



RIKSREVISIONEN

## Bilaga till granskningsrapport

Datum: 2022-11-02

Diarienummer: 3.1.1-2021-0773

RiR 2022:23

# Bilaga 2

## Att operationalisera (bristande) likvärdighet mellan skolor



Skolverkets  
statistikuppdrag  
– information, uppföljning och  
utvärdering

## Att operationalisera (bristande) likvärdighet mellan skolor

Den övergripande bedömningsgrunden för granskningen är skollagens bestämmelser om alla elevers rätt att få tillgång till en likvärdig utbildning. Till saken hör att det inte är självklart att bestämma hur likvärdiga skolor är då skollagen inte ger någon tydlig vägledning.

I regel görs det operationaliseringar i form av ”olika indikatorer som belyser skilda aspekter av likvärdighetsbegreppet.”<sup>1</sup> Den mest vanliga operationalisering är att översätta likvärdighet med socioekonomiska förutsättningar. Skolverket gör det i sina egna uppföljningar, men det är också vanligt förekommande i internationella sammanhang, till exempel i jämförelsen av skolresultat från internationella kunskapsmätningar i Pisa eller Timms.<sup>2</sup> Vi har valt att tillämpa samma slags operationalisering och ser till (skillnader i) socioekonomiska förhållanden bland skoleleverna som ett godtagbart tillvägagångssätt för att uppskatta graden av likvärdighet mellan skolor.

På skolområdet har det sedan länge funnits en stor grad av samsyn om vad som utgör relevanta socioekonomiska faktorer. Det kan exemplifieras med följande litteraturgenomgång på vad som utgör socioekonomisk status (inom den engelskspråkiga litteraturen förkortat som SES): “Traditionally, a student’s SES has included, as components, parental educational attainment, parental occupational status, and household or family income, with appropriate adjustment for household or family composition. [...] Education, occupation and income are sometimes referred to as the “big three” (Willms and Tramonte 2019); Sirin (2005, p. 418) notes that there is substantial agreement among researchers on the ‘tripartite nature of SES that incorporates parental income, parental education, and parental occupation as the three main indicators of SES’. Ensminger and Fothergill (2003) note that ‘education, income and occupation’ are the ‘three most common measures of SES’.”<sup>3</sup>

Också bland (svenska) nationalekonomer som har analyserat utbildningsområdet hittar man jämförbara kategoriseringar av socioekonomi, se till exempel Brandén med flera (2018) som använder sig av ett index bestående av föräldrars utbildning,

---

<sup>1</sup> *Samling för skolan-nationell strategi för kunskap och likvärdighet*, SOU 2017:35, s.92.

<sup>2</sup> Se Skolverket, PISA 2015 15-åringars kunskaper i naturvetenskap, läsförståelse och matematik, 2016.

<sup>3</sup> Se Avisati, “The measure of socio-economic status in PISA: a review and some suggested improvements”, 2020.

inkomst och yrke för att uppskatta socioekonomisk nivå.<sup>4</sup> Ett annat exempel hittar man i Björklund med flera (2010).<sup>5</sup>

Också i andra sammanhang har de ovan nämnda faktorerna pekats ut som de mest relevanta för att kartlägga socioekonomiska förhållanden. Delegationen mot segregation (Delmos) och SCB redovisade 2021 ett regeringsuppdrag ”att utveckla ett rikstäckande index som belyser hur olika områden förhåller sig till varandra avseende socioekonomisk status”. De kom fram till att indexet skulle innehålla mått på inkomst, utbildning och arbetsmarknadsposition.<sup>6</sup>

I Skolverkets egen uppföljning av familjebakgrundens betydelse för skolresultat från 2018 inkluderas vid sidan om föräldrarnas utbildningsnivå också föräldrarnas bidragsberoende och inkomst.<sup>7</sup> Valet av variabler motiveras med att de ”har sin grund i forskning om vad som har störst påverkan på skolresultaten”. När det gäller sambanden och resultaten noteras att ”[a]v de socioekonomiska bakgrundsfaktorerna är det föräldrarnas utbildningsnivå som har den största betydelsen för betygsresultaten, men utbildningsnivåns betydelse har inte ökat över tid. Däremot har föräldrarnas inkomst blivit allt viktigare, och framstår som den centrala förklaringen till den ökade betydelsen av socioekonomisk bakgrund för betygsresultaten. För utlandsfödda elever har även föräldrarnas utbildningsnivå och grad av bidragstagande bidragit till den ökande betydelsen av socioekonomisk bakgrund för skolresultaten.”<sup>8</sup>

## Att bestämma en modells prediktionsförmåga

Enligt vedertagen praxis inom utvärderingslitteraturen bör en statistisk modell i första hand försöka förhålla sig till de aspekter som enligt teoretiska överväganden eller utifrån beprövad erfarenhet påverkar det studerade utfallet. Eller med andra ord: I en modellskattning bör man först och främst försöka ha med relevanta faktorer utifrån vetenskap om samspelet mellan olika faktorer samt modellens syfte och tänkta användning, dock inte (åtminstone inte enbart) faktorernas inverkan på förklaringsgrad i en regressionsmodell. Det är dock det senare som Skolverket framför i sina redogörelser om Salsa-modellens träffsäkerhet. Att förklaringsgraden i en regressionsmodell är hög kan i det här fallet kopplas till ett ömsesidigt beroende av variablerna som ingår i höger- och vänsterledet. Det är väl

<sup>4</sup> Brandén, Lindahl och Öckert, “The Importance of Nature–Nurture Interactions for Socioeconomic Outcomes: Evidence from a Large Sample of Adoptees”, 2018.

<sup>5</sup> Björklund, Lindahl, och Lindquist, “What More Than Parental Income, Education and Occupation? An Exploration of What Swedish Siblings Get from Their Parents”, 2010.

<sup>6</sup> Se *Segregation i Sverige – ÅRSRAPPORT 2021 OM DEN SOCIOEKONOMISKA BOENDESEGREGATIONENS UTVECKLING*.

<sup>7</sup> Se Skolverket, Analyser av familjebakgrundens betydelse för skolresultaten och skillnader mellan skolor, 2018.

<sup>8</sup> Se Skolverket, Analyser av familjebakgrundens betydelse för skolresultaten och skillnader mellan skolor, 2018.

känt att föräldrar med högre utbildning är mer benägna att välja skolor till sina barn, rimligen skolor som kan visa på högre meritvärden. Samtidigt bidrar bättre utbildade föräldrar till att deras barn uppnår bra resultat i skolan, till exempel genom att de på ett bättre sätt kan hjälpa till med läsläsningen. Detta innebär att sambandet mellan föräldrarnas utbildningsbakgrund och skolornas resultat (så som skolornas meritvärde) är resultatet av en växelverkande process. Sådan ömsesidighet medför att korrelationen mellan variabler som skolornas meritvärde och föräldrarnas utbildningsnivå är hög, vilket då också medför en hög förklaringsgrad i form av ett högt R<sup>2</sup>-värde i en regressionsmodell. En regressionsmodell allena tillåter inte att urskilja vad som påverkar vad när denna slags ömsesidighet råder, och därmed bör modellens ”förklaringsgrad” inte tolkas som att föräldrarnas utbildningsnivå förklarar skolornas resultat, så som det antyds på Skolverkets webbplats.<sup>9</sup> Själva ordet förklaringsgrad avser snarare ett statistiskt förhållande om hur mycket av variansen i den beroende variabeln (t.ex. variation i meritvärden mellan skolorna) som fångas upp av de inkluderande kontrollvariablerna i modellen. Föräldrarnas utbildningsbakgrund är högt korrelerad med föräldrarnas inkomst och arbetsmarknadsposition, varför dessa två faktorer inte ökar R<sup>2</sup>-värdet särskilt mycket då de inkluderas i regressionsmodellen vid sidan om föräldrarnas utbildningsbakgrund. Det betyder dock inte att dessa faktorer inte skulle kunna ha betydelse för skolornas predikterade utfall som tas fram utifrån regressionsmodellen. Eftersom Salsa-utfall bygger på en prediktion av skolornas resultat (meritvärde respektive andel elever som uppnår målen) med hjälp av koefficientestimat från regressionsmodellen missar en ensidig fokus på R<sup>2</sup> en väsentlig aspekt av modellens funktion.<sup>10</sup>

I två PM redovisar Skolverket uppskattningar av prediktionsförmåga för fördelningsnyckeln till grundskolan.<sup>11</sup> Detta görs genom att redovisa hur ofta de statistiska prediktionerna leder till korrekta skattningar av huruvida elever lyckas bli behörig till gymnasieskolan (närmare bestämt till nationellt program) eller ej. I PM:en anges att prediktionen blir rätt i 9 fall av 10. Vi har förbehåll mot denna form av redovisning av modellernas prediktionsförmåga. För det första tycks det som att testen utfördes på samma dataunderlag som själva modellen är skattad på (det framgår inte tydligt ur redovisningen). Ett vederhäftigt test för att fastställa en modells prediktionsförmåga bör utgå utifrån ett annat statistikunderlag än det

<sup>9</sup> Se SALSAs statistiska modell - Skolverket (2022-04-01) För att kunna fastställa ett kausalt samband skulle det behövas en situation där barn fördelas slumpmässigt till olika skolor och man i efterhand mäter hur elevernas skolresultat korrelerar med föräldrarnas utbildningsbakgrund.

