

Vilken betydelse har utländsk bakgrund för resultatet på högskoleprovet?

2

II

SVEN-ERIC REUTERBERG

II

2

2

III

Högskoleverkets rapportserie 2003:23 R

 **HÖGSKOLEVERKET**
National Agency for Higher Education

Vilken betydelse har utländsk bakgrund för resultatet på högskoleprovet?



SVEN-ERIC REUTERBERG

Högskoleverket • Birger Jarlsgatan 43 • Box 7851, 103 99 Stockholm
tfn 08-563 085 00 • fax 08-563 085 50 • e-post hsv@hsv.se • www.hsv.se

**Vilken betydelse har utländsk bakgrund för resultatet på
högskoleprovet? II**

Utgiven av Högskoleverket

Högskoleverkets rapportserie 2003:23 R

ISSN 1400-948X

Innehåll: **Sven-Eric Reuterberg**; Högskoleverket, utredningsavdelningen, **Nils
Olsson**

Formgivning: Högskoleverkets informationsavdelning

Tryck: Högskoleverkets vaktmästeri, Stockholm, juni 2003

Tryckt på miljömärkt papper

Innehållsförteckning

Förord	5
Sammanfattning	7
Inledning	9
Syfte	15
Metod	17
Undersökningsgrupper	17
Kontroll av allmän studieförmåga	22
Kontroll av selektionseffekter till högskoleprovet	26
Medelvärdesskillnader i delprovets generella faktorer	31
Vad betyder ålder för invandring för medelvärdena i de generella delprovsfaktorerna?	38
Skillnader i samband mellan delprovets generella faktorer och allmän studieförmåga	41
På jakt efter biasfaktorer	45
Innehåller ORD, NOG och DTK biasfaktorer?	45
Finns det textfaktorer i ELF och LÄS som gynnar grupp SS?	47
Den generella högskoleprovsfaktorns roll för gruppskillnaderna	51
Sammanfattande diskussion	55
Bilaga 1	65
Bilaga 2	67
Referenser	69

Förord

Allt fler personer med utländsk bakgrund söker idag till den svenska högskolan. Högskoleprovet är tillsammans med betyget från gymnasiet det viktigaste instrumentet för urvalet när antalet sökande övergår antalet studieplatser. För att öka kunskaperna om den språkliga bakgrundens betydelse för prestationen på högskoleprovet gav Högskoleverket professor Sven-Eric Reuterberg, vid institutionen för pedagogik och didaktik, Göteborgs universitet, uppdraget att studera eventuella skillnader i provresultat mellan personer med olika språklig bakgrund.

I en första delrapport med titeln *Vilken betydelse har utländsk bakgrund för resultatet på högskoleprovet?* visades att det fanns tydliga skillnader mellan de olika grupperna vad gäller resultat på provet. De som var födda i Sverige och hade minst en svenskfödd förälder nådde högre resultat än övriga grupper. Efter det att rapporten publicerats var det angeläget att närmare studera orsakerna till de skillnader som framkom.

I föreliggande rapport har författaren använt sig av faktoranalys för att studera vilka faktorer på högskoleprovet som försvårar provgenomförandet för den som har en annan språklig bakgrund än den svenska. Författaren svarar för innehållet och slutsatser i rapporten.

Sigbrit Franke
Universitetskansler

Sammanfattning

I denna rapport jämförs provtagare med olika språklig bakgrund vad gäller resultatet på det svenska högskoleprovet.

Resultaten visar att det finns tydliga skillnader i provresultatet mellan olika grupper. De högsta resultaten når de som är födda i Sverige med minst en svenskfödd förälder och de lägsta resultaten får i allmänhet de med minst en utlandsfödd förälder eller som har kommit till Sverige och har två svenskfödda föräldrar. Enda undantaget från dessa tendenser är det engelska läsförståelseprovet där de med utlandsfödda föräldrar har bättre provresultat än de svenskfödda som har minst en svensk förälder. Vidare visar resultaten att de med utländsk bakgrund som kommit till Sverige i 10-årsåldern eller senare genomgående har lägre resultat än de som kommit i yngre åldrar.

De skillnader i provresultat som finns mellan grupper kan inte hänföras till att enstaka uppgifter i högskoleprovet har ett innehåll som gynnar vissa grupper framför andra. I stället får skillnaderna tillskrivas de mer generella förmågor som mäts av delproven eller det totala provet.

Av resultaten att döma är det framförallt färdigheterna i svenska och kännedom om svenska förhållanden som är huvudorsaker till skillnaderna. För detta talar att det engelska läsförståelseprovet är det enda delprov som inte är på svenska eller som refererar till svenska förhållanden. Ytterligare ett stöd utgör det faktum att invandrare som kommit till Sverige i låg ålder klarar provet betydligt bättre än de som invandrat senare i livet.

Inledning

För att öka kunskaperna om den etniska och kulturella bakgrundens betydelse för prestationen på högskoleprovet gav Högskoleverket 1999 mig medel för en kunskapsöversikt inom området.

Den klassificering av språklig bakgrund jag använt mig av innefattar fyra grupper¹:

Grupp SS	personer födda i Sverige med minst en förälder född i Sverige
Grupp SU	personer födda i Sverige med båda föräldrarna födda utomlands
Grupp IS	personer födda utomlands med båda föräldrarna födda i Sverige
Grupp IU	personer födda utomlands med minst en förälder född utomlands

En första delstudie avrapporterades år 2001 (Reuterberg & Hansen, 2001). Resultaten i denna studie visade i korthet följande:

- grupperna SS och IS har högst andel provtagare, medan grupp IU har lägst
- grupperna SS och IS är mer benägna att upprepa sitt provtagande än övriga grupper
- grupp SS överträffar samtliga andra grupper vad gäller den högsta uppnådda normerade totalpoängen på provet och även på vart och ett av delproven
- invandrarstatus spelar en något större roll för män än för kvinnor vad gäller såväl den högsta uppnådda totalpoängen som de resultat som uppnåtts på delprovsnivå
- det finns inga tecken på att skillnaderna mellan de olika grupperna förändrats systematiskt under den tid som studerats (1991–1996) och denna stabilitet är i linje med vad man funnit i USA
- av de enskilda delproven är det DTK (diagram, tabeller och kartor) som uppvisar störst skillnader och ELF (engelsk läsförståelse) som visar minst skillnader

1. Grupperna SS och IU är heterogena såtillvida att de personer som tillhör dem kan ha en eller båda föräldrarna födda i Sverige respektive utomlands. För grupp SS gäller emellertid att 91,4 procent av personerna har föräldrar vilka båda är födda i Sverige. Av de personer som tillhör grupp IU har 81,9 procent föräldrar, vilka båda är födda utomlands. Grupp IS, slutligen, består till mycket stor del av personer som adopterats av svenskfödda föräldrar.

