



**BØRNE- OG
UNDERVISNINGS-
MINISTERIET**
STYRELSEN
FOR IT OG LÆRING

De socioøkonomiske referencer for gymnasiekarakterer

Metodenotat

Indhold

Indledning	3
De socioøkonomiske referencer	4
Den statistiske model	6
1.1 Baggrundsoplysninger	6
1.2 Modellen for de socioøkonomiske referencer	8
Datagrundlag.....	11
Studenter	11
Fag, prøveform og niveau.....	11
Afdelingsniveau	12

Indledning

Dette metodenotat indeholder en gennemgang af metoden anvendt for de socioøkonomiske referencer. Dette indebærer både en beskrivelse af den statistiske model samt anvendt data og baggrundsvariable.

Den socioøkonomiske reference er et statistisk udtryk for, hvordan studenter på landsplan med samme baggrundsforhold som uddannelsesstedets studenter har klaret prøverne. *Socioøkonomisk* refererer til studenternes sociale og økonomiske baggrund, mens *reference* fortæller, at tallet kan bruges som et sammenligningsgrundlag for uddannelsesstedes faktisk opnåede eksamensresultater og prøvekarakterer.

Man taler også om den socioøkonomiske reference som uddannelsesstedets *forventede karakter* eller *løfteevne*, når der er taget højde for studenternes baggrund.

De efterfølgende afsnit vil give en baggrund for de socioøkonomiske referencer, en gennemgang af den statistiske model samt baggrundsvariable. Afslutningsvis indeholder indeværende metodenotat et afsnit om data og datagrundlaget for modellen.

De socioøkonomiske referencer

Det enkelte uddannelsessteds eksamensresultat og eksamenskarakterer har sammenhæng med mange forskellige forhold. Der er både forhold, som uddannelsesstedet har indflydelse på, og forhold som uddannelsesstedet ikke har indflydelse på. Uddannelsesstedets elevgrundlag i bred forstand ligger uden for dets indflydelse.

Mange undersøgelser har vist, at en væsentlig forklaring på de opnåede karakterer på de gymnasiale uddannelser knytter sig til studenternes faglige niveau fra grundskolen målt ved afgangskaraktererne fra 9. klasse (Folkeskolens prøver (FP)) samt til studenternes socioøkonomiske baggrund. Den socioøkonomiske baggrund skal forstås som de sociale og økonomiske kår, som studenten kommer fra. Især forældrenes uddannelsesmæssige baggrund har vist sig at have en markant sammenhæng med de opnåede karakterer, men andre faktorer som herkomst, forældrenes økonomiske forhold og familietype samt studentens adgangsvej til den gymnasiale uddannelse udviser også en sammenhæng.

Selv om elevgrundlaget i udgangspunktet ikke ligger inden for uddannelsesstedets indflydelse, er det vigtigt at huske på, at uddannelsesstedet også har mulighed for at indrette sin undervisning efter de elever, der går på den.

Ved at foretage statistiske beregninger, i andre sammenhænge ofte refereret til som korrektioner, er det muligt til dels at forklare de forskelle i uddannelsesstedets karaktergennemsnit, som skyldes forskelle i elevgrundlaget. Herved opnås et mere nuanceret billede af uddannelsesstedet niveau målt ved eksamensresultatet og eksamenskaraktererne. Beregningerne skal helst tage højde for alle de forhold, som uddannelsesstedet ikke har direkte indflydelse på, f.eks. elevernes faglige niveau inden start på den gymnasiale uddannelse, elevens køn, herkomst, forældrenes uddannelsesniveau mv.

Der er dog forhold, som ikke er direkte målbare og/eller statistisk registrerede, og som derfor ikke kan inddrages. Det gælder f.eks. motivation, familie og venners påvirkning, sygdom, handicap mv. Beregningerne skal som følge heraf betragtes med det forbehold, at der ikke nødvendigvis er taget højde for samtlige forhold uden for uddannelsesstedets indflydelse.

