

Dansk Erhvervs gymnasieanalyse – *Sådan gør vi*

FORMÅL

Formålet med analysen er at undersøge, hvor dygtige de enkelte gymnasier er til at løfte elevernes faglige niveau. Dette kan man ikke undersøge blot ved at sammenligne de rå gymnasiekarakterer, da de i høj grad afhænger af, hvilke elever der optages på de enkelte gymnasier. Derfor udregnes i stedet en *gymnasieeffekt*, hvor der tages højde for elevernes faglige og sociale baggrund.

Ved hjælp af økonometriske redskaber har Dansk Erhverv derfor undersøgt, hvor stor forskel der er på gymnasiernes evne til at løfte elevernes faglige niveau, efter der er taget højde for elevernes folkeskolekarakterer og andre sociale baggrundsvariable såsom forældrenes uddannelsesniveau og arbejdsmarkedsstatus.

Formålet er herved at tage højde for de forhold, der varierer mellem gymnasierne, og som gymnasiet ikke selv er herre over, så man kun måler effekterne af de forskelle, som gymnasierne selv kan skabe.

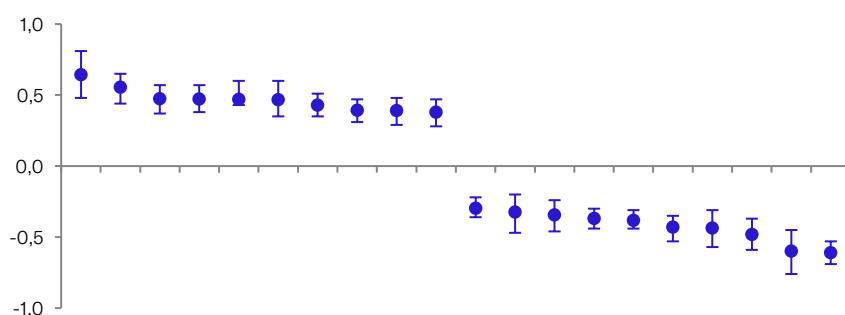
Resultaterne viser, at selv når man tager højde for den statistiske usikkerhed, er der betydelige forskelle mellem gymnasiers evne til at løfte elevernes faglighed, figur 1.

Rå karakterdata kan ikke bruges til sammenligninger på tværs af gymnasier.

Måler gymnasiernes løfteevne ved at tage højde for elevsammensætningen

Gymnasierne bør vurderes på de parametre, som de selv har mulighed for at påvirke.

Figur 1: **Gymnasieeffekt (2015-2017) for top 10 og bund 10 (STX)**



Signifikant forskel mellem de 10 bedste og de 10 dårligste gymnasier.

Kilde: Dansk Erhverv på baggrund af registerdata fra Danmarks Statistik

Anm.: Gymnasieeffekterne er angivet med et 95 pct. konfidensinterval

Data

Dansk Erhverv har ved hjælp af en særkørsel i Danmarks Statistiks *registerdatabase* data for samtlige elever, der i perioden 2010-2017 har afsluttet en gymnasieuddannelse (STX, HHX eller HTX). Foruden elevens gymnasiekarakterer indeholder data oplysninger om elevens folkeskolekarakterer og socioøkonomiske baggrund (forældres baggrund, økonomi, m.v.).

Gymnasieeffekten beregnes hen over en treårig periode, 2015-2017. På den måde kan den estimerede effekt sættes direkte i relation til et gymnasieforløb for den typiske elev.

Gymnasieeffekten beregnes med udgangspunkt i de skriftlige eksamenskarakterer, hvilket betyder, at samtlige standpunktskarakterer samt mundtlige eksamenskarakterer er udeladt af modellen. Dette er et bevidst valg, da det vurderes, at karaktererne fra de skriftlige eksamener i højere grad er et objektive mål for elevernes faglighed på tværs af gymnasier, eftersom de skriftlige eksamener er udarbejdet fra centralt hold af Undervisningsministeriet, hvormed disse er ens for alle, ligesom bedømmelserne foretages af to tilfældige eksterne censorer.

Grundet mængden af baggrundsvariable i regressionsmodellen, opstår der nemt problemer med manglende enkeltobservationer for flere elever, hvilket i givet fald betyder, at forholdsvis mange elever udelades. For at undgå at udelade en stor mængde af observationer, medtages elever med manglende oplysninger for én eller flere baggrundsvariable i estimationen ved, at den manglende oplysning sættes til 0, og der medtages en "missing value"-indikatorvariabel, som sættes til én. På denne måde indgår eleven ikke i estimationen af koefficienten på den manglende variabel, men eleven beholdes i data-sættet og indgår i estimationen af de øvrige koefficienter. Dette er sædvanlig standard i denne type beregninger.

Der rapporteres kun resultater for gymnasier, som for hvert år i perioden 2015-2017, har haft afgivet et minimum på enten 200 eksamenskarakterer og haft mindst 45 elever til en skriftlig eksamen (og mindst 150 elever over de tre år samle set). Dette skyldes, at der er større usikkerhed ved estimation af gymnasieeffekten for små gymnasier.