- ålder för inflyttning till Sverige har stor betydelse såtillvida att de som flyttat in före 10 års ålder tar provet i större utsträckning, upprepar sitt provtagande i högre grad och når en högre normerad poäng på provet jämfört med dem som flyttat in senare i livet
- inom grupp IU finns tydliga skillnader mellan grupper med olika födelse-land i benägenhet att ta provet, att upprepa sitt provtagande samt i provresultat
- också i USA och i Israel finns tydliga etniska skillnader både vad gäller benägenhet att genomgå antagningsprov till högre utbildning och vad gäller de resultat som uppnås
- kvinnornas högre benägenhet att genomgå högskoleprovet och männens högre provresultat är också utfall vilka överensstämmer väl med vad som gäller i USA.

I det pressmeddelande som lämnades då rapporten presenterades säger universitetskanslern att det är mycket viktigt att följa upp dessa resultat och närmare studera orsakerna till dem. Likaså säger utbildningsminister Thomas Östros i riksdagen som svar på en interpellation:

Jag har erfarit att Högskoleverket – med utgångspunkt i den nu publicerade rapporten – avser att initiera ytterligare studier i frågan. Bl.a. avser verket att undersöka om de observerade skillnaderna i resultat på högskoleprovet beror på att individerna har olika utbildningsbakgrund, dvs. valt olika program i gymnasieskolan. Verkets fortsatta studier ska också visa om det finns vissa slag av uppgifter i provet som möter särskilda svårigheter för dem som har utländsk bakgrund samt vad det kan vara som kännetecknar sådana uppgifter.

(Riksdagsprotokoll 2000/01:59, s. 37).

Mot denna bakgrund har jag fått i uppdrag av Högskoleverket att närmare studera bakgrunden till de tidigare påvisade skillnaderna mellan grupper med olika invandrarstatus.

I såväl den inbjudan till forsknings- och utvecklingsarbete som Högskoleverket gav 1999 som i utbildningsministerns riksdagsvar, betonas att det är viktigt att granska högskoleprovet med avseende på huruvida det finns sådana inslag i provet som skapar särskilda svårigheter för provtagare med en annan kulturell bakgrund än den svenska. Om det skulle finnas sådana inslag i provet föreligger vad som i engelskspråkig litteratur kallas ”bias”. Enligt Shepard, Camilli och Averill (1981) är ett test eller en uppgift ”biased” om personer med samma förmåga men tillhörande olika grupper inte har lika goda möjligheter att prestera bra på testet eller att klara uppgiften. När man här talar om lika förmåga syftar man naturligtvis på den förmåga som testet eller uppgiften är avsedd att mäta. För att man skall tala om bias måste således skillnaderna grupperna emellan vara orsakade av sådant som är irrelevant för mätningen. Cole och Moss (1989) uttrycker detta på följande vis:

A critical difficulty in interpreting item-bias studies is that the conclusion of bias does not directly follow the statistical results. One must still judge whether the basis of the differences on items is irrelevant to the construct and, hence, bias or whether it is relevant to the construct and, therefore not an issue of bias.

(a.a. s. 211).

En annan viktig distinktion är den mellan bias på uppgiftsnivå ("item bias") och bias på testnivå ("test bias") (Humphreys, 1986). Som påpekas av Roznowski och Reith (1999) har de två begreppen vanligen diskuterats som två separata fenomen och de två har sällan behandlats i ett och samma sammanhang. De, liksom Humphreys, varnar för att dra direkta slutsatser om bias på testnivå utifrån studier av bias på uppgiftsnivå:

The relationship between test and item bias often has been assumed, although there is little direct empirical evidence to support such an assumption.

(Roznowski & Reith, 1999, s. 249).

Roznowski och Reith hävdar att bias skall studeras och mätas på testnivå. Det är ju inte utfallet på de enskilda uppgifterna som ligger till grund för en bedömning av individen, utan utfallet på det totala testet. Man kan ju tänka sig det enkla fallet att vi har ett test med ett antal biasuppgifter, där somliga uppgifter gynnar en grupp medan andra gynnar en annan. Om så är fallet behöver ingalunda det totala testresultatet vara biased. Så är däremot fallet om samtliga eller huvuddelen av uppgifterna är biased till fördel för en och samma grupp.

Ett annat problem med att studera bias på uppgiftsnivå är att finna den verkliga anledningen till bias. I en tidig sammanställning av forskning rörande bias på uppgiftsnivå i amerikanska antagningsprov visar Boldt (1983) att man i somliga undersökningar kunde finna meningsfulla orsaker till snedvridningen, medan så inte var fallet i andra. Av de undersökningar som Boldt granskade hörde flertalet till den sistnämnda kategorin. Det visade sig också att när man funnit en meningsfull förklaring till bias i vissa undersökningar kunde denna inte verifieras i andra.

En något senare undersökning av samma typ som Boldts är den som genomfördes av Schmitt och Dorans (1990). Denna avsåg endast studier av Scholastic Assessment Test (SAT). På den matematiska delen av SAT fann Schmitt och Dorans få uppgifter som uppvisade några större skillnader mellan olika etniska grupper, medan antalet sådana uppgifter var större på den verbala delen. Enligt Schmitt och Dorans kunde tre orsaker till bias identifieras, nämligen grad av relevans för de olika grupperna, uppgifternas placering i provet – uppgifter som låg i slutet visade oftare bias – samt ords stavning och innebörd. När det gäller den sistnämnda förklaringen fann man att s.k. homografer (ord med lika stavning men med olika innebörd) tenderade att missgynna minoritetsgrupperna.

Det faktum att grad av bekantskap med uppgiftens innehåll kan vara en orsak till bias bekräftas av Freedle och Kostin (1997). De kallar denna förklaring ”hypotesen om kulturell bekantskap”. Likaså bekräftar Schmitt och Bleistein (1987) att bias på uppgiftsnivå kan ha sin grund i en snabbhetsfaktor så att skillnader uppträder på uppgifter i slutet av provet. Denna ”differential speededness” kan enligt författarna emellertid ha sin förklaring att uppgifter i slutet av provet skiljer sig från övriga uppgifter också i en rad andra avseenden.