<sup>10</sup> För en fördjupad diskussion kring skillnaden mellan en regressionsmodells prediktionsförmåga och dess förklaringsgrad (R<sup>2</sup>) se Shmueli, ”To Explain or to Predict”, 2010.

<sup>11</sup> Se Skolverket, *Redovisning av regeringsuppdraget Uppdrag om att ta fram förslag inför fördelning av medel*, dnr 2015:00898, 2015, samt PM\_Fördelningsnyckeln\_2017\_grundskolan, internt PM, mottaget via epost 2021-11-03.

som modellens parameterestimat är skattade på.<sup>12</sup> Vår andra kritik är att redovisningen av prediktionsförmågan inte förhåller sig till andelen elever som rent faktiskt blir behöriga; i stället görs ett implicit antagande att det skulle vara en 50%-chans att en elev blir behörig till gymnasieskolan.<sup>13</sup> I själva verket var det dock mellan 80 och 90 procent (variation utifrån årsmedelvärden) av eleverna som blev behöriga under den studerade perioden. Det innebär att även en till sin konstruktion innehållslös regressionsmodell (en modell endast innehållande en konstant) skulle kunna prediktera rätt i nio av tio fall enbart genom att reproducera medelvärdet.<sup>14</sup> Det ska här nämnas att liknande kritik kan riktas mot många andra studier som redovisar en statistikmodells prediktionsförmåga, däribland den som görs i SOU 2017:35 som beskriver det socioekonomiska indexets prediktionsförmåga. När det gäller redovisningen av fördelningsnyckeln till gymnasieskolan redovisas där inte motsvarande prediktionstillförlitlighet.<sup>15</sup> Rent generellt bedömer vi att fördelningsnyckeln till gymnasieskolan i ett väsentligt avseende skiljer sig från fördelningsnyckeln till grundskolan i det att den inkluderar gymnasieelevernas respektive meritvärde från grundskolan, samt en uppdelning över programtyp på gymnasiet som kontrollvariabler. Programtyp avser endera introduktionsprogram, yrkesprogram respektive högskoleförberedande program. Uppgifter om meritvärden från grundskolan samt programtyp är rimligen starkt korrelerad med socioekonomiska faktorer i den utsträckning som elevens socioekonomiska bakgrund har betydelse för skolresultat i grundskolan respektive program som eleven sökt till. Modellen är därmed till sin konstruktion väsensskild från fördelningsnyckeln för grundskolan. Vi ser ingen direkt anledning att ifrågasätta modelspecifikationen av fördelningsnyckeln till gymnasieskolan. Därutöver saknar vi möjlighet att stämma av modellen mot en given referens-modell, så som vi bedömer kunna göra när det gäller både Salsa och fördelningsnyckeln till grundskolan. Av den anledningen kommer vi inte adressera fördelningsnyckeln till gymnasieskolan något mer i den fortsatta redogörelsen.

<sup>12</sup> Så görs till exempel i IFAU 2007:20 som anges som referensstudie i *Redovisning av regeringsuppdraget Uppdrag om att ta fram förslag inför fördelning av medel*, s.5.

<sup>13</sup> Rent konkret görs det genom att elever som har prediktionsvärden mellan 0–0.5 tilldelas värdet 0, vilket betyder att de skattas ej uppnå behörighet; elever med värdena 0.5–1 tilldelas värdet 1, vilket betyder att de skattas uppnå behörighet.

<sup>14</sup> Detta förklaras av att alla ingående elever i en sådan modell skulle få samma predikterade sannolikhet, nämligen medelvärdet av alla ingående elevers andel att bli behörig. Om medelvärdet skulle vara 90% och man skulle sätta en gräns att alla elever som har en predikterad sannolikhet större än 50% ska klassas som att de klarar behörighet blir det rätt i 90% av fallen, och fel i 10% av fallen. Det betyder att en sådan naiv modell skulle prediktera rätt i 90% av fallen.

<sup>15</sup> Se PM Fördelningsnyckeln 2020\_gymnasieskolan, internt PM, mottaget via epost 2021-07-05.

## Hur väl lämpar sig Salsa och fördelningsnyckel till grundskolan för att fånga in socioekonomiska skillnader mellan skolor?

Vi har tittat närmare på modellskattningar för Salsa och fördelningsnyckeln för grundskolan. Dessa två modeller liknar varandra i bemärkelsen att den enda indikatorn som tydligt fångar in socioekonomiska aspekter är föräldrarnas utbildningsnivå. En aspekt som skiljer modellerna åt är, förutom deras tänkta användning, att föräldrarnas utbildningsnivå ingår som en enda variabel i Salsa-modellen. Med det menas att utbildningsstegen förgymnasial-, gymnasial-, respektive eftergymnasial-utbildning är kodad som en sifferföljd (1, 2, 3) i en enda utbildningsvariabel. Det här sättet att hantera utbildningsbakgrund i en regressionsmodell avviker från gängse praxis inom utvärderingslitteraturen. Där låter man i regel respektive utbildningssteg utgöra en egen (indikator-) variabel. Med en sådan flexibel ansats tillåter man att betydelsen av föräldrarnas utbildning kan skilja sig åt (dvs *hur* gymnasieutbildning står sig i relation till förgymnasial eller eftergymnasial utbildning i relation till utfallsvariabeln tvingas inte in i en på förhand fastställd ordning). I Skolverkets fördelningsnycklar till grundskolan används däremot en ansats som är i linje med praxis inom utvärderingslitteraturen.

Härnäst belyser vi vilka konsekvenser urvalet av inkluderade (socioekonomiska) faktorer och val av modellspecifisering har för klassificeringen av skolor. Till det ändamålet krävs en referens-modell som vi kan jämföra med. Vi har valt att använda oss av modellen som ligger till grund för SCB:s socioekonomiska index som utgångspunkt. Utifrån våra intervjuer med tjänstemän på SCB är vår bedömning att arbetet med att ta fram modellen kännetecknades av ett konstruktivt samarbete mellan SCB och företrädare för Skolkommissionen. I det arbetet samlades det också in synpunkter från externa sakkunniga. Vi tar det som intäkt för att SCB:s socioekonomiska index är noggrant utarbetat. Med det sagt vill vi understryka att det inte finns ett enda sätt att beskriva och kategorisera en elevs socioekonomiska situation. Olika viktningar och prioriteringar kan behöva göras utifrån modellens syfte eller också vilken information som är möjlig att fånga in i register i olika situationer.

Vi har fått tillgång till uppgifter om Salsa-värden och bakomliggande variabler för alla berörda skolenheter. När det gäller SCB:s socioekonomiska index och Skolverkets fördelningsnyckel för grundskolan har vi inte haft motsvarande tillgång till de (predefinierade) variabeluppgifterna. En anledning till det är att i Salsa-modellen används uppgifter på skolenhetsnivå, medan Skolverkets fördelningsnyckel bygger på uppgifter på elevnivå. Dock har vi haft tillgång till detaljerade beskrivningar om respektive modells konstruktion, utifrån vilka vi har

tagit fram jämförelseskattningar.<sup>16</sup> I det arbete valde vi att göra justeringar i ett antal avseenden när det gäller modellen som relaterar till det socioekonomiska indexet från SCB. För det första använder vi oss av utbildningsnivå av föräldrarna, inte vårdnadshavare, som kontrollvariabel. Anledningen till det är att öka jämförbarheten med Skolverkets fördelningsnyckel som utgår ifrån föräldrarnas utbildning, inte vårdnadshavarnas. Enligt Skolverkets mall är det den högsta utbildning för respektive förälder som används i fall det finns uppgifter om dem. När det gäller förhållanden i elevens omliggande bostadsområde har vi utgått ifrån DeSO-områden, inte SAMS-områden, med uppgifter om befolkningens sammansättning med avseende på andelen utrikesfödda (alla åldrar), utbildningsnivå (25–64 år) samt medianinkomst (alla åldrar). DeSO-områden lanserades 2018 av SCB och är den officiella geografiska uppdelningen av Sverige i syfte att kartlägga socioekonomiska förhållanden.<sup>17</sup> När det gäller utbildningsnivå har vi tagit andelen i DeSO-område som har tre-års eftergymnasial utbildning, inte genomsnittlig utbildningsnivå. Vi har inte heller genomfört en principalkomponentanalys av dessa tre faktorer, så som gjordes i beräkningarna till SCB:s socioekonomiska index. Den främsta anledningen till att vi inte gjorde det var att säkerställa transparensen i skattningarna, vilket vi bedömer blir tydligare med en modell som till sin specifikation är så enkel som möjligt. Det är också viktigt att hålla fram att SCB:s socioekonomiska index i sin ursprungliga form skattas som en logistisk regressions-skattning på elever i årskurs 9 med utfall om behörighet till gymnasiet, som därefter aggregeras till skolenhetsnivå. Salsa-skattningar däremot utgår ifrån uppgifter på skolenhetsnivå, utifrån två utfall: genomsnittliga meritvärden, respektive andelen elever som uppnått kunskapskraven i alla ämnen. Bägge två utfall redovisas på Skolverkets webbplats. För att kunna åstadkomma en jämförelse av modellen som ligger till grund för SCB:s socioekonomiska index med Salsa använder vi oss av OLS-regressionsskattningar med motsvarande variabeluppdelning som används i SCB:s socioekonomiska index, dock som genomsnittsvärden på skolenhetsnivå och med genomsnittliga meritvärden som beroende variabel. Det är därmed fråga om en jämförelse av Salsa och en statistisk modell som med avseende på inkluderade variabler följer SCB:s socioekonomiska index, men som inte är identiska med denna.

När det gäller jämförelsen av Skolverkets fördelningsnyckel till grundskolan skattas den på elevnivå med sannolikhet att uppnå behörighet till nationellt program i årskurs 9 som utfall. Det görs genom att utgå ifrån om eleverna uppnår yrkesgymnasiebehörighet, eller inte. De genom (logit-)skattningarna predikterade

---

<sup>16</sup> Vi använde oss av underlag och mejlsvar från Skolverket för att modellera fördelningsnyckeln till grundskolan samt den utförliga tekniska bilagan i SOU 2017:35 för modeller som bygger på SCB:s socioekonomiska index.

<sup>17</sup> Se SCB, *DeSO – Demografiska statistikområden* för förtydliganden.

värden, som är på elevnivå, aggregeras därefter till skolenhetsnivå. Motsvarande tillvägagångssätt görs utifrån variabelsuppgifterna som ingår i SCB:s socioekonomiska index för att på så sätt kunna jämföra modellens prediktion med Skolverkets fördelningsnyckel till grundskolan.

I följande avsnitt redovisas jämförelser av resultat kopplade till Skolverkets statistiska modeller med avseende på Salsa respektive fördelningsnyckeln för grundskolan med modeller som bygger på SCB:s socioekonomiska index. Frågan som adresseras är om det blir approximativt lika uppskattningar av skolornas resultat, respektive behov av stöd då man använder sig av Skolverkets modeller eller en modell som bygger på SCB:s beräkningsmodell av socioekonomiskt index (som vi i följande genomgång kommer att referera till som ”SCB-modellen”). Vi börjar med att redovisa likheter/skillnader då man ser till skolornas resultat utifrån Skolverkets Salsa-värden och SCB modellen. Alla redovisade siffror avser grundskolor med avseende på elever i årskurs 9 för läsåret 2018/19.<sup>18</sup>

**Redovisning av överensstämmelsen av Skolverkets modeller med avseende på socioekonomiska faktorer: betygsgenomsnitt i Salsa**

Tabellen nedan visar resultaten av skattningar utifrån Skolverkets Salsa-modell och SCB-modellen. Utfallsvariabeln är skolornas genomsnittliga meritvärde och i skattningarna ingår uppgifter som avser läsåret 2018/2019. Modellerna tar som utgångspunkt regressionsskattningar som predikterar ett förväntat värde för varje skola givet sina ”egenskaper” utifrån de ingående variablerna och deras respektive koefficientestimat. Nedan visas underliggande regressionsskattningar.<sup>19</sup> När Skolverket tar fram Salsa-estimat exkluderas skolor som saknar uppgifter om elevernas föräldrarnas utbildning för mer än 25% av eleverna i en skola, eller då utfall om skolresultat är extremt låga (andelen godkända elever mindre än 20% eller genomsnittligt meritvärde för skolan under 120). Dessutom exkluderas skolor som har (uppgifter för) mindre än 15 elever.<sup>20</sup> För att öka jämförbarheten har vi valt att använda motsvarande avgränsning i skattningar utifrån SCB-modellen.

<sup>18</sup> Här görs ingen extrapolering till elever i lägre årskurser, så som det görs i den praktiska tillämpningen av fördelningsnycklarna i bestämningen av storleken på riktade statliga medel till skolhuvudmännen.

<sup>19</sup> Vi konstaterar att det finns en del signifikanta men också icke-signifikant koefficientestimat, men det är inte något vi kommer gå in på närmare.

<sup>20</sup> Ifall en elev saknar en uppgift, t.ex. föräldrarnas utbildning, ingår denna inte i skattningarna.



**Tabell 1** Resultat av regressionskattningar för Skolverkets Salsa-modell och modell baserad på SCB SE index, läsår 2018/19. Utfallsvariabel skolornas genomsnittliga meritvärde.

	SCB baserad 2019 koefficient (standardfel)	Salsa_SV 2019 koefficient (standardfel)
Andel elever med föräldrar som har högst förgymnasial utbildning	-63.8568*** (15.302)	
Andel elever med föräldrar som har högst gymnasial utbildning	-70.7814*** (5.938)	
Andel elever med föräldrar som har eftergymnasial utbildning	referens	
Andel elever med föräldrar vars utbildningsnivå är okänd	-80.0134* (36.332)	
Andel elever med föräldrar som har mindre än 60% av medianinkomst	-44.6296*** (10.703)	
Andel elever med föräldrar som har 60–80% av medianinkomst	-34.2420** (12.048)	
Andel elever med föräldrar som har 80–100% av medianinkomst	-18.4737 (10.785)	
Andel elever med föräldrar som har 100–120% av medianinkomst	-36.7055*** (9.703)	
Andel elever med föräldrar som har 120–150% av medianinkomst	-48.3903*** (10.748)	
Andel elever med föräldrar som har mer än 150% av medianinkomst	referens	
Andel tjejer i skolan	24.9247*** (6.508)	
Andel elever som bor med båda föräldrar (ja/nej)	16.7561* (6.839)	
Andel elever med antal syskon lika med 0	-98.5877* (49.563)	
Andel elever med antal syskon lika med 1	referens	
Andel elever med antal syskon lika med 2	-6.4540 (8.586)	
Andel elever med antal syskon lika med 3	-6.6093 (8.832)	

	SCB baserad 2019 koefficient (standardfel)	Salsa_SV 2019 koefficient (standardfel)
Andel elever med antal syskon lika med 4	-6.6236 (13.406)	
Andel elever med antal syskon lika med 5 eller fler	2.6929 (16.021)	
Andel elever med i hushåll som fått ekonomiskt bistånd (ja/nej)	-38.8344** (12.082)	
Andel elever som bort längre än sex år i Sverige	referens	
Andel elever invandrade de senaste tre åren	-86.3017*** (14.722)	
Andel elever invandrade de senaste fyra till sex åren	-41.0529** (14.120)	
Andel högskoleutbildade i elevens DeSO	-3.4998 (8.663)	
Andel utlandsfödda i elevens DeSO	37.6692*** (6.354)	
Medianinkomst i elevens DeSO	0.0731* (0.029)	
Andel nyinvandrade i skolan		-0.8749*** (0.100)
Föräldrarnas genomsnittliga utbildning i skolan		73.8606*** (2.829)
Andel pojkar i skolan		-0.1997** (0.068)
Konstant	254.5516*** (15.838)	75.2257*** (8.391)
R-kvadrat	0.685	0.624
R-kvadrat-just	0.680	0.623
Antal skolor	1462	1462

Anm.: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Utifrån de skattade koefficientestimaten och skolornas medelvärden för respektive variabel räknas det fram förväntade värden för alla skolor som ingår i skattningarna.<sup>21</sup> Därefter beräknas differensen mellan de faktiska och de förväntade värden. Om en skola är bättre än det förväntade värdet blir differensen

<sup>21</sup> Med begreppet förväntade värden avses här en statistik storlek, inte någon bedömning av vilka krav som ska ställas på en skola utifrån skollagens intentioner.

positivt (en positiv residual) och vice versa. Dessa residual-värden rangordnas därefter för alla grundskolor som ingår. Vi har valt att dela in skolorna utifrån fem intervaller i stigande ordning. De 20 procent av skolorna som har lägst residual-värden samlas i gruppen "sämst", de efterföljande 20 procent blir klassade som "under (den statistiska) förväntan", de 20 procent som fördelar sig runt 0-värdet (dvs där predikterad och faktiska utfall är ungefär lika) klassas som "enligt förväntan"; på motsvarande sätt fås de 20 procent av skolorna som blir klassade som "bättre än (den statistiska) förväntan" och slutligen de 20 procent av skolor som hamnar i gruppen "bäst". Notera att det är två separata uppdelningar av skolorna som redovisas nedan, en som tagits fram utifrån Salsa-modellen och en som tagits fram utifrån modellen som liknar SCB:s socioekonomiska index. Utifrån respektive uppdelning jämförs hur samstämmiga modellerna är när det gäller att klassificera skolornas resultat.

Tabell 2 visar hur ofta en skolas prestation (sämst, under förväntan, enligt förväntan, över förväntan, bäst) är samstämmig enligt de två modellerna. Värdet 0 indikerar samma klassificering i båda måtten, värdet 1 att det är avvikande bedömningar i ett steg, t.ex. att Salsa måttet indikerar skolans prestation vara under förväntan medan SCB-måttet landar i enligt förväntan. Ett värde på 2 (3) indikerar att det är en skillnad om två (tre) stegs skillnad.

Det finns tydliga skillnader mellan utfallen av de två modellerna, där 49 procent av skolorna får olika placeringar i kvintil-fördelningen utifrån respektive modellberäkning.

**Tabell 2** Skillnader i bedömningen av skolornas resultat utifrån Salsa-modellen och modell baserad på SCB:s SE index. Baserad på kvintil-indelning för respektive beräkningsmått. Läsår 2018/19.

Skillnad i klassificering av skolor mellan Skolverkets och SCB:s beräkningsmodeller	Frekv.	Procent	Aggr.
0	747	51,09	51,09
1	567	38,78	89,88
2	133	9,10	98,97
3	15	1,03	100,00
<b>Total</b>	1 462	100,00	

Nedan visas en mer detaljerad redovisning över hur utfallen skiljer sig åt. Tabellvärden ska läsas enligt följande: ifall en skola enligt Salsa-beräkningarna är inordnad i kvintilen (femtedelen) om sämst presterande skolor gäller motsvarande klassificering utifrån beräkningarna som bygger på SCB:s socioekonomiska index i 211 av totalt 292 fall. I 64 fall klassas de sämsta skolorna enligt Salsa som *under*

förväntan i SCB modellen, i 15 fall klassas de som *enligt förväntan* i SCB modellen, i två fall *över förväntan*. På motsvarande sätt klassas av de 292 skolor som hamnar i den lägsta ("sämsta") kvintilen enligt modellen som bygger på SCB:s socioekonomiska index 211 som *sämst*, 59 *under förväntan*, 18 *enligt förväntan* och 4 *över förväntan* enligt Salsa-modellen.

**Tabell 3** Fördelningen i bedömningen av skolornas resultat utifrån Salsa-modellen och modell baserad på SCB:s SE index, läsår 2018/19

Standardiserat utifrån SCB-modellen	Standardiserat utifrån Salsa-modellen					totalt
	sämst	under förväntan	enligt förväntan	över förväntan	bäst	
sämst	211	59	18	4	0	292
under förväntan	64	116	63	44	6	293
enligt förväntan	15	91	109	64	16	292
över förväntan	2	24	86	111	70	293
bäst	0	3	16	73	200	292
<b>Totalt</b>	292	293	292	293	292	1 462

Resultaten visar på rätt stora skillnader i predikterade värden för skolenheterna. Det betyder att regressionsmodeller kan leda till mycket olika prediktioner beroende på vilka kontrollvariabler som inkluderas och hur dessa har blivit specificerade i regressionsmodellen. Detta gäller fastän modellerna inte skiljer sig särskilt mycket med avseende på deras respektive R<sup>2</sup>-värde. Enligt de redovisade värden i tabell 1 fås för Salsa-modellen ett (justerat) R<sup>2</sup>-värde på 62,3%, medan motsvarande i SCB:s modell är 68,0%, en skillnad på mindre än sex procentpunkter. Så trots att skillnaden i R<sup>2</sup>-värden är relativt liten föreligger det avsevärda skillnader i klassificering av skolenheter. Som vi skrev ovan hänvisar Skolverket i sin redovisning av Salsa-modellens tillförlitlighet till R<sup>2</sup>-värden ("förklaringsgrad"). Utifrån de resultat som redovisas här framstår denna praxis som inte speciellt tillförlitlig.

### Hur ser det ut inom Stockholms län?

Skolornas förutsättningar och resultat skiljer sig över landet, inte minst mellan skolor i landsbygdsområden och storstadsområden.<sup>22</sup> Men även inom storstadsområden finns det stora skillnader i skolornas uppskattade resultat. För att se hur olika modellprediktioner i de här undersökta modellerna spelar roll då man "zooma in" och ser på mer avgränsade områden visas i detta avsnitt en jämförelse av skolor som ligger inom Stockholms län. Anledningen till att vi har

<sup>22</sup> Se till exempel IFAU, RAPPORT 2014:25, 7.1 Skillnader i skolresultat.

valt Stockholms län är att det är regionen med flest grundskolor med årskurs 9, vilket gör det lättare att se hur mycket prediktioner från de två olika modellerna liknar respektive skiljer sig från varandra. Det är samma predikterade utfall som togs fram på riksnivå (i föregående redovisning) men där vi här enbart redovisar resultaten för skolorna inom Stockholms län.

Som framgår ifrån tabell 4 är det betydligt fler skolor som blir klassade som bäst utifrån Salsa med 143 skolor jämfört med SCB-modellens 87 av de 326 skolenheter som redovisas för Stockholms län. Detta tyder på att Salsa-modellen, i jämförelsen med SCB-modellen, inte i lika stor utsträckning justerar för olikheter i socioekonomiska förutsättningarna i Stockholmsområdet. Standardiseringen av en skolas betygsgenomsnitt i förhållande till socioekonomiska faktorer skiljer sig därmed avsevärt beroende på vilken modell som används.

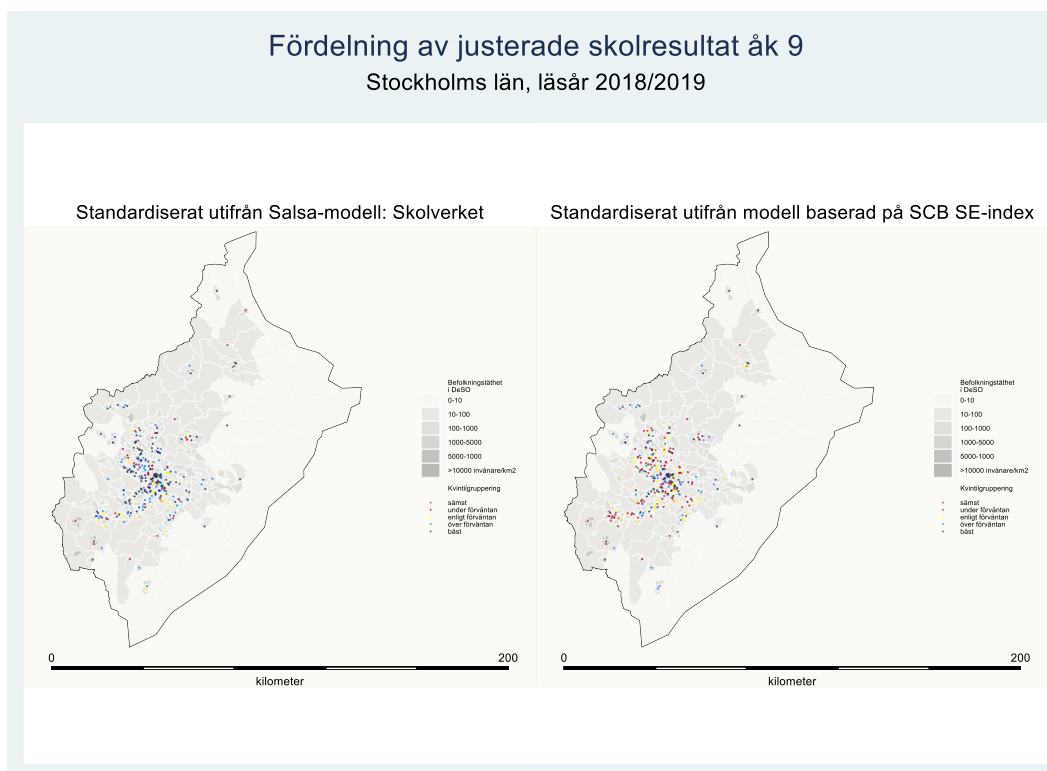
**Tabell 4** Fördelningen i bedömningen av skolornas resultat utifrån Salsa-modellen och modell baserad på SCB:s SE index. Läsår 2018/19. Redovisning av resultat för grundskolor i Stockholms län.

Standardiserat utifrån SCB-modellen	Standardiserat utifrån Salsa-modellen					bäst	totalt
	sämst	under förväntan	enligt förväntan	över förväntan	över		
sämst	25	14	10	2	0	51	
under förväntan	3	8	21	30	6	68	
enligt förväntan	0	1	6	37	14	58	
över förväntan	0	0	2	20	40	62	
bäst	0	1	0	3	83	87	
Totalt	28	24	39	92	143	326	

Ett alternativt sätt att se på skillnaderna i modellernas prediktioner är att se på skillnader i fördelning utifrån skolornas geografiska lokalisering.

I nedanstående karta visas resultat för skolor enligt motsvarande uppdelning som ovan. Skolor som hamnar i lägsta kvintilen markeras i röd färg, mittkategorien (medianen) gult, och de skolorna som hamnar högst i respektive beräkningsalgoritm markeras som blått. Det visar sig att färgspridningen är mer mångfacetterad och mer utspridd utifrån resultaten som genereras av modellen som bygger på SCB:s socioekonomiska index jämfört med skattningar som fås utifrån Salsa-modellen. Då man fokuserar på de mer centrala (befolkningstäta) delarna på kartan blir det tydligt att där enligt Salsa-modellens prediktioner majoriteten av skolorna är blå-klassade blir det enligt SCB:s modell fler inslag av röda prickar, dvs att fler skolor där hamnar i nedre delen av fördelningen.

**Figur 1** Utfall av beräkningar genom Salsa-modellen och (variant på) SCB:s socioekonomiska index. Uppdelning av justerade värden uppdelade i kvintiler. De 20 procent av grundskolor i Sverige som har lägsta värden (första kvintilen) benämns sämst, de 20 procent som följer benämns under förväntan, de skolor i tredje kvintilen (mittkvintilen) benämns enligt förväntan, skolor som ligger ovanför benämns över förväntan och skolorna i den högsta kvintilen benämns bäst. Redovisning av resultat för grundskolor i Stockholms län.



### Känslighetsanalys: alternativ klassificering av skolor

Klassificering ovan har gjorts utifrån en uppdelning i kvintiler (fem klasser). Som en känslighetsanalys har vi också tagit fram alternativa uppdelningar. Dels redovisas här uppdelning i kvartiler (fyra klasser), dels en uppdelning i decentiler (tio klasser).

**Tabell 5** Skillnader i bedömningen av skolornas resultat utifrån Salsa-modellen och modell baserad på SCB:s SE index. Baserad på kvartil-indelning för respektive beräkningsmått. Läsår 2018/19.

Skillnad i klassificering av skolor mellan Skolverkets och SCB:s beräkningsmodeller	Frekv.	Procent	Aggr.
0	857	58,62	58,62
1	536	36,66	95,28
2	67	4,58	99,86
3	2	0,14	100,00
<b>Total</b>	<b>1 462</b>	<b>100,00</b>	

**Tabell 6** Skillnader i bedömningen av skolornas resultat utifrån Salsa-modellen och modell baserad på SCB:s SE index. Baserad på decentil-indelning för respektive beräkningsmått. Läsår 2018/19.

Skillnad i klassificering av skolor mellan Skolverkets och SCB:s beräkningsmodeller	Frekv.	Procent	Aggr.
0	448	30,64	30,64
1	544	37,21	67,85
2	256	17,51	85,36
3	132	9,03	94,39
4	58	3,97	98,36
5	18	1,23	99,59
6	4	0,27	99,86
7	2	0,14	100,00
<b>Total</b>	<b>1 462</b>	<b>100,00</b>	

Det visar sig att en klassning i fyra kvartiler leder till något större överensstämmelser mellan utfall generade genom Salsa-modellen och SCB-modellen. Det är en följd av att färre alternativ leder till större samstämmighet. Motsatsen uppstår då man ser till en uppdelning i decentiler.

Vi har också utfört skattningar och uppdelningar som motsvarar rating av leverantörer inom Arbetsförmedlingens program ”Stöd och matchning”. Betygsättningen inom ratingen utgår ifrån medelvärdet och standardavvikelser, där de leverantörer som har residual-värden som ligger under en standardavvikelse räknad från medelvärdet får en stjärna, de med värden inom en standardavvikelse lägre än medelvärdet två stjärnor, de inom en standardavvikelse större än medelvärdet tre stjärnor och de leverantörerna med residual-värden mer än en standardavvikelse större än medelvärdet får fyra stjärnor.<sup>23</sup> En applicering av motsvarande uppdelning leder till följande resultat.

**Tabell 7** Fördelning i bedömningen av skolornas resultat utifrån Salsa-modellen och modell baserad på SCB:s SE index. Baserad på en fyrgradig uppdelning som följer uppdelningen i Arbetsförmedlingens rating-betygsättning. Läsår 2018/19.

Standardiserat utifrån SCB-modellen	Standardiserat utifrån Salsa-modellen				totalt
	En stjärna	Två stjärnor	Tre stjärnor	Fyra stjärnor	
En stjärna	134	46	6	0	186
Två stjärnor	62	347	125	5	539
Tre stjärnor	4	142	324	68	538
Fyra stjärnor	0	0	69	130	199
<b>Totalt</b>	200	535	524	203	1 462

Sammantaget är det 527 skolor som får andra betyg (antal stjärnor) då man använder sig av en modell som bygger på SCB:s socioekonomiska index jämfört med en uppdelning som bygger på Salsa-modellen. Det motsvarar 36 procent av skolorna.

<sup>23</sup> Se RiR 2020:13 för en granskning av hur systemet med rating inom Arbetsförmedlingens ”Stöd och matchning” fungerade, och hur ratingen påverkade val av privata leverantör genom arbetssökande.



## Redovisning av överensstämmelsen av Skolverkets modeller med avseende på socioekonomiska faktorer: fördelningsnyckel för grundskolan

Härnäst jämförs Skolverkets fördelningsnyckel för grundskolan med en referensmodell, som också här utgår ifrån SCB:s socioekonomiska index. Utfallsvariabeln som testas i respektive modell är elevernas sannolikhet att klara behörighet till yrkesgymnasium. Grundskattningar utförs på elevnivå, inte skolenhetsnivå som i föregående jämförelse av Salsa-modellen. Resultaten aggregeras därefter från elevnivå till skolenhetsnivå. Nedan visas underliggande skattningsresultat. Dessa bygger på logistisk regression (redovisningen i tabell 8 innehåller både koefficientestimat och odds-kvoter). Som framgår från tabell 8 innehåller Skolverkets fördelningsnyckel till grundskolan en mycket detaljerade uppdelning utifrån elevens tid i Sverige respektive hans egen och föräldrarnas ursprungsland.

**Tabell 8** Resultat av logit-skattningar med elevernas sannolikhet att uppnå yrkesgymnasiebehörighet i årkurs 9 som utfall; koefficient estimat och odds kvoter. Modell baserad på SCB:s SE index respektive Skolverkets fördelningsnyckel till grundskolan. Läsår 2018/19.