Metode

Dansk Erhverv benytter den såkaldte *residual-metode* til at estimere de gymnasiale undervisningseffekter. Modellen er en standard OLS-model, hvor elevens gymnasiekaraktererⁱ forklares enkeltvis ved hjælp af henholdsvis elevens folkeskolekarakterer fra 9. klasse samt elevens sociale baggrundskarakteristika. Den del af gymnasiekarakteren, der ikke kan forklares ved hjælp af henholdsvis elevens folkeskolekarakterer og sociale baggrundskarakteristika, dvs. residualet, defineres som *gymnasieeffekten*.

Registerdata fra Danmarks Statistik

I modellen indgår studenter i perioden 2015-2017.

Skriftlige karakterer vurderes at være det mest objektive mål for faglighed på tværs af gymnasier.

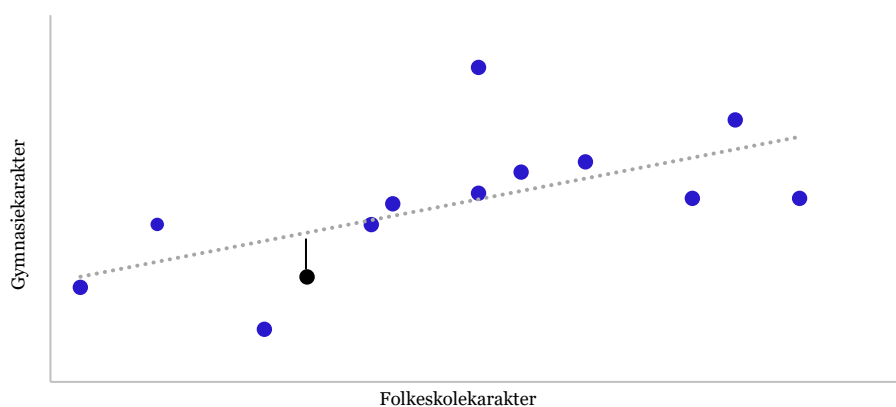
Elever med manglende observationer for én eller flere baggrundsvariable indgår i modellen vha. en "missing value"-indikatorvariabel.

Små gymnasier udelades grundet statistisk usikkerhed.

Gymnasieeffekt: Afvigelse mellem elevens faktiske karakter og forventede karakter.

Meget forsimplet er dette illustreret i figur 2. Den stiplede linje i grafen viser den generelle sammenhæng mellem karakterniveau i folkeskolen og karakteren på gymnasiet. Jo højere karakterniveau i folkeskolen des højere karakterniveau i gymnasiet. Der er dog naturligvis elever, der får karakterer, der er højere eller lavere, end det forventede. Dette er den såkaldte *gymnasieeffekt* og er illustreret ved afstanden mellem den stiplede linje og den enkelte sorte prik. Ligger man over linjen, er der en positiv gymnasieeffekt. Ligger man under linjen, er der en negativ effekt.

Figur 2: **Gymnasieeffekten i en forsimplet illustration**



Kilde: Dansk Erhverv.

Der er naturligvis en variation i karaktererne, som ikke kan forklares ved hjælp af baggrundsoplysninger (to søskende med samme karaktergennemsnit i folkeskolen vil eksempelvis ikke nødvendigvis få samme gymnasiekarakter), men ved at summere på tværs af alle karakterer for alle elever på et gymnasium, kan man se de systematiske forskelle mellem de enkelte gymnasier.

Systematiske forskelle mellem gymnasiernes undervisningseffekt.

På matematisk form er modellen bygget op på nedenstående måde og en detaljeret oversigt af modellens forklarende variable fremgår af tabel 1.

$$gymnasiekar_{ifg} = \beta_0 + \beta_1 d_f + \beta_2 fagniveau + \beta_3 folkeskolekar_i + \beta_4 d_f folkeskolekar_i + \beta_5 10.klasse_i + \beta_6 socialbag_i + \beta_7 afgangsår + \beta_8 skolekammerat + \varepsilon_{ifg}$$

hvor $gymnasiekar_{ifg}$ angiver karakteren i gymnasiefaget f for person i på gymnasium g , d_f er en vektor af dummys, som angiver, hvilken gymnasiekarakter der forklares og $fagniveau$ angiver, hvilket niveau faget er taget på. $Folkeskolekar_i$ er en vektor bestående af folkeskolekarakterer for person i og $d_f * folkeskolekar_i$ er et interaktionsled mellem de to førnævnte. $10.klasse_i$ angiver om den pågældende elev har gået i 10. klasse, $socialbag_i$

Beskrivelse af modellens opbygning.

er en vektor bestående af sociale baggrundsvariable for eleven og $afgangsår_i$ angiver året eleven blev student. Slutteligt er $skolekammerat_g$ den såkaldte skolekammerateffekt og ε_{ifg} er regressionens restled (residual).