Ytterligare en sammanställning av tidigare forskningsresultat med syfte att identifiera egenskaper hos prov och provuppgifter som missgynnar etniska minoriteter är den som genomfördes av Carlton och Harris (1992). Denna undersökning bekräftar till stora delar vad Boldt (1983) samt Schmitt och Dorans (1990) tidigare visat. Carlton och Harris sammanställning visar vidare att antalet biasuppgifter var begränsat samt att skillnaderna mellan olika grupper var måttliga på dessa uppgifter. Som exempel på detta kan nämnas att mindre än 5 procent av uppgifterna uppvisade en skillnad i lösningsfrekvens som översteg 10 procentenheter.

De sammanställningar som gjorts av amerikanska undersökningar tyder knappast på att bias är en viktig orsak till skillnader i provresultat mellan olika etniska grupper. Kan man då på grundval av detta dra den slutsatsen att bias inte heller är en viktig anledning till de skillnader som Reuterberg och Hansen (2001) fann mellan grupper med olika invandrarstatus i Sverige? Enligt min mening måste man vara försiktig med en sådan slutsats eftersom förhållandena är så olika i Sverige jämfört med USA. Dessutom finns det stora metodiska problem när det gäller att studera bias.

Ett sådant hänger samman med att man kan tala om bias först när man finner skillnader mellan olika *grupper som har samma förmåga*. I de amerikanska undersökningar som studerat bias på uppgiftsnivå har man på traditionellt sätt tagit det totala provresultatet som mått på förmåga. Dvs. man jämför svarsfrekvensen på en viss uppgift för dem som har samma totala provresultat. Om det finns många uppgifter som uppvisar samma typ av bias blir naturligtvis också det totala provresultatet påverkat och man kan då enbart upptäcka sådana uppgifter som uppvisar särskilt hög bias.

Ett annat problem är att enskilda uppgifter i ett prov sällan är endimensionella i den meningen att de mäter en enda bakomliggande förmåga (Gustafsson, 1988). Detsamma kan gälla också för hela provresultat. Som visas av Reuterberg (1999) kan man t.ex. i det engelska läsförståelseprovet (ELF) i högskoleprovet identifiera dels en generell faktor med anknytning till samtliga provuppgifter, dels innehållsfaktorer som anknyter till de enskilda frågor vilka är kopplade till en och samma text. Om man i ett sådant fall finner bias kan det vara svårt att med traditionell metodik avgöra huruvida skillnaderna har sitt ursprung i den generella faktorn eller i de mer specifika textfaktorerna. Ett annat problem med nära anknytning härtill är att avgöra om den faktor som uppvisar skillnader är en faktor som testet är avsett att mäta. Om

så är fallet betraktas skillnaden som relevant och det handlar då som tidigare nämnts inte om bias. Är däremot faktorn irrelevant handlar det om bias.

Enligt min mening är ett faktoranalytiskt angreppssätt en mera fruktbar väg att gå än att arbeta med observerade lösningsfrekvenser och poäng. En fördel med det faktoranalytiska angreppssättet är att man får en bild av uppgiftens och hela testets faktorstruktur. I synnerhet gäller detta om man arbetar med en "nested" faktormodell (Gustafsson & Balke, 1993). Att analysera en "nested" faktormodell innebär att man först tar ut den generella faktorn. Den bildas av den för alla uppgifterna gemensamma kovariansen. Av de kovarianser som sedan återstår mellan enskilda uppgifter bildas residualfaktorer, vilka är uttryck för mer specifika bakomliggande faktorer eller för förmågor som är snävare än den generella förmåga som hela testet mäter. Efter det att de olika faktorerna identifierats kan de olika gruppernas prestationer jämföras på faktornivå, vilket ger en mer nyanserad bild av hur skillnaderna i totalresultat uppstår.

Ett exempel kan möjligen underlätta förståelsen av detta.

En skolklass med 30 elever har genomgått ett matematikprov. Provet består av 20 uppgifter. Hälften av uppgifterna är benämnda (försedda med text) medan den andra hälften är rena räkneuppgifter (utan text).

I klassen ingår en grupp av elever med läs- och skrivsvårigheter, medan de övriga inte har några sådana problem.

En analys av uppgifternas interkorrelationer visar att samtliga 20 uppgifter korrelerar positivt inbördes. Detta indikerar att provresultaten påverkas av en generell förmåga som återspeglar elevernas matematiska förmåga (M-faktorn). När vi tagit ut M-faktorn återstår klara positiva korrelationer mellan 8 av de 10 benämnda uppgifterna. Detta visar att de 8 uppgifterna förutom M-faktorn också mäter en annan mer specifik faktor. Eftersom samtliga dessa uppgifter är benämnda är det rimligt att anta att denna specifika faktor utgör ett mått på elevernas läsförmåga (L-faktorn).

Ett intressant resultat vid faktoruppdelningen är att enbart 8 av 10 benämnda talen laddar i L-faktorn. En möjlig förklaring till detta kan vara att texten till de övriga 2 uppgifterna varit så enkel att de inte hämmat gruppen med läs- och skrivsvårigheter.

Provresultaten påverkas således inte av en enda förmåga utan av två. Provet är med andra ord inte endimensionellt.

En annan fördel med en sådan modell är att den generella faktorn och alla specifika faktorer är okorrelerade med varandra. Därigenom kan den totala testvariansen delas upp på de olika faktorerna vilket visar varje faktors bidrag till den totala variansen i provet. Samtidigt separeras också felvariansen ut från testets totala varians vilket innebär att de skillnader man finner på faktornivå är uttryck för sanna skillnader.

Som visats av Reuterberg och Gustafsson (1992) ökar en faktors vikt i totalresultatet dels med uppgifternas laddning i faktorn, dels med antalet uppgifter som laddar i faktorn ifråga. När det gäller det sistnämnda är det så, att vikten i totalvariationen bestäms av kvadraten på antalet uppgifter som laddar i faktorn. En faktor som uppvisar bias och har många uppgifter knutna till sig får således mycket allvarligare konsekvenser för det totala provresultatet jämfört med om faktorn har enbart några få uppgifter. Allra allvarligast är givetvis de fall där den generella faktorn uppvisar bias eftersom den påverkar samtliga uppgifter i testet.

