsmåtar."* Vi antar at det ikke bør by på nevneverdige problemer å sammenstille de aktuelle problemstillingene.

Vi har for øvrig allerede lastet ned informasjonen om standpunkt-karakterer ved hjelp av webscraping. Nå ønsker vi å se på forholdet mellom standpunkt og eksamens-karakterer, ettersom det bør være av offentlig interesse å få kunnskap om eventuelle avvik.

Med hilsen

John Bones  
Journalist VG

Denne eposten har blitt skannet for virus av Utdanningsdirektoratets antivirusprogramvare.  
Denne eposten har blitt skannet for virus av Utdanningsdirektoratets antivirusprogramvare.

# De skjulte karakterene VEDLEGG

16.desember fikk vi beskjed om at vi også fikk tilgang til karakterene fra videregående, slik det meddeles i eposten under:

