



Metodenotat: De socioøkonomiske referencer for grundskolekarakterer

Senest opdateret: december 2022

Indhold

Baggrund	1
Socioøkonomiske baggrundsvariable	3
Datagrundlag	4
Den statistiske model	5
Socioøkonomisk reference på kommuneniveau	7
Bilag 1: Socioøkonomiske baggrundsoplysninger	8

Baggrund

Den enkelte skoles faktiske karaktergennemsnit i 9. klasse har sammenhæng med mange forskellige forhold. Der er både forhold, som skolen har indflydelse på, og forhold, som skolen ikke har indflydelse på. Skolens elevgrundlag i bred forstand ligger uden for skolens indflydelse.

Mange undersøgelser har vist, at en væsentlig forklaring på de opnåede karakterer knytter sig 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 har også indflydelse på karaktererne.

Selv om elevgrundlaget i udgangspunktet ikke ligger inden for skolens indflydelse, er det dog vigtigt at huske på, at skolen 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 at reducere de forskelle i skolernes karaktergennemsnit, som skyldes forskelle i skolernes elevgrundlag. Herved opnås et mere nuanceret billede af skolernes niveau målt ved karaktergennemsnittet. Beregningerne skal helst tage højde for alle de forhold, som skolen ikke har indflydelse på, fx 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 skolens indflydelse.

Vurderingen af den enkelte skoles karaktergennemsnit set i forhold til elevernes sociale baggrund tilvejebringes ved, at det *faktisk* opnåede karaktergennemsnit sammenlignes med et statistisk beregnet udtryk for, hvad elever på landsplan med lignende baggrundsforhold forudsiges at opnå. Dette tal defineres som en socioøkonomisk reference.

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

Den socioøkonomiske reference er et statistisk udtryk for, hvordan elever på landsplan med samme baggrundsforhold som skolens elever har klaret 9.-klasseprøverne. *Socioøkonomisk* refererer til elevernes sociale og økonomiske baggrund, mens *reference* fortæller, at tallet kan bruges som et sammenligningsgrundlag for skolens faktiske opnåede karakterer.

Når en skole 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. Skolens størrelse, målt ved antallet af elever med en prøvekarakter, 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 skolens faktiske karaktergennemsnit og den beregnede socioøkonomiske reference er reel, eller om den blot kan være et udtryk for tilfældig variation.

Der beregnes socioøkonomiske referencer for såvel det seneste skoleår 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.

¹ Der anvendes et 95%-sikkerhedsinterval. Se nærmere under den statistiske model.

Socioøkonomiske baggrundsvARIABLE

Beregningen af de socioøkonomiske referencer er baseret på oplysninger på individniveau. Oplysningerne hentes fra Danmarks Statistik.

Baggrundsoplysningerne er:

- Elevens køn
- Elevens alder
- Elevens herkomst og oprindelsesland
- Forældrenes højeste fuldførte uddannelse
- Forældrenes arbejdsmarkedsstatus
- Forældrenes bruttoindkomst
- Forældrenes ledighedsgrad
- Familietype
- Antal børn og placering i børneflokk

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

Datagrundlag

Alle beregninger er foretaget på individniveau, men opgøres udelukkende på institutionsniveau. Til og med skoleåret 2015/16 blev den socioøkonomiske reference beregnet pr. hovedinstitution, men fra skoleåret 2016/17 beregnes denne på afdelingsniveau.

Beregningen af de socioøkonomiske referencer er baseret på følgende elever:

- Elever på skoler af typen folkeskoler, frie og private grundskoler og efterskoler. Elever på efterskoler med samlet særligt tilbud indgår ikke.
- Elever i specialklasser indgår ikke.
- Elever, der har aflagt 9.-klasseprøver som privatister, indgår ikke.

Resultater vises ikke for institutioner med færre end 5 elever med en eksamenskarakter pr. år. Beregningen af de socioøkonomiske referencer er foretaget for såvel de bundne prøvefag som for prøvefag til udtræk². Derudover beregnes også en socioøkonomisk reference for elevernes samlede ikke-vægtede gennemsnit af de 8 bundne prøvediscipliner. Kun elever, der har aflagt minimum 4 af disse prøver, får beregnet et gennemsnit.

Den socioøkonomiske reference beregnes én gang årligt umiddelbart efter afslutning af skolernes karakterindberetning af sommerens prøver i september. De 1-årige socioøkonomiske referencer er således baseret på indberettede karakterer fra ordinære prøver og indeholder ikke karakterer fra senere afholdte sygeprøver.

Bemærk

Beregningen af socioøkonomiske referencer for skoleårene 2019/2020, 2020/2021 og 2021/2022 er foretaget i en begrænset version. Dette skyldes de særlige regler for prøveafvikling og karaktergivning grundet enten fuld eller delvis aflysning af prøverne i sommeren 2020, 2021 og 2022.

I disse tre skoleår er der kun foretaget 1-årige beregninger af socioøkonomiske referencer for det samlede gennemsnit i bundne prøvefag. Resultaterne for disse skoleår omfatter derfor ikke:

- 1-årige socioøkonomiske referencer for de enkelte prøver
- 3-årige socioøkonomiske referencer

² Eksklusive fransk, hvor relativt få elever aflægger prøver.

Den statistiske model

Den anvendte statistiske model er en såkaldt multilevel-model:

$$y_{ij} = x_{ij}\beta + u_j + e_{ij} \quad (1)$$

hvor y_{ij} er karakterresultatet for elev i på institution j , x_{ij} er elevens værdier på baggrundsvariable, og β er de tilhørende parameterestimer, u_j er varianskomponenten svarende til variationen mellem skolerne, og e_{ij} er residualet svarende til variationen mellem eleverne på skolen.

Beregningerne er foretaget for såvel de bundne prøfefag som for prøfefag til udtræk (eksklusive fransk, hvor relativt få elever aflægger prøve).

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

Alle de socioøkonomiske baggrundsvariable beskrevet i Bilag 1: Baggrundsoplysninger medtages i modellen sammen med interaktionen mellem forældrenes uddannelse og elevens herkomst, interaktionen mellem elevens køn og herkomst samt andelen på skoleniveau af hhv. indvandrere og efterkommere og andelen af elever med forældre med en mellemlang eller lang videregående uddannelse.

Interaktionerne medtages, for at undersøge om effekten af forældrenes uddannelse og elevens køn afhænger af elevens herkomst.

Skoleniveauvariablen vedrørende andelen af indvandrere og efterkommere medtages, for at undersøge om der er en ekstra effekt af en høj andel af indvandrere og efterkommere på samme institution. Tilsvarende for andelen af elever med forældre, der har en mellemlang eller lang videregående uddannelse.

Øvrige interaktioner og skoleniveauvariable er ikke medtaget, dels for ikke at øge modellernes kompleksitet, dels viser undersøgelser, at disse ikke øger modellernes forklaringskraft væsentligt.

Ikke alle elever har fyldestgørende oplysninger på de benyttede baggrundsvariable. Der er i disse tilfælde indsat 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 multipel 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.

I den statistiske model for beregning af socioøkonomiske referencer for 3-årsperioden medtages endvidere skoleåret som en forklarende variabel. Herved tages der højde for, at karakterniveauet kan være forskelligt fra år til år.

For hver elev beregnes en socioøkonomisk referenceværdi 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 skoleniveau er den estimerede forskel mellem karaktergennemsnittet og den socioøkonomiske reference i en ordinær regressionsmodel gennemsnittet af disse elevresidualer på den enkelte skole, $\bar{r}_j$. I de her anvendte to-niveau-modeller, hvor vi både har variation mellem skoler ('Between'), og variation mellem elever inden for skolen ('Within'), bliver residualerne på skoleniveau:

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

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

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

hvor B er variationen mellem skoler ('Between'), W er residualvariationen mellem elever inden for skolen ('Within'), og n_j er antallet af elever på den j 'te skole.

For forskellen mellem opnået karakter og den socioøkonomiske reference for hver skole er der endvidere udregnet et 95%-sikkerhedsinterval. Hvis dette sikkerhedsinterval ligger over 0, opnår skolens 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 skole et karaktergennemsnit på niveau med 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 skole. Således bliver sikkerhedsintervallet mindre for den enkelte skole, når en 3-årig periode betragtes, idet der her indgår elever for 3 år i beregningerne.

Socioøkonomisk reference på kommuneniveau

Opgørelserne på kommuneniveau foretages for gennemsnittet af elevernes karakterer i de bundne prøvfag på baggrund af beregningerne af den socioøkonomiske reference på institutionsniveau. Opgørelsen på kommuneniveau indeholder udelukkende resultaterne fra folkeskolerne.