Vid strukturell ekvationsmodellering (SEM) och med okorrelerade faktorer får man automatiskt en sådan variansuppdelning. I vårt exempel visade det sig att M-faktorn svarade för 60 procent av den totala variansen och L-faktorn för 25 procent. Den återstående variansandelen får hänföras till mycket svaga specifika faktorer samt till rena mätfel.

För en introduktion till strukturell ekvationsmodellering hänvisas till Maruyama (1998) och till Marcoulides och Hershberger (1997).

Med en uppdelning av totalvariansen på olika faktorer har man vidare möjlighet att jämföra olika gruppers resultat på var och en av faktorerna i stället för att basera jämförelsen på den observerade poängen som utgör en blandning av samtliga faktorer. Man får med andra ord mycket mer exakt information om vilka de bakomliggande förmågorna är som orsakar bias. Därmed blir det också lättare att sluta sig till orsakerna till bias. Att göra detta genom att enbart utgå ifrån uppgifternas yttre karakteristika är svårt. Två uppgifter kan ha flera gemensamma yttre egenskaper utan att de mäter samma bakomliggande förmåga. Likaså kan två till synes olika uppgifter mäta en och samma bakomliggande förmåga. Genom att studera uppgifternas kovarianser i en "nested" faktormodell får man information om i vad mån uppgifterna faktiskt mäter samma bakomliggande förmåga.

En jämförelse mellan de två elevgrupperna visar en skillnad till nackdel för dem med läs- och skrivsvårigheter på L-faktorn, men på M-faktorn finns inga skillnader mellan de två grupperna.

Resultaten visar således att det inte finns några skillnader mellan de två grupperna vad gäller matematisk förmåga, utan de skillnader som visat sig i totalpoängen beror på skillnader i L-faktorn. Utan en konfirmatorisk faktoranalys hade det legat nära till hands att dra den slutsatsen att gruppen med läs- och skrivsvårigheter sämre matematisk förmåga än övriga.

Vanligen har bias betraktats som ett rent kvantitativt fenomen. Man har jämfört grupper med avseende på lösningsfrekvenser på enskilda uppgifter eller totalpoäng på ett helt test. Enligt min mening kan bias också ha en kvalitativ aspekt. Man kan inte ta för givet att ett och samma test mäter exakt samma bakomliggande faktorer för alla grupper. Exempel på att så inte behöver vara

fallet är Freedles och Kostins (1997) hypotes om kulturell bekantskap samt Schmitts och Bleisteins (1987) snabbhetsfaktor. Sådana faktorer kan ha en klar inverkan på resultatet för vissa grupper, men helt sakna betydelse för andra. Det finns därför skäl att i analysen av bias också innefatta sådana jämförelser och här är naturligtvis den faktoranalytiska metodiken synnerligen användbar.

I vårt exempel är provet inte avsett att mäta läsförmåga. De skillnader som förekommer är således en effekt av bias.

Om man väger in en sådan kvalitativ aspekt i biasbegreppet kommer också validitetsfrågan i fokus. Som Cole och Moss (1989) hävdar kan man tala om bias enbart i de fall då gruppskillnader orsakas av faktorer som är irrelevanta för mätningen dvs. sådana faktorer som påverkar testresultatet, men som saknar betydelse för det kriterium mot vilket testresultatet skall ställas. Om inslaget av sådana faktorer varierar mellan olika grupper kommer också testets validitet att variera mellan grupperna.

Syfte

Syftet med föreliggande undersökning är att finna ut orsakerna till de skillnader i provresultat vid högskoleprovet som påvisats av Reuterberg och Hansen (2001). Beror skillnaderna på att de olika grupperna presterar olika i de generella faktorer som vart och delproven mäter, eller har skillnaderna sitt ursprung i att enskilda uppgifter gynnar vissa grupper på ett orättmätigt sätt? Om det senare visar sig vara fallet uppstår frågan huruvida man då kan finna grupper av biasuppgifter som indikerar att skillnaderna orsakas av någon mer specifik delförmåga som delprovet mäter vid sidan av dess generella faktor.

Metod

Att studera bias på testnivå implicerar två huvudproblem av metodisk art. Det första är förknippat med definitionen av bias, nämligen att vi har bias enbart i de fall då man finner skillnader mellan olika grupper som är lika i den förmåga som testet är avsett att mäta. När det gäller bias på uppgiftsnivå har detta problem, som tidigare nämnts, lösts så att man tar det totala testresultatet som mått på denna förmåga – en lösning som emellertid förutsätter att testresultatet i sig inte är biased. Då man studerar bias på testnivå måste man mäta förmågan på något annat sätt än med hjälp av det test som skall analyseras. Det mått på den allmänna förmågan som jag valt är grundskolebetyg och jag kommer i detta avsnitt att diskutera hur väl betygen fungerar som en kontrollvariabel för allmän studieförmåga.

Det andra metodiska huvudproblemet är att den grupp som genomgått testet ofta är selekterad på ett eller annat sätt. När det gäller högskoleprovet är provtagarna positivt selekterade ur totalpopulationen vad gäller allmän studieförmåga. De som lyckats mindre bra i sina tidigare studier planerar inte att genomgå en högre utbildning och har därför ingen anledning att genomgå provet. När det gäller jämförelser mellan olika grupper tillkommer det faktum att selektionen till provet kan vara olika sträng inom olika grupper, och om så är fallet kommer de skillnader man finner i provresultat att vara annorlunda än de man skulle finna om samtliga i populationen genomgått provet. För en mer ingående diskussion om effekterna av sådan differentiell selektion hänvisas till Reuterberg (1996). Att vid differentiell selektion studera bias bland enbart dem som genomgått provet kan således leda till helt felaktiga slutsatser om bias i provet. Jag kommer i detta avsnitt att granska hur selektionen till provet är relaterad till invandrarstatus samt att kort presentera en metod med vars hjälp man kan undvika detta problem.

Innan jag går in på dessa metodproblem skall emellertid undersökningsgrupperna presenteras.