	SCB baserad 2019 koefficient (standardfel)	SCB baserad 2019 odds-ratio (standardfel)	Skolverket 2019 koefficient (standardfel)	Skolverket 2019 odds-ratio (standardfel)
elever med föräldrar som har högst förgymnasial utbildning	-1.3041*** (0.038)	0.2714*** (0.010)	-1.7802*** (0.042)	0.1686*** (0.007)
elever med föräldrar som har högst gymnasial utbildning	-0.8741*** (0.023)	0.4172*** (0.010)	-1.1694*** (0.024)	0.3105*** (0.007)
elever med föräldrar som har eftergymnasial utbildning	referens	referens	referens	referens
elever med föräldrar vars utbildningsnivå är okänd	-1.0008*** (0.054)	0.3676*** (0.020)	-1.2643*** (0.075)	0.2824*** (0.021)
elever med föräldrar som har mindre än 60% av median inkomst	-1.1577*** (0.048)	0.3142*** (0.015)		
elever med föräldrar som har 60–80% av medianinkomst	-0.9086*** (0.046)	0.4031*** (0.019)		
elever med föräldrar som har 80–100% av medianinkomst	-0.8235*** (0.046)	0.4389*** (0.020)		
elever med föräldrar som har 100–120% av medianinkomst	-0.5864*** (0.046)	0.5563*** (0.025)		
elever med föräldrar som har 120–150% av medianinkomst	-0.3024*** (0.048)	0.7390*** (0.035)		
elever med föräldrar som har mer än 150% av medianinkomst	referens	referens		
kvinnlig elev	0.2076*** (0.021)	1.2308*** (0.026)	0.1709*** (0.022)	1.1864*** (0.026)
elever som bor med båda föräldrar (ja/nej)	0.3354*** (0.023)	1.3984*** (0.033)		

	SCB baserad 2019 koefficient (standardfel)	SCB baserad 2019 odds-ratio (standardfel)	Skolverket 2019 koefficient (standardfel)	Skolverket 2019 odds-ratio (standardfel)
elever med antal syskon lika med 0	-0.3957*** (0.104)	0.6732*** (0.070)		
elever med antal syskon lika med 1	referens	referens		
elever med antal syskon lika med 2	0.0720** (0.027)	1.0747** (0.029)		
elever med antal syskon lika med 3	-0.1105*** (0.031)	0.8954*** (0.028)		
elever med antal syskon lika med 4	-0.3348*** (0.040)	0.7155*** (0.028)		
elever med antal syskon lika med 5 eller fler	-0.5317*** (0.050)	0.5876*** (0.029)		
elever i hushåll som fått ekonomiskt bistånd (ja/nej)	-0.5358*** (0.036)	0.5852*** (0.021)		
elever som bort längre än sex år i Sverige	referens	referens		
elever invandrade de senaste tre åren	-1.9912*** (0.044)	0.1365*** (0.006)		
elever invandrade de senaste fyra till sex åren	-0.7938*** (0.042)	0.4521*** (0.019)		
Andel högskoleutbildade i elevens DeSO	0.7394*** (0.190)	2.0947*** (0.397)		
Andel utlandsfödda i elevens DeSO	0.2822* (0.136)	1.3261* (0.180)		
Medianinkomst i elevens DeSO	0.0038*** (0.001)	1.0038*** (0.001)		
<i>Födelseland elev</i>				
Central- och Sydamerika			-0.3833* (0.166)	0.6816* (0.113)
Centrala, Östra och Södra Afrika			-1.1299*** (0.113)	0.3231*** (0.036)
EU-länder			-0.0457 (0.108)	0.9553 (0.103)

	SCB baserad 2019 koefficient (standardfel)	SCB baserad 2019 odds-ratio (standardfel)	Skolverket 2019 koefficient (standardfel)	Skolverket 2019 odds-ratio (standardfel)
Nordafrika			-0.9444*** (0.196)	0.3889*** (0.076)
Nordamerika			-0.2176 (0.281)	0.8044 (0.226)
Mellanöstern			-0.7935*** (0.086)	0.4523*** (0.039)
Oceanien			0.6034 (1.029)	1.8284 (1.881)
Skandinavien			0.1366 (0.179)	1.1464 (0.206)
Västafrika			-0.4661 (0.261)	0.6275 (0.164)
Övriga Europa			-0.3424** (0.118)	0.7101** (0.084)
Övriga länder i Asien			-0.3877*** (0.106)	0.6786*** (0.072)
Sverige			referens	referens
Okänd			0.2285 (0.557)	1.2567 (0.700)
<i>Födelseland mamman</i>				
Central- och Sydamerika			-0.3399*** (0.093)	0.7118*** (0.067)
Centrala, Östra och Södra Afrika			0.1107 (0.099)	1.1171 (0.110)
EU-länder			-0.1911** (0.074)	0.8260** (0.061)
Nordafrika			0.0113 (0.151)	1.0113 (0.152)
Nordamerika			0.0671 (0.277)	1.0694 (0.296)
Mellanöstern			0.1074 (0.075)	1.1134 (0.084)
Oceanien			-0.1446 (0.479)	0.8654 (0.415)
Skandinavien			-0.0922 (0.084)	0.9120 (0.077)

	SCB baserad 2019 koefficient (standardfel)	SCB baserad 2019 odds-ratio (standardfel)	Skolverket 2019 koefficient (standardfel)	Skolverket 2019 odds-ratio (standardfel)
Västafrika			-0.2534 (0.206)	0.7762 (0.160)
Övriga Europa			0.0041 (0.075)	1.0041 (0.075)
Övriga länder i Asien			0.0027 (0.071)	1.0027 (0.071)
Sverige			referens	referens
Okänd			0.1873* (0.094)	1.2059* (0.113)
<i>Födelseland pappan</i>				
Central- och Sydamerika			-0.5256*** (0.092)	0.5912*** (0.054)
Centrala, Östra och Södra Afrika			-0.2710** (0.091)	0.7626** (0.069)
EU-länder			-0.0642 (0.071)	0.9379 (0.066)
Nordafrika			-0.5260*** (0.147)	0.5910*** (0.087)
Nordamerika			0.4157 (0.269)	1.5154 (0.407)
Mellanöstern			-0.5474*** (0.070)	0.5785*** (0.041)
Oceanien			0.5550 (0.505)	1.7419 (0.879)
Skandinavien			-0.2061** (0.079)	0.8137** (0.065)
Västafrika			-0.7346*** (0.161)	0.4797*** (0.077)
Övriga Europa			-0.4297*** (0.069)	0.6507*** (0.045)
Övriga länder i Asien			0.0702 (0.087)	1.0727 (0.094)
Sverige			referens	referens
Okänd			-0.2431*** (0.069)	0.7842*** (0.054)

	SCB baserad 2019 koefficient (standardfel)	SCB baserad 2019 odds-ratio (standardfel)	Skolverket 2019 koefficient (standardfel)	Skolverket 2019 odds-ratio (standardfel)
<i>Antal år boendes i Sverige</i>				
0 år i Sverige			-3.3387*** (0.244)	0.0355*** (0.009)
1 år i Sverige			-3.3705*** (0.163)	0.0344*** (0.006)
2 år i Sverige			-2.6930*** (0.146)	0.0677*** (0.010)
3 år i Sverige			-2.1339*** (0.149)	0.1184*** (0.018)
4 år i Sverige			-1.7700*** (0.152)	0.1703*** (0.026)
5 år i Sverige			-1.1888*** (0.156)	0.3046*** (0.047)
6 år i Sverige			-1.0034*** (0.152)	0.3666*** (0.056)
7 år i Sverige			-0.5567*** (0.162)	0.5731*** (0.093)
8 år i Sverige			-0.5042** (0.179)	0.6040** (0.108)
9 år i Sverige			-0.3074 (0.173)	0.7353 (0.127)
10 år i Sverige			-0.2985 (0.168)	0.7419 (0.125)
11 år i Sverige			-0.4933** (0.169)	0.6106** (0.103)
12 år i Sverige			-0.4804** (0.173)	0.6186** (0.107)
13 år i Sverige			-0.3899* (0.175)	0.6771* (0.119)
14 år i Sverige			-0.6264*** (0.164)	0.5345*** (0.088)
15 år i Sverige			referens	referens
16 år i Sverige			-0.6707*** (0.165)	0.5114*** (0.084)

	SCB baserad 2019 koefficient (standardfel)	SCB baserad 2019 odds-ratio (standardfel)	Skolverket 2019 koefficient (standardfel)	Skolverket 2019 odds-ratio (standardfel)
17 år i Sverige			-2.6992*** (0.177)	0.0673*** (0.012)
18 år i Sverige			-2.8509*** (0.428)	0.0578*** (0.025)
19 eller fler år i Sverige			-2.1665*** (0.499)	0.1146*** (0.057)
<b>Konstant</b>	1.7711*** (0.147)	5.8776*** (0.864)	3.5075*** (0.170)	33.3651*** (5.662)
<b>chi2 (df_m)</b>	10454.6864 (21.0000)	10454.6864 (21.0000)	10382.2956 (59.0000)	10382.2956 (59.0000)
<b>r2_p</b>	0.2321	0.2321	0.2170	0.2170
<b>Antal elever</b>	111 093	111 093	111 129	111 129

Anm.: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . Robusta standardfel klustrade på skolenhetsnivå.

