



De socioøkonomiske referencer for gymnasiekarakterer

Baggrund

Den enkelte institutions eksamensresultat og eksamenskarakterer har sammenhæng med mange forskellige forhold. Der er både forhold, som institutionen har indflydelse på, og forhold som institutionen ikke har indflydelse på. Institutionens elevgrundlag i bred forstand ligger uden for institutionens indflydelse.

Mange undersøgelser har vist, at en væsentlig forklaring på de opnåede karakterer knytter sig til elevernes faglige niveau fra grundskolen målt ved afgangskaraktererne fra 9. klasse (Folkeskolens prøver (FP)) samt til elevernes socioøkonomiske baggrund. Den socioøkonomiske baggrund skal forstås som de sociale og økonomiske kår, som eleven 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 elevens adgangsvej til den gymnasiale uddannelse udviser også en sammenhæng.

Selv om elevgrundlaget i udgangspunktet ikke ligger inden for institutionens indflydelse, er det dog vigtigt at huske på, at institutionen 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 institutionernes karaktergennemsnit, som skyldes forskelle i elevgrundlaget. Herved opnås et mere nuanceret billede af institutionernes niveau målt ved eksamensresultatet og eksamenskaraktererne. Beregningerne skal helst tage højde for alle de forhold, som institutionen ikke har direkte indflydelse på, fx 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 fx 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 institutionens indflydelse.

Vurderingen af den enkelte institutions gennemsnit af eksamensresultaterne og eksamenskaraktererne set i forhold til elevernes faglige og sociale baggrund tilvejebringes ved, at det faktisk opnåede eksamensresultat sammenlignes med et statistisk beregnet udtryk for, hvad elever 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 institutionens elever kommer fra hjem med et bestemt uddannelsesniveau, eller at institutionens elever har en bestemt fordeling med hensyn til herkomst. Disse forhold er der taget højde for.

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

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

Den socioøkonomiske reference er et statistisk udtryk for, hvordan elever på landsplan med samme baggrundsforhold som institutionens elever har klaret prøverne. *Socioøkonomisk* refererer til elevernes sociale og økonomiske baggrund, mens *reference* fortæller, at tallet kan bruges som et sammenligningsgrundlag for institutionens faktisk opnåede eksamensresultater og -karakterer

Både institutioner 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 en institution med et højt eksamensgennemsnit ligger lavere end dens socioøkonomiske reference, vil det typisk være en institution, som har overvejende 'ressourcestærke' elever, og hvor de statistiske modeller forudsiger, at institutionen burde opnå et endnu højere gennemsnit, end den faktisk gør.

Tilsvarende kan institutioner med eksamensresultatgennemsnit i den lave ende ligge højere end deres socioøkonomiske reference, når der er taget højde for elevsammensætningen. Dette kan fx være institutioner, hvor elevernes karaktergennemsnit fra 9. klasse er lavt og/eller institutioner, 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 institutioner med et højt gennemsnit også ligge højere end deres socioøkonomiske reference, og institutioner med et lavt gennemsnit kan på samme måde også ligge lavere.

Når en institutions eksamensresultater vurderes i forhold til dens socioøkonomiske reference, er det vigtigt, at der også tages højde for den tilfældige variation, der er mellem elevernes karakterer, således at der ikke konkluderes på et resultat, som lige så godt kunne være opstået ved en tilfældighed. Institutionens størrelse, målt ved antallet af elever med en eksamenskarakter, har en væsentlig betydning for, hvor sikkert den socioøkonomiske reference er bestemt. Færre elever giver mere usikre resultater. Derfor beregnes et sikkerhedsinterval¹, således at det er muligt at afgøre, hvorvidt en forskel mellem institutionens eksamensresultatgennemsnit og

¹ Der anvendes et 95 % sikkerhedsinterval.

den beregnede socioøkonomiske reference er reel, eller om den blot kan være et udtryk for tilfældig variation.

Dette notat beskriver resultaterne af disse statistiske analyser og de overordnede resultater på landsplan. Herudover beskrives det anvendte datagrundlag.

Der er beregnet socioøkonomiske referencer for såvel den seneste prøvetidspunkt som for den 3-årige periode, der dækker det aktuelle år samt de to foregående. Ved at anvende en 3-årig periode undgås en stor del af år til år variationen, der naturligt er på en institution.

Socioøkonomiske baggrundsvariable

Beregningen af de socioøkonomiske referencer er baseret på oplysninger på individniveau. Karakterdataene er baseret på indberetninger til STIL, mens alle oplysningerne vedrørende elevens baggrundsforhold er hentet fra registre hos Danmarks Statistik. Baggrundsoplysningerne er:

- Elevens 9. klasse FP karakterer (gennemsnit af de bundne prøvefag)
- Elevens køn
- Elevens alder
- Elevens herkomst
- Forældrenes højeste fuldførte uddannelse
- Forældrenes bruttoindkomst
- Familiestatus
- Elevens adgangsvej inden start på uddannelsen

En detaljeret gennemgang af baggrundsoplysningerne og den valgte kodning ses i bilag 1 i bilagsnotatet.

Institutionsniveau

Alle beregninger er foretaget på individniveau, enkeltvis for hver af de gymnasiale uddannelser, men opgøres udelukkende på institutionsniveau. Til og med bevisår 2016 er beregningerne foretaget på hovedinstitutionsniveau, mens det fra og med 2017 er på afdelingsniveau.