Undersökningsgrupper

I denna studie har vi använt en databas som innefattar bakgrundsuppgifter och utbildningsuppgifter för samtliga personer som är födda under åren 1972–1984. Totalt ingår 1,3 miljoner personer i databasen. Av tids- och kostnadsskäl kan hela denna grupp inte ingå i den föreliggande undersökningen, utan jag har valt dem som är födda 1977 respektive 1982. Den förstnämnda gruppen är en av de sista grupperna där alla har fått grundskolebetyg enligt den gamla 5-gradiga skalan, och de stod i begrepp att söka till högre utbildning då det fortfarande var mycket vanligt att man försökte kvalificera sig till högre utbildning antingen via gymnasiebetyg eller högskoleprovsresultat. De

födda 1982 har däremot fått de nya kriterierrelaterade betygen med fyra olika betygssteg, och då dessa skulle söka till en högre utbildning hade förutsättningarna ändrats på så sätt att man fått möjlighet att kvalificera sig också via komplettering av gymnasiebetygen. Denna möjlighet är en bidragande faktor till att betydligt färre genomgått högskoleprovet i den yngre gruppen.

De bakgrundsuppgifter som är aktuella i denna undersökning är personernas egna och deras föräldrars födelseland samt, i de fall där så är aktuellt, tidpunkten för invandring till Sverige. Med ledning av dessa uppgifter har jag skapat variabeln invandrarstatus och den omfattar fyra olika kategorier, vilka är identiska med dem i Reuterbergs och Hansens tidigare studie:

Grupp SS	personer födda i Sverige med minst en förälder född i Sverige
Grupp SU	personer födda i Sverige med båda föräldrarna födda utomlands
Grupp IS	personer födda utomlands med båda föräldrarna födda i Sverige
Grupp IU	personer födda utomlands med minst en förälder född utomlands

Gruppen IU kommer i vissa analyser att delas upp ytterligare med avseende på tidpunkt för invandring. Två grupper kommer att särskiljas, nämligen de som invandrat före tio års ålder och de som invandrat senare i livet. Anledningen härtill är att ålder för invandring har en klar inverkan på benägenheten att genomgå högskoleprovet (Gustafsson, Andersson & Hansen, 2000) liksom på provresultaten (Reuterberg & Hansen, 2001).

Det totala resultatet på högskoleprovet normeras så att det resultat man uppnått vid ett provtillfälle görs jämförbart med resultat från andra tillfällen. Någon motsvarande normering görs däremot inte av de resultat som uppnås på de enskilda delproven. Detta innebär att provtillfället måste hållas under kontroll i de analyser som skall göras här. Jag har då valt vårproven det år då provtagarna fyller 19 dvs. 1996 respektive 2001. Motiveringen för detta val är att det är det provtillfälle där vi har det största antalet provtagare. Gruppernas storlek framgår av tabellerna 1 och 2 och i tabell 3 anges andelen provtagare i relation till invandrarstatus och kön.

Tabell 1. Populationens* storlek samt antal provtagare i relation till invandrarstatus och kön. Födda 1977.

Grupp	Population				Provtagare			
	Män	Kvinnor	Totalt	%	Män	Kvinnor	Totalt	%
SS	43 645	41 070	84 715	90,8	7 914	9 811	17 725	93,2
SU	1 786	1 861	3 647	3,9	253	325	578	3,0
IS	580	825	1 405	1,5	94	171	265	1,4
IU	1 796	1 685	3 481	3,7	222	229	451	2,4
Samtliga	47 807	45 441	93 248	100,0	8 483	10 536	19 019	100,0

* Med population avses här de personer som har kompletta uppgifter.

Tabell 2. Populationens storlek samt antal provtagare i relation till invandrarstatus och kön. Födda 1982.

Grupp	Population				Provtagare			
	Män	Kvinnor	Totalt	%	Män	Kvinnor	Totalt	%
SS	41 743	39 424	81 167	90,6	4 659	5 204	9 863	91,7
SU	2 029	1 986	4 015	4,5	209	221	430	4,0
IS	608	680	1 288	1,4	58	77	135	1,3
IU	1 583	1 530	3 113	3,5	169	159	328	3,0
Samtliga	45 963	43 620	89 583	100,0	5 095	5 661	10 756	100,0

Tabell 3. Andel provtagare i relation till födelseår, invandrarstatus och kön. (I procent.)

Grupp	Födda 1977			Födda 1982		
	Män	Kvinnor	Totalt	Män	Kvinnor	Totalt
SS	18,1	23,4	20,9	11,2	13,2	12,2
SU	14,2	17,5	15,8	10,3	11,1	10,7
IS	16,2	20,7	18,9	9,5	11,3	10,5
IU	12,4	13,6	13,0	10,7	10,4	10,5
Samtliga	17,7	23,2	20,3	11,1	13,0	12,0

Trots att de båda årskullarna i stort sett är lika stora, 99 711 födda 1977 och 99 666 födda 1982, innefattar den yngre gruppen nästan 4 000 färre personer som framgår av tabellerna 1 och 2. Anledningen härtill är dels att det är fler personer i den yngre gruppen som saknar uppgift om invandrarstatus, dels att fler saknar uppgift om de nya betygen. En delförklaring till det sistnämnda är att 800 personer lämnat grundskolan redan 1997 eller tidigare, då betygen sattes enligt den gamla femgradiga skalan. Avsaknaden av betyg och uppgift om invandrarstatus är för övrigt den huvudsakliga förklaringen till att antalet personer i populationen understiger det totala antalet personer i respektive kohort.

Av tabellerna 1 och 2 framgår också att grupp SS är överrepresenterad bland provtagarna jämfört med populationen, medan främst grupperna SU och IU är underrepresenterade.

Det faktum att den yngre gruppen drabbats av ett större bortfall kan emellertid inte förklara den stora nedgången i antalet högskoleprovstagare. Som tabellerna 1 och 2 visar har detta antal sjunkit från 19 000 till knappt 11 000, vilket motsvarar 20,3 respektive 12,0 procent av samtliga i undersökningsgrupperna. Denna nedgång är inte unik för 19-åringar utan gäller generellt för högskoleprovet. Som jämförelse kan nämnas att antalet provtagare totalt sett sjönk från 144 061 år 1997 till drygt 100 000 år 2000 (Stage & Ögren, 2001, s. 1). En anledning till denna stora nedgång är, som tidigare nämnts, att det numera är möjligt att i efterhand komplettera sitt gymnasiebetyg, vilket många troligen anser vara en säkrare väg att förbättra sina möjligheter till att bli antagna till en högskoleutbildning än att upprepa sitt provtagande (Svensson, 2002). Andra faktorer som bidragit till denna nedgång är att årskullarnas storlek minskat samt att antalet utbildningsplatser ökat, vilket gjort det lättare att komma in på högskolan.