På kommuneniveau kan forskellen mellem elevernes gennemsnitskarakterresultat og den estimerede socioøkonomiske reference beregnes som det vægtede gennemsnit af de beregnede forskelle på institutionsniveau:

$$(\sum_1^m n_i u_i)/n \quad (2)$$

hvor u_i er den estimerede forskel, n_i er antallet af elever med et karaktergennemsnit af de bundne prøver på i 'te institution, og m er antallet af institutioner i den enkelte kommune. n er antallet af elever i alt med et karaktergennemsnit af de bundne prøver på kommunens folkeskoler.

Tilsvarende beregnes variansen på kommuneniveau³:

$$Var(\sum_1^m n_i u_i/n) = \sum_1^m n_i^2 Var(u_i)/n^2 \quad (3)$$

hvor $Var(u_i)$ er variationen på den estimerede forskel på i 'te institution.

Da residualerne på institutionsniveau, u_i , i model (1) antages at være normalfordelte, kan der efterfølgende beregnes et 95%-sikkerhedsinterval. Hvis dette sikkerhedsinterval ligger over 0, opnår kommunens folkeskoleelever samlet 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 i den pågældende kommune samlet 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 og den enkelte kommune. Således bliver sikkerhedsintervallerne mindre for store kommuner, hvorfor relativt små forskelle mellem karaktergennemsnit og den socioøkonomiske reference kan blive statistisk signifikante.

³ Under antagelse af uafhængighed mellem institutionerne.

Bilag 1: Socioøkonomiske baggrundsoplysninger

Baggrundsoplysning	Valgt kodning
Køn	Dreng Pige
Alder	15 år eller yngre 15-16 år 16-17 år Ældre end 17 år
Herkomst¹ og oprindelsesland²	Dansk herkomst Indvandrere fra Vestlige lande Indvandrere fra Ikke-vestlige lande Efterkommere fra Vestlige lande Efterkommere fra Ikke-vestlige lande
Forældrenes højeste fuldførte uddannelse³	Grundskolen Gymnasial uddannelse Erhvervsfaglig uddannelse Kort videregående uddannelse Mellemlang videregående uddannelse Lang videregående uddannelse (inkl. ph.d. mv.)
Faderens arbejdsmarkedsstatus⁴	Beskæftiget Ledig Øvrige uden for arbejdsmarkedet
Moderens arbejdsmarkedsstatus	(som faderens arbejdsmarkedsstatus)
Forældrenes gennemsnitlige bruttoindkomst⁵	1. kvartil 2. kvartil 3. kvartil 4. kvartil
Faderens ledighedsgrad⁶	0 % 1 – 50 % 50+ %
Moderens ledighedsgrad	(som faderens ledighedsgrad)
Familietype⁷	Par Enlig Ikke hjemmeboende børn

Baggrundsoplysning	Valgt kodning
Antal børn og placering i børneflokkene⁸	Enebarn Ældste barn i børneflok Midterbarn i børneflok på højst 4 børn Yngste barn i børneflok på højst 4 børn Øvrige

Noter:

- 1) 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, hvor 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.
- 2) Kodningen 'Indvandrere/efterkommere fra Ikke-vestlige lande' indeholder også de meget få, som har uoplyst oprindelsesland, er statsløse m.m.
- 3) Den højest fuldførte uddannelse blandt elevens mor og far målt i det aktuelle år, medmindre eleven er ældre end 13 år. Da indgår oplysningen fra det år, eleven fyldte 13 år.
- 4) Arbejdsmarkedsstatus er et 'øjebliksbillede' målt ultimo november det aktuelle år, med mindre eleven er ældre end 13 år. Da indgår oplysningen fra det år, eleven fyldte 13 år.
- 5) Den valgte kodning er baseret på medianen og nedre og øvre kvartil af bruttoindkomsten blandt forældrene i elevpopulationen, som de socioøkonomiske referencer beregnes på baggrund af. Bruttoindkomsten er fra det aktuelle år, medmindre eleven er ældre end 13 år. Da indgår oplysningen fra det år, eleven fyldte 13 år.
- 6) Grad af ledighed i året beregnes som personens timer i tilstanden (hverdage) i forhold til årets referencetimer, som er antal hverdage i året ganget med 7,4. En ledighedsgrad på 0 % svarer til at personen ikke har været ledig i året. En ledighedsgrad på 50+ % svarer til, at personen har været ledig mellem halvdelen og hele året.
- 7) Familietypen par indeholder ægtepar, registreret partnerskab, samlevende og samboende par.
- 8) Antal børn og placering i børneflokkene er sammensat af to variable: 'Antal børn i familien' og 'Nummer i søskendeflokken'. Antal børn i familien dækker antal hjemmeboende børn under 18 år, som er børn af mindst én af de op til to voksne i familien.