I efterföljande steg aggregeras uppgifter för elever till deras respektive skolenhet. Nedan redovisas hur samstämmiga resultaten är när det gäller att bedöma skolornas aggregerade risk (avseende elever i årskurs 9) att inte uppnå yrkesgymnasiebehörighet (vilket också ange behörighet till nationellt program enligt Skolverkets terminologi). Det är flera skolor som ingår i redovisningen här (se tabell 9) än i jämförelsen av Salsa (jämför med tabell 2). Detta förklaras av att det i Skolverkets fördelningsnyckel till grundskolan inte görs motsvarande avgränsningar som det görs i beräkningar av Salsa.

De 20 procent av skolorna som har lägst predikterat andel samlas i gruppen "lägst" risk, de efterföljande 20 procent blir klassade som "under medianen", de 20 som fördelar sig runt 0-värdet (dvs där predikterad och faktiska utfall är ungefär lika) klassas som "enligt förväntan"; därpå följer de 20 procent av skolorna som blir klassade som "över medianen" och slutligen de 20 procent av skolor som hamnar i gruppen "högst" risk.

Som framgår av tabellvärden i tabell 9 är diskrepansen mellan de två beräkningsmodellerna betydligt mindre än vi såg i jämförelsen av Salsa-modellen och en modell som bygger på det socioekonomiska indexet (se tabell 2). Skolverkets fördelningsnyckel och SCB modellen är lika utifrån en kvintilindelning i två tredjedelar (66,4%) av skolorna. Dock föreligger också en del avvikelser, i tre fall att samma skolor klassas som lägst (enligt SCB:s modell) respektive högst (enligt Skolverkets modell).

**Tabell 9** Skillnader i bedömningen av skolornas predikterad andel av elever att inte uppnå gymnasiebehörighet utifrån Skolverkets fördelningsnyckel till grundskolan och modell baserad på SCB:s SE index. Baserad på kvartil-indelning för respektive beräkningsmått. Läsår 2018/19.

Skillnad i klassificering av skolor mellan Skolverkets och SCB:s beräkningsmodeller	Frekv.	Procent	Aggr.
0	1 088	66,42	66,42
1	498	30,40	96,83
2	41	2,50	99,33
3	8	0,49	99,82
4	3	0,18	100,00
<b>Totalt</b>	<b>1 638</b>	<b>100,00</b>	

I efterföljande tabell 10 redovisas en mer detaljerad redovisning av skillnader i predikterade utfall.

**Tabell 10** Fördelningen i bedömningen av skolornas predikterad andel av elever att inte uppnå gymnasiebehörighet utifrån Skolverkets fördelningsnyckel till grundskolan och modell baserad på SCB:s SE index. Läsår 2018/19.

Standardiserat utifrån SCB-modellen	Standardiserat utifrån Skolverkets fördelningsnyckel för grundskolan					totalt
	lägst	under medianen	median kvintilen	över medianen	högst	
Lägst	256	60	5	3	3	327
under medianen	60	179	78	7	4	328
median kvintilen	11	80	167	63	7	328
över medianen		8	75	209	36	328
Högst		1	3	46	277	327
<b>Totalt</b>	<b>327</b>	<b>328</b>	<b>328</b>	<b>328</b>	<b>327</b>	<b>1 638</b>



## Hur ser det ut inom Stockholms län?

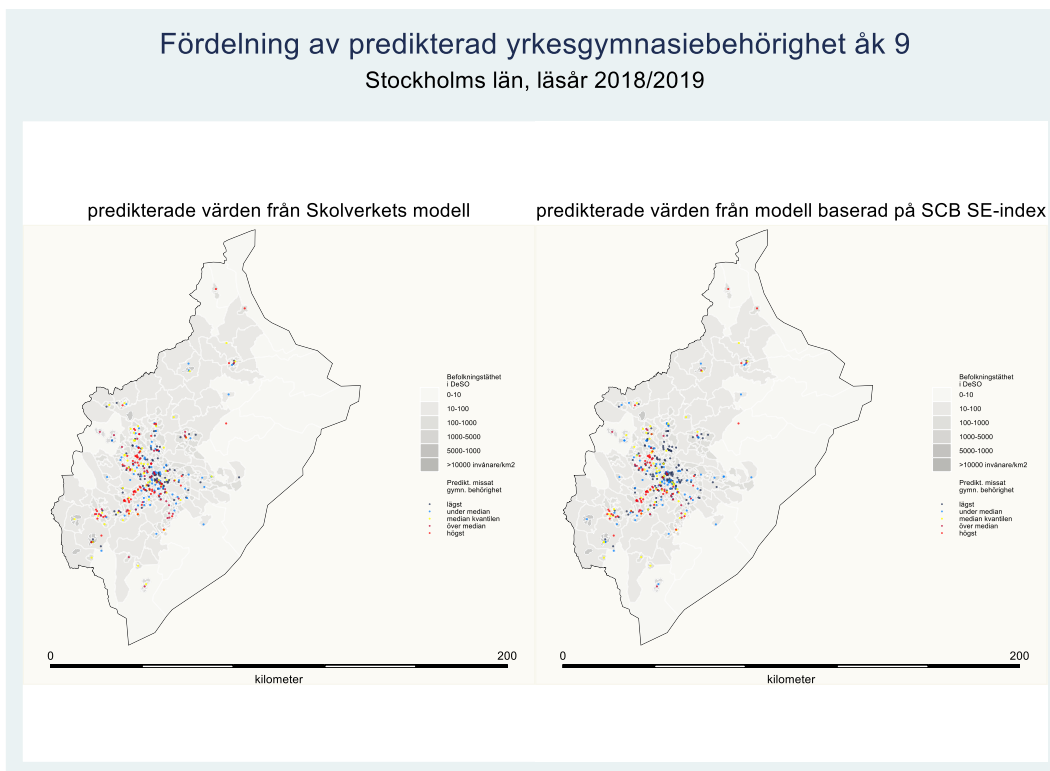
När man avgränsar jämförelsen och ser till Stockholms län ges följande bild. Det är 147 skolor inom Stockholms län som har lägst uppskattad andel elever att inte uppnå yrkesgymnasiebehörighet, enligt SCB modellen. Utifrån Skolverkets fördelningsnyckel är motsvarande antal 104 skolor. Enligt SCB modellen skulle därmed färre skolor inom Stockholms län få statsbidrag jämfört med skattningar som bygger på Skolverkets fördelningsnyckel för grundskolan. Detta gäller i den mån som statsbidrag fördelas kompensatorisk där skolor med högst uppskattat risk att eleverna inte uppnår behörighet till nationellt program tillräknas mer och de med lägst risk tillräknas mindre i statsbidrag.<sup>24</sup>

**Tabell 11** Fördelningen i bedömningen av skolornas predikterad andel av elever att inte uppnå gymnasiebehörighet utifrån Skolverkets fördelningsnyckel till grundskolan och modell baserad på SCB:s SE index. Läsår 2018/19. Redovisning av resultat för grundskolor i Stockholms län.

Standardiserat utifrån SCB-modellen	Standardiserat utifrån Skolverkets fördelningsnyckel för grundskolan					totalt
	lägst	under medianen	median kvintilen	över medianen	högst	
Lägst	95	42	4	3	3	147
under medianen	8	38	33	5	3	87
median kvintilen	1	1	22	27	3	54
över medianen	0	1	0	19	13	33
Högst	0	0	0	0	56	56
<b>Totalt</b>	<b>104</b>	<b>82</b>	<b>59</b>	<b>54</b>	<b>78</b>	<b>377</b>

<sup>24</sup> För att vara tydlig ska här betonas att riktade statliga bidrag i regel går till skolhuvudmännen som ansvarar för en eller flera skolor, inte till enskilda skolor. Det är skolhuvudmännen själva som avgör hur de tilldelade bidragsbelopp fördelas till sina skolor, utifrån givna ramar.

**Figur 2** Utfall av beräkningar genom fördelningsnyckel till grundskolan och modell baserad på SCB:s socioekonomiska index. Uppdelning av justerade värden uppdelade i kvintiler. De 20 procent av grundskolor i Sverige som har lägst värden (första kvintilen) benämns lägst, de 20 procent som följer benämns under median, de skolor i tredje kvintilen (mittkvintilen) benämns median kvintilen, skolor som ligger ovanför benämns över median och skolorna i den högsta kvintilen benämns högst. Redovisning av resultat för grundskolor i Stockholms län.