Gymnasiernes samlede *gymnasieeffekt* fremkommer herefter ved at beregne et gennemsnit af residualen for samtlige skriftlige eksamenskarakterer for eleverne på det pågældende gymnasium.

Gymnasieeffekten er teoretisk set lig OLS-modelens residual

Tabel 1: **Oversigt over modellens baggrundsvariable**

Variabel	Variabelinformation
Information om eleven	
Køn	
Forældreidentifikation	Er forældrene observeret (fx hvis eleven ikke har nogen mor)
Familiemarkering	Angiver om en person bor på samme adresse som begge sine forældre, kun moderen, kun faderen eller ingen af dem
Herkomst	Dansk eller vestlig/ikke-vestlig indvandrer/efterkommer
Folkeskolekarakter	Folkeskolekaraktererne indgår som et gennemsnit af skriftlige eksamenskarakterer i fagene Dansk og Matematik.
Har eleven haft tysk eller fransk	
Har eleven gået i 10. klasse	
Information om forældrene	
Uddannelsesniveau	
Arbejdsmarkedsstatus	
Alder ved barns fødsel	
Andre baggrundsvariable	
Fagniveau karakteren er opnået i	Har betydning om en given karakter er opnået i fx et A-niveau eller B-niveau fag
Skolekammerateffekt	Gymnasiegennemsnit af elevernes folkeskolekarakterer på den pågældende årgang

Skolekammerateffekten angiver elevsammensætningen på det pågældende gymnasium. I vores model er skolekammerateffekten beregnet som et gennemsnit af elevernes gennemsnitlige folkeskolekarakterer i skriftlige eksamener. Skolekammerateffekten medtages, da tidligere studier viser, at elevsammensætningen på en skole har betydning for den enkelte elevs karakterer. Med andre ord vil en given elev klare sig bedre, såfremt denne går på et gymnasium med mange fagligt stærke elever, fremfor på et gymnasium med mange fagligt svage elever. Eftersom gymnasiernes elevsammensætning ikke er noget gymnasiene direkte kan kontrollere, er det derfor væsentligt at korrigere for denne effekt, når gymnasieeffekten skal beregnes.

Skolekammerater påvirker hinandens faglighed.

Usikkerheder og robusthedstest

Det metodiske valg, der uden sammenligning har størst betydning for resultaterne, er valget af kun at medtage elevernes skriftlige karakterer. De øvrige metodiske valg har også betydning, men ændrer dog kun gymnasieeffekterne i begrænset omfang, og flytter derfor kun gymnasierne få pladser på listen.

Som beskrevet tidligere, er det mere retvisende at begrænse modellen til de skriftlige karakterer, da de vurderes at være bedre mål for fagligheden. Som konsekvens heraf er vores model derfor kun i stand til at måle den faglighed, der kan måles med de skriftlige eksamener, upåagtet af, at andre faglige kundskaber også har betydning for elevernes overordnet faglige niveau.

I modellen medtages en række baggrundskarakterer for at sikre, at de forhold, gymnasiet ikke selv er herre over (bl.a. hvem der optages), ikke har betydning for den udregnede gymnasieeffekt. Da man kan anvende folkeskolekarakterer som kontrolvariabel, kan dette foretages mere effektivt, end eksempelvis tilsvarende analyser af undervisningseffekterne på folkeskoleområdet. Det betyder også, at små ændringer i andre baggrundsfaktorer vedr. forældre mv. ikke har særlig stor betydning for modellens forklaringssevne. Det er dog trods alt ikke muligt statistisk at bortkontrollere alle forskelle (to tvillinger med samme folkeskolekarakterer, vil givetvis ikke opnå helt samme gymnasie-karakterer). Dette er dog ikke problematisk, så længe disse forskelle ikke er systematiske.

Resultaterne er, som i alle statistiske modeller, behæftet med en statistisk usikkerhed. Denne er beregnet vha. af den såkaldte *bootstrap metode*. Resultaterne er derfor afrapporteret som konfidensintervaller omkring den enkelte gymnasieeffekt. Konfidensintervallet angiver det interval, inden for hvilket, man kan være 95 pct. sikker på, at den reelle gymnasieeffekt ligger.

Valg af karakterdata har størst betydning for analysens resultater.

Modellen måler kun den faglighed, som opfanges i de skriftlige eksamener.

Folkeskolekarakterer kan forklare gymnasiekarakterne forholdsvis præcist.

Benytter bootstrap metode til at udregne konfidensintervaller.

NOTER

¹ Det er den enkelte karakter per elev, der estimeres, og ikke et simpelt gennemsnit af en elevs gymnasie-karakter. Derved indgår den enkelte elev i modellen det antal gange, der svarer til antallet af skriftlige eksamenskarakterer eleven har modtaget. Fordelen ved at estimere den enkelte karakter og ikke et karaktergennemsnit er, at man ved et gennemsnit mister varians i data.