Vurderingen af det enkelte uddannelsessteds gennemsnit af eksamensresultaterne og eksamenskaraktererne set i forhold til studenternes faglige og sociale baggrund tilvejebringes ved, at det faktisk opnåede eksamensresultat sammenlignes med et statistisk beregnet udtryk for, hvad studenter på landsplan med lignende baggrundsforhold forudsiges at opnå. Dette tal defineres som en socioøkonomisk reference.

Forskellen mellem det faktisk opnåede eksamensresultat og den socioøkonomiske reference kan nu beregnes. En eventuel forskel skyldes nu ikke, at uddannelsesstedets studenter kommer fra hjem med et bestemt uddannelsesniveau, eller har en bestemt fordeling med hensyn til herkomst. Disse forhold er der taget højde for.

Da studenternes faglige niveau fra grundskolen målt ved afgangskaraktererne i 9. klasse indgår i modellen, kaldes denne type model for 'value added models' eller model for *uddannelsesstedets løfteevne*.

Udtrykket 'socioøkonomisk reference' anvendes til at beskrive, hvilket karakterniveau man statistisk set ville forvente på uddannelsesstedet ud fra dens elevsammensætning.

Både uddannelsessteder med eksamensresultatgennemsnit i den lave og i den høje ende vil kunne ligge lavere end deres socioøkonomiske reference, når der tages højde for elevsammensætningen. Hvis et uddannelsessted med et højt eksamensgennemsnit ligger lavere end dens socioøkonomiske reference, vil det typisk være et uddannelsessted, som har overvejende 'resourcestærke' studenter, og hvor de statistiske modeller forudsiger, at uddannelsesstedet burde opnå et endnu højere gennemsnit, end det faktisk gør.

Tilsvarende kan uddannelsessteder med gennemsnit af eksamensresultat i den lave ende ligge højere end deres socioøkonomiske reference, når der er taget højde for elevsammensætningen. Dette kan f.eks. være uddannelsessteder, hvor studenternes karaktergennemsnit fra 9. klasse er lavt og/eller uddannelsessteder, hvor andelen af forældre med ingen eller kort uddannelse er relativt høj, og hvor institutionens resultat overstiger det, man ville forvente i forhold til elevsammensætningen.

På samme måde kan uddannelsessteder med et højt gennemsnit også ligge højere end deres socioøkonomiske reference, og uddannelsessteder med et lavt gennemsnit kan på samme måde også ligge lavere.

Når et uddannelsessteds eksamensresultater vurderes i forhold til dens socioøkonomiske reference, er det vigtigt, at der tages højde for den tilfældige variation, der er mellem studenternes karakterer, således at der ikke konkluderes på et resultat, som lige så godt kunne være opstået ved en tilfældighed. Uddannelsesstedets størrelse, målt ved antallet af studenter med en prøvekarakterer, har en væsentlig betydning for, hvor sikkert den socioøkonomiske reference er bestemt. Færre studenter giver mere usikre resultater. Derfor beregnes et sikkerhedsinterval¹, således at det er muligt at afgøre, hvorvidt en forskel mellem uddannelsesstedets gennemsnit af eksamensresultat og den beregnede socioøkonomiske reference er reel, eller om den blot kan være et udtryk for tilfældig variation.

¹ Der anvendes et 95 % sikkerhedsinterval.

Den statistiske model

1.1 Baggrundsoplysninger

Tabel 1 Baggrundsoplysninger

Baggrundsoplysning	Kodning
9. klasse FSA karaktergennemsnit	Gennemsnit af bundne prøver (ekskl. dansk or- den) ² .
Køn	Dreng Pige
Alder på tidspunkt for afslutning af uddannelsen	18 år eller yngre 19 år 20 år 21 år eller ældre
Herkomst³	Dansk Indvandrer Efterkommer
Forældrenes højest fuldførte uddannelse⁴	Grundskolen Gymnasial uddannelse

² Eleven skal have karakter i mindst 4 ud af 8 (2007->) eller 4 ud af 7 (2002-06) bundne prøvefag.

³ Der skelnes mellem indvandrere, efterkommere og personer med dansk oprindelse. Personer med dansk oprindelse (dansk) er personer, hvor mindst en af forældrene er dansk statsborger og født i Danmark.