Som tabell 3 visar har andelen provtagare sjunkit mest inom grupperna SS och IS, vilket medför att skillnaderna i andel provtagare mellan de fyra grupperna är betydligt lägre i den yngre gruppen jämfört med i den äldre. Även könsskillnaderna har reducerats påtagligt beroende på att nedgången i antal provtagare varit särskilt stor bland kvinnorna. Fortfarande är det dock något vanligare bland kvinnor än bland män att genomgå högskoleprovet. Kvinnornas större benägenhet att genomgå högskoleprovet gör att de är i majoritet bland provtagarna. Särskilt tydligt är detta inom grupp IS där kvinnorna utgör nästan två tredjedelar av provtagarna födda 1977 och 57 procent av provtagarna i den yngre årskullen. Dessa sneda könsfördelningar inom grupp IS får konsekvenser för hur denna grupp skiljer sig från övriga grupper i resultat på främst de kvantitativa delproven eftersom det finns särskilt stora könsskillnader till männens fördel på dessa prov.

För att undersöka de skillnader som finns beroende på könsfördelning kommer vissa analyser att genomföras dels för den totala gruppen, dels med uppdelning på kön.

Avslutningsvis finns det skäl att också granska vilka skillnader som föreligger mellan de olika invandrargrupperna vad gäller totalpoäng på högskoleprovet.

Tabell 4. Totalpoäng på högskoleprovet i relation till kön och invandrarstatus. Födda 1977. Medelvärde (M) och standardavvikelse (S).

Grupp	Män		Kvinnor		Totalt	
	M	S	M	S	M	S
SS	78,79	16,31	71,92	17,48	74,99	17,31
SU	72,56	18,56	63,99	18,23	67,71	18,84
IS	71,78	19,58	63,87	17,39	66,67	18,55
IU	68,96	19,16	63,02	18,03	65,95	18,81
Samtliga	78,27	16,62	71,35	17,64	74,44	17,53

Tabell 5. Totalpoäng på högskoleprovet i relation till kön och invandrarstatus. Födda 1982.

Grupp	Män		Kvinnor		Totalt	
	M	S	M	S	M	S
SS	70,67	16,96	64,97	16,68	67,65	17,05
SU	63,69	17,63	56,69	16,26	60,13	17,29
IS	60,00	14,38	56,44	15,67	57,96	15,18
IU	58,94	18,26	56,92	14,80	57,96	16,67
Samtliga	69,86	17,22	64,30	16,75	66,93	17,20

De medelvärden som redovisas i tabellerna 4 och 5 gäller enbart provtagargruppen och de är därför påverkade av såväl selektion som differentiell selektion till provet. De ger således inte en rättvisande bild av vilken roll invandrarstatus spelar för resultaten på högskoleprovet. Detta innebär dock inte att de är ointressanta, för de återspeglar ju de faktiska skillnader i provresultat och därmed också hur möjligheterna att bli uttagen till en utbildning varierar mellan grupperna.

Som tabellerna 4 och 5 visar uppnår grupp SS genomgående högst resultat, medan skillnaderna mellan de övriga tre grupperna är små och det sistnämnda gäller i synnerhet för kvinnorna. Även männen uppvisar små skillnader, men för dem är rangordningen den samma i båda årskullarna och den innebär att grupp SU har högst medelvärde och grupp IU lägst. Ser man

till de skillnader som föreligger mellan grupp SS och övriga grupper är den så stor att den får betydande konsekvenser för möjligheterna att bli antagen till en högre utbildning.

Förutom skillnaderna mellan de fyra grupperna finns också en klar skillnad mellan årskullarna och mellan manliga och kvinnliga provtagare. Skillnaden mellan årskullarna är så stor, ca sju poäng, att den knappast kan förklaras av skillnader i provens svårighetsgrad. Däremot kan den mycket väl hänga samman med skillnader i selektion till provet. När det slutligen gäller köns-skillnaderna är de av samma natur som en rad tidigare undersökningar visat, nämligen att männen uppnår högre resultat än kvinnorna på provet. I den äldre årskullen uppgår skillnaden till nästan sju poäng, men i den yngre har den minskat till fem och en halv.

Kontroll av allmän studieförmåga

Enligt definitionen på bias skall man, som tidigare nämnts, jämföra grupper som är lika med avseende på den förmåga som testet är avsett att mäta. När det gäller bias på uppgiftsnivå brukar man ta det totala testresultatet som mått på denna förmåga. Vid studier av bias på testnivå måste däremot förmågan mätas på annat sätt. Eftersom högskoleprovet används som ett urvalsinstrument till högre utbildning är det personernas studieförmåga som skall kontrolleras. En möjlighet hade här varit att ta gymnasiebetygen som mått på studieförmågan. Gymnasiebetygen är ju det urvalsinstrument som används vid sidan av högskoleprovet. Tyvärr är i varje fall de gamla relativa gymnasiebetygen inte möjliga att arbeta med här. Detta hänger samman med att de studerande på varje gymnasielinje eller program utgör sin egen referensgrupp (jämförelsegrupp) vid betygsättningen, vilket medför att betyg från en linje/program inte är jämförbara med dem från en annan. Däremot är grundskolebetygen jämförbara för alla och sambandet mellan grundskolebetyg och gymnasiebetyg är så högt att de kan tjäna som kontrollvariabel i denna undersökning.

De fem delprov som ingår i högskoleprovet är konstruerade för att mäta olika sidor av studieförmågan. Vi har å ena sidan NOG och DTK som båda mäter en kvantitativt analytisk förmåga medan de övriga tre delproven ORD, LÄS och ELF mäter en språklig förmåga. Som visats av Gustafsson, Wedman och Westerlund (1992) är emellertid dessa båda sidor korrelerade vilket visar att det också finns en generell förmåga som manifesterar sig i båda delförmågorna. Eftersom urvalet till högre utbildning sker utifrån den poängsumma som den sökande uppnått på samtliga fem delprov sammanlagt blir det den generella studieförmågan som blir utslagsgivande oavsett till vilken utbildning urvalet sker.