Resultater vises ikke for institutioner med færre end 5 elever med en eksamenskarakter pr. år.

Den statistiske model

Den anvendte statistiske model er en såkaldt multilevel model:

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

hvor y_{ij} er karakteren for elev i på institution j , x_{ij} er elevens baggrundsvariable og β er de tilhørende parameterestimater, u_j er varianskomponenten svarende til variationen mellem institutionerne, og e_{ij} er residualet svarende til variationen mellem eleverne på institutionen. 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 eleverne, og hvor stor en del der skyldes forskelle mellem institutionerne. Dette kaldes modellens forklaringsgrad.

Alle de socioøkonomiske baggrundsvariable beskrevet i bilag 1 (Bilagstabel 1) medtages i modellen.

Desuden er betydningen af interaktionen mellem forældrenes uddannelse og elevens herkomst samt betydningen af diverse institutions variable undersøgt. De undersøgte institutions variable er: andelen af indvandrere og efterkommere på institutionen, andelen af elever med forældre hvis højeste fuldførte uddannelse mindst er en mellemlang videregående uddannelse samt andelen af elever med forældre, der har en lav gennemsnitlig bruttoindkomst. Interaktionen medtages for, at undersøge om effekten af forældrenes uddannelse afhænger af elevens herkomst. Institutionsniveau variablene medtages for at undersøge, hvorvidt der er en 'klassekamerat effekt', dvs. om der er en øget effekt af, at der er en høj/lav andel af elever med en bestemt baggrund. Hverken interaktions variablen eller institutions variablene er statistisk signifikante i den multiple model og medtages derfor ikke i den endelige beregning af de socioøkonomiske referencer. Når der således er taget højde for elevernes individuelle baggrundsoplysninger, kan der ikke i disse modeller konstateres en øget statistisk signifikant effekt af, at der er en høj/lav andel af elever med en bestemt baggrund på den enkelte institution.

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.

Ikke alle elever 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 så tilføjes en indikatorvariabel, der fortæller hvorvidt, der er tale om en uoplyst værdi. Omfanget af ukendt oplysning er dog ikke uafhængig af fx herkomst. Oplysningsgraden for alle baggrundsvariablene er størst for elever af dansk herkomst. Der er således ikke tale om tilfældige 'huller' i informationen. Det er fx specielt udtalt blandt indvandrere, at informationen om forældrenes uddannelse mangler. Da uddannelsesniveaet blandt de indvandrere og efterkommere, hvor uddannelsesoplysningerne kendes, er en del anderledes end blandt danskere, 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 elev med ukendt baggrundsoplysning gættes 5 gange, således at modellen bliver estimeret på baggrund af en sandsynlig fordeling. Metoden kaldes 'multiple imputation'. Gættene baseres på fordelingen blandt elever med oplyst baggrundsvariabel, hvor der er taget hensyn til elevens køn, alder og herkomst. Den socioøkonomiske reference for hver elev er så et gennemsnit af de 5 beregnede værdier.

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

På institutionsniveau er den estimerede forskel mellem karaktergennemsnittet og den socioøkonomiske reference i en ordinær regressionsmodel gennemsnittet af disse elevresidualer på den enkelte institution, \bar{r}_j . I de her anvendte to-niveau modeller, hvor vi både har variation mellem institutioner ('Between'), og variation mellem elever indenfor institutionen ('Within'), bliver residualerne på institutionsniveau:

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

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

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

B er variationen mellem institutioner ('Between'), W er residualvariationen mellem elever indenfor institutionen ('Within') og n_j er antallet af elever på den j 'te institution.

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

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

Bilag 1: Socioøkonomiske baggrundsoplysninger

Bilagstabel 1 Baggrundsoplysninger

Baggrundsoplysning	Valgt kodning
9. klasse FSA karaktergennemsnit	Gennemsnit af bundne prøver (ekskl. dansk or-den) ² .
Køn	Dreng Pige
Alder på tidspunkt for afslutning af uddannelsen	19 år eller yngre 20 år 21 eller ældre
Herkomst³	Dansk herkomst Indvandrere Efterkommere
Forældrenes højeste fuldførte uddannelse⁴	Grundskolen Gymnasial uddannelse Erhvervsfaglig uddannelse Kort videregående uddannelse Mellemlang videregående uddannelse Lang videregående uddannelse (inkl. ph.d. mv.)
Forældrenes gennemsnitlige bruttoindkomst⁵	- 300.000 kr. 300 - 400.000 kr. 400 - 500.000 kr. 500.000+ kr.

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

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

Efterkommere er født i Danmark. Ingen af forældrene er danske statsborgere, født i Danmark. Når en eller begge forældre, der er født i Danmark, opnår dansk statsborgerskab, vil deres børn ikke 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, opfattes personen også som efterkommer.

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

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

⁵ Bruttoindkomsten er i året for start på uddannelsen.

Bilagstabel 1 Baggrundsoplysninger (fortsat)

Baggrundsoplysning	Valgt kodning
Familiestatus⁶	Hjemmeboende hos par Hjemmeboende hos enlig Udeboende
Adgangsvej⁷	Folkeskole Friskole og privat grundskole Efterskole Anden
10. klasse⁸	Ja Nej

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

⁸ Gennemført 10. klasse.