Efterkommere er født i Danmark. Ingen af forældrene er danske statsborgere og født i Danmark. Når en eller begge forældre, der er født i Danmark, opnår dansk statsborgerskab, vil deres børn ikke længere blive klassificeret som efterkommere. Fastholder danskfødte forældre imidlertid begge et udenlandsk statsborgerskab, vil deres børn blive klassificeret som efterkommere. Hvis der ikke findes oplysninger om nogen af forældrene, og personen er udenlandsk statsborger og født i Danmark, opfattes personen også som efterkommer.

Indvandrere er født i udlandet. Ingen af forældrene er danske statsborgere og født i Danmark. Hvis der ikke findes oplysninger om nogen af forældrene, og personen er født i udlandet, opfattes personen som indvandrer.

⁴ Den højest fuldførte uddannelse blandt elevens mor og far oplyst ved start på uddannelsen. Hvis uoplyst, da indgår oplysningen fra det år eleven fyldte 13 år.

Baggrundsoplysning	Kodning
	Erhvervsfaglig uddannelse Kort videregående uddannelse Mellemlang videregående uddannelse Lang videregående uddannelse (inkl. ph.d. mv.)

Tabel 1 Baggrundsoplysninger (fortsat)

Baggrundsoplysning	Kodning
Forældrenes gennemsnitlige bruttoindkomst⁵	- 300.000 kr.
	300.000 - 400.000 kr.
	400.000 - 500.000 kr.
	500.000+ kr.
Familiestatus⁶	Hjemmeboende hos par
	Hjemmeboende hos enlig
	Udeboende
Adgangsvej⁷	Folkeskole
	Friskole og privat grundskole
	Efterskole
	Anden
10. klasse⁸	Nej
	Ja

1.2 Modellen for de socioøkonomiske referencer

Den anvendte statistiske model er en såkaldt multilevel model:

$$y_{ij} = x_{ij}\beta + u_j + e_{ij}$$

hvor y_{ij} er karakteren for student i på uddannelsessted j , x_{ij} er studentens baggrundsvariable og β er de tilhørende parameterestimer, u_j er varianskomponenten svarende til variationen mellem uddannelsesstedet, og e_{ij} er residualen svarende til variationen mellem eleverne på

⁵ Bruttoindkomsten er i året for start på uddannelsen.

⁶ Familiestatus i året for start på uddannelsen er en kombination af familietype og plads. Familietypen par indeholder ægtepar, registreret partnerskab, samlevende og samboende par. Plads er enten hjemmeboende eller udeboende. Hvis uoplyst da indgår oplysningen fra det år eleven fyldte 13 år.

⁷ Seneste grundskoletype (9. klasse eller afsluttet 10. klasse) inden start på uddannelsen.

⁸ Gennemført 10. klasse.

uddannelsesstedet. For hver kombination af fag, prøveform og niveau foregår beregningerne på individniveau for hver af de gymnasiale uddannelser hver for sig.

Ved anvendelse af modellen fås et estimat over, hvor stor en del af variationen mellem karaktererne der skyldes forskelle mellem studenter, og hvor stor en del der skyldes forskelle mellem uddannelsesstederne. Dette kaldes modellens forklaringsgrad.

Alle de socioøkonomiske baggrundsvariable beskrevet i tabel 1 (side 6) medtages i modellen sammen med interaktionen mellem forældrenes uddannelse og studentens herkomst. Desuden medtages variable samlet for uddannelsesstedet i form af andelen af indvandrere og efterkommere på uddannelsesstedet, andelen af studenter med forældre hvis højest fuldførte uddannelse mindst er en mellemlang videregående uddannelse samt andelen af studenter med forældre, der har en lav gennemsnitlig bruttoindkomst. Interaktionen medtages for at undersøge, om effekten af forældrenes uddannelse afhænger af studentens herkomst. Variablene for uddannelsesstedet medtages for at undersøge, hvorvidt der er en 'klassekammerat effekt', dvs. om der er en øget effekt af, at der er en høj/lav andel af studenter med en bestemt baggrund. Øvrige interaktioner og variable for uddannelsesstedet er ikke medtaget, dels for ikke at øge modellernes kompleksitet, dels viser undersøgelser, at disse ikke øger modellernes forklaringskraft væsentligt. I den statistiske model for beregning af socioøkonomiske referencer for 3-års perioden medtages endvidere dimittendåret som en forklarende variabel. Herved tages der højde for, at karakterniveauet kan være forskelligt fra år til år.