Den variabel som här skall fungera som kontrollvariabel skall följaktligen också den avse generell studieförmåga. Som mått på denna valdes först grundskolebetyg i samtliga teoretiska ämnen. Förutom kravet på att ha en

kontrollvariabel som så väl som möjligt täcker begreppet studieförmåga måste kontrollvariabeln ha samma innebörd för alla fyra grupperna. Detta innebär att det inte får finnas några större skillnader mellan grupperna med avseende på den vikt, med vilken ett visst betyg ingår i den generella studieförmågan. Då detta kontrollerades visade det sig att betyget i svenska hade starkt varierande laddningar i den generella faktorn för de olika grupperna. Den kvadrerade laddningen anger betygets reliabilitet och för grupp SS i den äldre populationen hade betyget i svenska en reliabilitet som uppgick till 0,682, vilket skall jämföras med 0,574 för grupp SU och 0,497 för grupp IU. Av detta skäl uteslöts betyget i svenska vid mätningen av den generella studieförmågan. Vilka ämnesbetyg som därefter återstår framgår av tabellerna 6 och 7 där också ämnesbetygens reliabiliteter redovisas för var och en av grupperna. I tabell 8 redovisas den generella studieförmågans reliabilitet med uppdelning också på kön.

Tabell 6. Ämnesbetygens reliabilitet för personer med olika invandrarstatus. Födda 1977.

Ämnesbetyg	SS	SU	IS	IU
Engelska	.536	.539	.506	.507
Matematik	.555	.540	.545	.566
Biologi	.688	.681	.648	.696
Fysik	.650	.625	.617	.665
Kemi	.677	.656	.644	.671
Teknik	.365	.376	.365	.387
Geografi	.781	.785	.710	.796
Historia	.805	.804	.772	.811
Religion	.792	.795	.750	.789
Samhällskunsk.	.807	.813	.771	.810

Tabell 7. Ämnesbetygens reliabilitet för personer med olika invandrarstatus. Födda 1982.

Ämnesbetyg	SS	SU	IS	IU
Engelska	.480	.485	.417	.489
Matematik	.477	.477	.414	.508
Biologi	.600	.597	.547	.589
Fysik	.563	.578	.499	.593
Kemi	.577	.571	.503	.594
Teknik	.348	.350	.342	.397
Geografi	.739	.766	.719	.745
Historia	.773	.783	.736	.799
Religion	.755	.777	.703	.744
Samhällskunsk.	.772	.799	.730	.794

Tabell 8. Den generella studieförmågans reliabilitet för personer med olika invandrarstatus med uppdelning på födelseår och kön.

		SS	SU	IS	IU
Födda 1977	Män	.952	.950	.941	.951
	Kvinnor	.953	.952	.946	.954
Födda 1982	Män	.938	.942	.923	.941
	Kvinnor	.942	.943	.930	.947

Som tabellerna 6 och 7 visar har de flesta ämnesbetygen en något lägre reliabilitet i grupp IS jämfört med övriga grupper, vilket leder till att den totala betygsskalan också får en något lägre reliabilitet för denna grupp (tabell 8). Skillnaderna är emellertid små och de beror inte på att felvariansen är högre, utan på det förhållandet att den sanna variansen är lägre inom denna grupp än inom de övriga. Detta visar att det snarare är gruppens sammansättning än betygets mätegenskaper som ligger bakom de något lägre reliabiliteterna för grupp IS. Denna tolkning har för övrigt ett starkt stöd i den tidigare rapporten om högskoleprovsresultat och invandrarstatus (Reuterberg & Hansen, 2001). Där visas nämligen att grupp IS är en avsevärt mer homogen grupp än övriga vad gäller social sammansättning. Av samtliga i grupp IS kommer nämligen inte mindre än 90 procent från någon av socialgrupperna 1 eller 2, vilket kan jämföras med knappt 70 procent för grupp SS, som är den grupp som visar näst minst spridning i social sammansättning. Ytterligare ett bevis

på detta finner vi i tabellerna 11 och 12 nedan, där grupp IS uppvisar den lägsta spridningen vad gäller generell studieförmåga.

Mot denna bakgrund blir slutsatsen att det inte finns några väsentliga skillnader mellan grupperna vad gäller innebörden av generell studieförmåga så som den mäts av de utvalda grundskolebetygen. De tio utvalda betygen blir således de som ligger till grund för mätningen av generell studieförmåga.

Innan undersökningens huvudresultat redovisas finns det skäl att också studera vilka samband som råder mellan den generella studieförmågan mätt via grundskolebetygen och resultaten på de olika deltesten i högskoleprovet. För att här undvika inflytande av mätfel kommer sambanden att studeras via SEM (Strukturell EkvationsModellering), där de tio enskilda betygen definierar generell studieförmåga och där tre delskalor med slumpmässigt fördelade uppgifter definierar den generella delprovsfaktorn i vart och ett av de fem proven.

Tabell 9. Samband mellan generell studieförmåga och de generella delprovsfaktorerna. Provtagare.

Födda		ORD	NOG	LÄS	DTK	ELF
1977	Män	.509	.467	.554	.500	.478
	Kvinnor	.548	.559	.618	.562	.585
	Samtliga	.527	.479	.590	.487	.525
1982	Män	.450	.424	.433	.451	.429
	Kvinnor	.498	.492	.524	.508	.517
	Samtliga	.470	.409	.475	.421	.444

Tabell 10. Samband mellan generell studieförmåga och de generella delprovsfaktorerna. Population.

Födda		ORD	NOG	LÄS	DTK	ELF
1977	Män	.665	.622	.706	.656	.633
	Kvinnor	.680	.687	.744	.691	.714
	Samtliga	.696	.656	.751	.665	.698
1982	Män	.582	.552	.565	.581	.559
	Kvinnor	.632	.625	.659	.643	.651
	Samtliga	.638	.590	.641	.599	.617

En jämförelse mellan tabellerna 9 och 10 visar att sambanden ökar påtagligt när vi går från provtagargruppen till populationen. Tabell 10 bygger på en uppskattning av sambandet i hela populationen, utifrån det observerade sambandet bland provtagarna i tabell 9. Eftersom provtagarna inte är representativa för hela populationen får man olika stora samband. Detta är en fullt naturlig förändring mot bakgrund av att provtagarna utgör en positivt selekterad grupp med betydligt lägre spridning i såväl betyg som provresultat. Det finns också en viss skillnad mellan årskullarna, såtillvida att sambanden är lägre för den yngre gruppen, som fått betyg enligt den nya kriterierelaterade betygsskalan med fyra betygssteg, medan 77:orna däremot har betyg uttryckta i en femgradig normrelaterad betygsskala. Denna skillnad i betygsskalor leder till en lägre spridning för den yngre gruppen. Eftersom skalorna är helt olika visar sig denna skillnad dock inte i en lägre standardavvikelse för den yngre gruppen (jfr tabellerna 11 och 12), men om man ser till fördelningarna över de olika betygsstegen blir den tydlig. Som exempel kan nämnas att inte mindre än 85 procent av samtliga i den yngre gruppen återfinns på de två mellanstegen ”Godkänd” och ”Väl godkänd”. I den äldre gruppen fördelar sig 90 procent av samtliga på de tre mellanstegen 2, 3 och 4.