### Känslighetsanalys: alternativ klassificering av skolor

Klassificering ovan har gjorts utifrån en uppdelning i kvintiler (fem klasser). Som en känslighetsanalys har vi också tagit fram alternativa uppdelningar. Dels redovisas här uppdelning i kvartiler (fyra klasser), dels en uppdelning i decentiler (tio klasser).

**Tabell 12** Skillnader i bedömningen av skolornas predikterad andel av elever att inte uppnå gymnasiebehörighet utifrån Skolverkets fördelningsnyckel till grundskolan och modell baserad på SCB:s SE index. Baserad på kvartil-indelning för respektive beräkningsmått. Läsåret 2018/19.

Skillnad i klassificering av skolor mellan Skolverkets och SCB:s beräkningsmodeller	Frekv.	Procent	Aggr.
0	1 193	72,83	72,83
1	416	25,40	98,23
2	19	1,16	99,39
3	10	0,61	100,00
<b>Total</b>	<b>1 638</b>	<b>100,00</b>	

**Tabell 13** Skillnader i bedömningen av skolornas predikterad andel av elever att inte uppnå gymnasiebehörighet utifrån Skolverkets fördelningsnyckel till grundskolan och modell baserad på SCB:s SE index. Baserad på decentil-indelning för respektive beräkningsmått. Läsåret 2018/19.

Skillnad i klassificering av skolor mellan Skolverkets och SCB:s beräkningsmodeller	Frekv.	Procent	Aggr.
0	730	44,57	44,57
1	643	39,26	83,82
2	186	11,36	95,18
3	48	2,93	98,11
4	11	0,67	98,78
5	11	0,67	99,45
6	5	0,31	99,76
7	4	0,24	100,00
<b>Total</b>	<b>1 638</b>	<b>100,00</b>	

Det visar sig att en klassning i fyra kvartiler leder till något större överensstämmelser mellan skattningar baserade på Skolverkets fördelningsnyckel till grundskolan och SCB-modellen. Det är en följd av att färre alternativ leder till större samstämmighet. Ett motsatt förhållande uppstår då man ser till en

uppdelning i decentiler. Resultaten av känslighetsanalyserna framstår som inbördes konsistenta.

## Sammanfattande diskussion

Sammanfattningsvis konstaterar vi att de två här undersökta statistiska modeller från Skolverket skiljer sig utifrån sin träffsäkerhet då de jämförs med modeller som bygger på SCB:s socioekonomiska index. Fördelningsnyckeln till grundskolan leder till en relativt likartad klassificering som en modell baserad på SCB:s socioekonomiska index. I vårt huvudscenario, med en uppdelning i fem klasser, blir det samma klassificering för två tredjedelar av skolorna. Dock ser vi också en del skillnader, i enskilda fall diametralt olika bedömningar av skolornas position. Sådana olikheter i den statistiska bedömningen kan ha konkret betydelse, till exempel när det gäller tilldelningar av medel till skolhuvudmän. Detta gäller i den mån storleken av riktade statliga medel kopplas till inbördes ordning av skolor med avseende på deras uppskattade förutsättningar. Till exempel fördelas statsbidrag för hjälp med läxor och annat skolarbete utifrån en rangordning av alla huvudmän som sökt bidrag med hjälp av fördelningsnyckel för grundskolan.

Som vi har noterat har Salsa i en del fall använts för att jämföra och rangordna skolor utifrån tanken om skolornas kvalitet. Våra skattningar visar att en alternativ modell som utgår ifrån SCB:s socioekonomiska index och en uppdelning av resultat i kvintiler skulle leda till skilda resultat för nästan hälften av skolorna. En applicering av ratingbetygssystem som Arbetsförmedlingen har använt sig av i sin klassificering av leverantörer inom förmedlingstjänsten ”Stöd och matchning”, som i stället för att beskriva spridning i en fördelning utifrån kvintiler utgår ifrån medelvärden och standardavvikelser, visar att 36 procent av skolorna skulle få olika bedömningar (i form av antal stjärnor som varierar mellan 1 och 4) då en modell baserad på SCB:s socioekonomiska index skulle tillämpas i stället för Salsa. Utan att vi tagit ställning till vilken slags beräkningsmodell som skulle utgöra ett tillförlitligt mått på skolors kvalitet visar våra jämförelser att det är vanskligt att utgå ifrån Salsa som kvalitetsmått för grundskolor.

Resultatet att Skolverkets fördelningsnyckel för grundskolan i sina prediktioner är mer överensstämmande med modellen som bygger på SCB:s socioekonomiska index än vad Salsa förmår, har olika förklaringar. Dels används en flexibel uppdelning av föräldrarnas utbildning (enskilda indikatorer som fångar in föräldrarnas utbildning i stället för en enda variabel). Dessutom ingår mycket detaljerade uppdelningar av elevernas och deras föräldrarnas ursprungsland (uppdelade i regioner), samt tiden eleven har tillbringat i Sverige. Troligen samvarierar dessa med socioekonomiska faktorer, så som vårdnadshavares inkomstnivå eller arbetsmarknadsställning. Dock kvarstår faktumet att Skolverkets

fördelningsnyckel till grundskolan inte förhåller sig direkt till andra socioekonomiska faktorer än föräldrarnas utbildning.

Vår genomgång kan inte uttala sig om hur en modell som avser att adressera socioekonomiska förhållanden bör se ut i alla lägen. I slutändan är valet av faktorer i en statistisk modell som ska ligga till grund för bedömning av skolornas prestation och/eller behov av riktat stöd avhängiga av de ändamål som beslutsfattare avser att åstadkomma.

## Referenslista

### Litteratur

- Avvisati, F., "The measure of socio-economic status in PISA: a review and some suggested improvements", *Large-scale Assessments in Education*, 2020.
- Benmarker, H., Carling, K. och Forslund, A., *Vem blir långtidsarbetslös?*, IFAU RAPPORT 2007:20, 2007.
- Björklund, A., Lindahl, L. och Lindquist, M., "What More Than Parental Income, Education and Occupation? An Exploration of What Swedish Siblings Get from Their Parents", *The B.E. Journal of Economic Analysis & Policy*, vol. 10, nr 1, 2010.
- Brandén, G., Lindahl M. och Öckert, B., *The Importance of Nature–Nurture Interactions for Socioeconomic Outcomes: Evidence from a Large Sample of Adoptees, i Understanding Intergenerational Mobility*, IFAU DISSERTATION SERIES 2019:1.
- Holmlund, H., Häggblom, J., Lindahl, E., Martinson, S., Sjögren, A., Vikman, U. och Öckert, B., *Decentralisering, skolval och fristående skolor: resultat och likvärdighet i svensk skola*, IFAU-RAPPORT 2014:2, 2014.
- Shmueli, G., "To Explain or to Predict", *Statistical Science*, vol. 25, nr 3, 2010.

### Utredningar och rapporter

- DELMOS, *Segregation i Sverige – Årsrapport 2021 om den socioekonomiska boendesegregationens utveckling*, dnr DELMOS 2021/334, 2021.
- Riksrevisionen, *Stöd och matchning – ett valfrihetssystem för arbetssökande*, RiR 2020:13, 2020.
- Skolverket, *Redovisning av regeringsuppdraget Uppdrag om att ta fram förslag inför fördelning av medel*, dnr 2015:00898.
- Skolverket, *PISA 2015 15-åringars kunskaper i naturvetenskap, läsförståelse och matematik*, rapport 450, 2016.
- Skolverket, *PM\_Fördelningsnyckeln\_2017\_grundskolan*, internt PM, 2017.
- Skolverket, *Analyser av familjebakgrundens betydelse för skolresultaten och skillnader mellan skolor*, rapport 467, 2018.
- Skolverket, *Fördelningsnyckeln 2020\_gymnasieskolan*, internt PM, 2020.
- SOU 2017:35, *Samling för skolan – Nationell strategi för kunskap och likvärdighet*, 2017.

## Elektroniska källor

SCB, "DeSO – Demografiska statistikområden (scb.se)", <https://www.scb.se/hitta-statistik/regional-statistik-och-kartor/regionala-indelningar/deso---demografiska-statistikomraden/>, hämtad 2022-11-09.

Skolverket, "SALSA statistisk modell", <https://www.skolverket.se/skolutveckling/statistik/salsa-statistisk-modell>, hämtad 2022-04-01.