Følgende baggrundsvariable er ikke statistisk signifikante i den multiple model og medtages derfor ikke i den endelige beregning af de socioøkonomiske referencer: forældrenes arbejdsmarkedsstatus samt alle variable om andelen for uddannelsesstedet. Når der således er taget højde for studenternes individuelle baggrundsoplysninger, kan der ikke i disse modeller konstateres en øget effekt af, at der er en høj/lav andel af studenter med en bestemt baggrund på det enkelte uddannelsessted.

Ikke alle studenter har fyldestgørende oplysninger på de her benyttede baggrundsvariable. En ofte anvendt metode til håndtering af dette er 'missing value' indikator metoden, hvor uoplyst medtages som en selvstændig kategori på lige fod med baggrundsvariablenes øvrige kategorier, og der tilføjes en indikatorvariabel, der fortæller hvorvidt, der er tale om en uoplyst værdi.

Omfanget af ukendt oplysning er dog ikke uafhængig af herkomst. Oplysningsgraden for alle baggrundsvariablene er størst for studenter af dansk herkomst. Der er således ikke tale om tilfældige 'huller' i informationen. Det er f.eks. specielt udtalt blandt indvandrere, at informationen om forældrenes uddannelse mangler. Da uddannelsesniveauet blandt de indvandrere og efterkommere, hvor uddannelsesoplysningerne kendes, er en del anderledes end blandt personer med dansk herkomst, vil der være en stor risiko for, at 'missing value' indikator metoden fejlestimerer effekten af såvel herkomst som af uddannelse. Tilsvarende problematik eksisterer for de øvrige baggrundsvariable.

Derfor er der i stedet valgt at indsætte en værdi, et såkaldt bedste gæt, på alle de steder, hvor oplysningerne er ukendte. For hver student med ukendt baggrundsoplysning gættes fem gange, således at modellen bliver estimeret på baggrund af en sandsynlig fordeling. Metoden kaldes 'multiple imputation'. Gættene baseres på fordelingen blandt studenter med oplyst baggrundsvariabel, hvor der er taget hensyn til studentens køn, alder og herkomst. Den socioøkonomiske reference for hver student er så et gennemsnit af de fem beregnede værdier.

For hver student beregnes en socioøkonomisk reference ud fra værdierne af studentens baggrundsvariable og de estimerede parametre for baggrundsvariablene. Herefter beregnes forskellen mellem opnået karakter og den socioøkonomiske reference, det såkaldte residual r_{ij} , for hver student.

På hovedinstitutionsniveau er den estimerede forskel mellem karaktergennemsnittet og den socioøkonomiske reference i en ordinær regressionsmodel gennemsnittet af disse studentresidualer på det enkelte uddannelsessted, \bar{r}_j . I de her anvendte to-niveau modeller, hvor vi både har variation mellem uddannelsesstedet ('Between'), og variation mellem studenter inden for uddannelsesstedet ('Within'), bliver residualerne på hovedinstitutionsniveau:

$$\hat{u}_j = c_j \bar{r}_j$$

hvor c_j er den såkaldte 'shrinkage' faktor, defineret ved

$$c_j = \frac{B}{(B + W/n_j)}$$

B er variationen mellem uddannelsesstedet ('Between'), W er residualvariationen mellem studenter inden for uddannelsesstedet ('Within') og n_j er antallet af studenter på den j 'te uddannelsesstedet.

For forskellen mellem opnået karakter og den socioøkonomiske reference for hvert uddannelsessted er der endvidere udregnet et 95 % sikkerhedsinterval. Hvis dette sikkerhedsinterval ligger over 0, opnår uddannelsesstedets studenter et statistisk signifikant højere karaktergennemsnit end studenter på landsplan med lignende baggrundsforhold, mens det modsatte er tilfældet, hvis sikkerhedsintervallet ligger under 0. Indeholder 95 % sikkerhedsintervallet værdien 0, da opnår studenter på det pågældende uddannelsessted et karaktergennemsnit som studenter på landsplan med lignende baggrundsforhold.

Sikkerhedsintervallet afhænger dels af den estimerede variation og dels af antallet af studenter, der indgår i beregningerne for det enkelte uddannelsessted. Således bliver sikkerhedsintervallerne mindre for det enkelte uddannelsessted, når der betragtes en 3-årig periode, idet der her indgår 3 studenterårgange i beregningerne.

Datagrundlag

Beregningen af socioøkonomiske referencer er foretaget for studenter fra ét år (eksempelvis 2019) og studenter fra 3 år (eksempelvis 2017, 2018 og 2019). Beregningen for den 3-årige periode har det mindste sikkerhedsinterval, idet der indgår 3 studenterår gange i beregningen.

Karakterdataene er baseret på indberetninger til STIL, mens alle oplysningerne vedrørende studenternes baggrundsforhold er hentet fra registre hos Danmarks Statistik.

Studenter

Beregningen af de socioøkonomiske referencer er baseret på følgende studenter:

- Studenter på hf (2-årig), hhx, htx og stx. Det betyder at studenter på hf-enkeltfag og på studenterkursus ikke indgår.
- Studenterne skal have afsluttet eksamen (opnået eksamensbevis).
- Minimum fem studenter per uddannelsessted.
- Studenter med registrerede afgangskarakterer fra grundskolen⁹

Fag, prøveform og niveau

For at beregne socioøkonomiske referencer på fag, opstilles der et minimumskrav til antal prøvekarakter på landsplan i det pågældende fag blandt studenterne.

Beregningen er foretaget dels for det samlede eksamensresultat (inklusive bonus A) og dels for prøvekaraktererne i en række fag for hver uddannelse. Dvs. at det er prøvekaraktererne på eksamensbeviset, der er taget udgangspunkt i, og fagene kan derfor være taget et eller to år tidligere end selve bevisåret. Alle beregninger er foretaget på individniveau, enkeltvis for hver af de gymnasiale uddannelser, men opgøres udelukkende på afdelingsniveau.

Alle karakterer er angivet på 7-trins skalaen.

Beregningerne foretages for prøvekarakterer aflagt ved prøveformene: mundtlig, skriftlig, opgave/projekt og samlet vurdering.

⁹ Da der kun er inkluderet studenter med en registreret afgangskarakter fra grundskolen, kan karaktergennemsnit for enkelte institutioner afvige en anelse fra karaktergennemsnittet baseret på samtlige elever. Afvigelsen er størst for hf institutioner, da disse har en større andel af ældre elever, hvis grundskolekarakterer ikke findes i det eksisterende grundskolekarakterregister.

Grundet Covid-19 i 2020¹⁰, 2021¹¹ og 2022¹² har prøveafleggelsen været begrænset. Det betyder at studenter i visse tilfælde har fået opløftet deres årskarakterer til også at være prøvekarakterer. Der beregnes kun socioøkonomiske referencer for prøvekarakterer, og det har derfor været metodisk problematisk at lave analyser på en blanding af prøvekarakterer og ophøjede årskarakterer. Beregningerne er derfor afgrænset til de studenter, som faktisk har været til eksamen. Desuden er beregningerne af de socioøkonomiske referencer for studenter i 2020, 2021, 2022 og 2023 foretaget i en begrænset version med færre fag.

Læs mere her: [Tekniske bemærkninger til socioøkonomiske referencer under Covid-19](#)

Afdelingsniveau

Alle beregninger er foretaget på individniveau, enkeltvis for hver af de gymnasiale uddannelser, men opgøres udelukkende for uddannelsesstedet. Til og med bevisår 2016 er beregningerne lavet på hovedinstitutionsniveau. Fra og med 2017 er beregningerne i stedet lavet på afdelingsniveau for stx, hhx og htx, mens det for hf-uddannelser gælder fra og med 2018.

Resultater vises ikke for uddannelsessteder med færre end 5 elever med et eksamensbevis det givne bevisår.

¹⁰ [Regler for afvikling af eksaminer 2020](#)

¹¹ [Regler for afvikling af eksaminer 2021](#)

¹² [Regler for afvikling af eksaminer 2022](#)