En jämförelse mellan könen visar vidare att sambanden är något högre för kvinnor än för män och en jämförelse mellan de olika delproven visar att LÄS-provet i allmänhet uppvisar de högsta sambanden med betyg, medan de två kvantitativa delproven DTK och NOG har lägst samband.

När det gäller den allmänna nivån på sambanden är dessa betydligt högre än de samband som Svensson, Gustafsson och Reuterberg (2001) fann mellan högskoleprovsresultat och erhållna poäng efter det första studieåret på civilingenjörutbildningar, juristutbildningar och lärarutbildningar. De är även fullt i nivå med de samband som gymnasiebetygen uppvisade med motsvarande mått på studieframgång inom de utbildningar som hade de högsta sambanden, nämligen civilingenjörutbildningarna. Detta ger enligt min mening stöd för antagandet att grundskolebetygen fungerar väl som indikator på den allmänna studieförmågan.

Kontroll av selektionseffekter till högskoleprovet

De som genomgår högskoleprovet är naturligtvis personer som planerar att studera på högskolan och därigenom utgör de en positivt selekterad grupp vad gäller tidigare skolprestationer. Som visats av Mäkitalo och Reuterberg (1996) och Reuterberg (1998) är dessutom denna selektionsprocess olika stark för män och kvinnor och olika stark också i skilda socialgrupper. Denna differentiella selektion till provet leder till att de skillnader mellan olika grupper man finner i provtagargruppens resultat blir annorlunda än de man skulle finna om samtliga i populationen genomgått högskoleprovet. Så kan t.ex. en eventuell bias i provet maskeras av denna differentiella selektion. Vidare kommer sambandet mellan kontrollvariabler och provresultat att påverkas så att

det blir lägre inom provtagargruppen jämfört med i hela populationen. (Jfr tabellerna 9 och 10.)

Det kan här finnas skäl att närmare granska denna differentiella selektion till högskoleprovet. Som mått på denna selektion tas här summan av de betyg som konstituerar den generella studieförmågan.

Tabell 11. Medelvärden och standardavvikelser i generell studieförmåga samt selektion till högskoleprovet. Födda 1977.

	Population		Provtagare		Selektion
	M	S	M	S	Effect size
SS	33,30	8,24	39,83	5,36	0,79
SU	31,81	8,40	39,17	5,72	0,89
IS	30,90	7,56	37,69	5,57	0,82
IU	30,12	8,71	38,50	6,01	1,01
Samtliga	33,09	8,29	39,75	5,40	0,80

Tabell 12. Medelvärden och standardavvikelser i generell studieförmåga samt selektion till högskoleprovet. Födda 1982.

	Population		Provtagare		Selektion
	M	S	M	S	Effect size
SS	124,95	36,74	151,59	24,30	0,72
SU	115,73	40,38	145,62	25,66	0,80
IS	115,71	34,09	143,56	26,19	0,75
IU	114,11	41,39	146,57	26,47	0,87
Samtliga	124,03	37,15	151,09	24,50	0,70

Tabellernas sista kolumn, kallad ”Effektstorlek”, anger hur mycket högre provtagarnas medelvärde är jämfört med populationens. Skillnaderna uttrycks här som standardavvikelseenheter. Ju högre värde desto mer positivt selekterad är gruppen. För den äldre årskullen uppgår skillnaden mellan provtagare och population till 0,8 standardavvikelseenheter, medan motsvarande skillnad är 0,7 i den yngre. I båda fallen har vi följaktligen en stark positiv selektion till högskoleprovet – något starkare i den äldre än i den yngre kohorten trots att andelen provtagare var klart lägst i den senare gruppen. En möjlig förklaring både till den lägre andelen provtagare och den svagare se-

lektionen bland 82:orna är att denna grupp haft möjlighet att komplettera sina gymnasiebetyg medan, som tidigare nämnts, så inte varit fallet för den äldre gruppen.

En intressant iakttagelse i detta sammanhang är att den äldre kohorten uppvisar den starkaste selektionen och samtidigt når det högsta resultatet på provet (tabell 4 och 5). Detta innebär att i varje fall en del av skillnaderna kan förklaras med skillnader i selektion.

Vi har emellertid inte enbart en positiv selektion till högskoleprovet utan också en differentiell selektion, där grupperna SS och IS är mindre starkt selekterade än övriga grupper. Den starkaste positiva selektionen sker i båda kohorterna inom grupp IU. Denna differentiella selektion leder i sin tur till att medelvärdesskillnaderna mellan de fyra grupperna är klart lägre bland provtagare än bland samtliga i populationen. Det sistnämnda visar att en studie av eventuell bias i högskoleprovet baserad enbart på provtagargruppen skulle väsentligt underskatta denna bias.

Vi kan också konstatera att grupp SS uppnår de klart högsta resultaten på högskoleprovet trots att provtagarna i denna grupp är den minst positivt selekterade. Detta innebär att de skillnader i provresultat som visats i tabellerna 4 och 5 utgör en klar underskattning av de skillnader vi skulle finna om samtliga i populationen genomgått provet.

Som nämnts ovan har Mäkitalo och Reuterberg (1996) visat att männen är mer positivt selekterade än kvinnorna till högskoleprovet. Det kan därför finnas skäl att granska huruvida det föreligger könsskillnader också i den differentiella selektionen.

Tabell 13. Könjämförelser beträffande selektion till högskoleprovet. Effect size.

Grupp	Födda 1977		Födda 1982	
	Män	Kvinnor	Män	Kvinnor
SS	0,92	0,65	0,82	0,61
SU	1,07	0,72	0,92	0,69
IS	0,97	0,72	0,92	0,61
IU	1,09	0,92	0,94	0,81
Samtliga	0,93	0,67	0,83	0,62

Som framgår av tabell 13 har den positiva selektionen till högskoleprovet försvagats över tid för både män och kvinnor även om den är påtaglig även i den yngre gruppen. I likhet med Mäkitalo och Reuterberg (1996) kan vi också konstatera att männen är starkare selekterade än kvinnorna inom båda kohorterna. Det faktum att männen är starkast selekterade till provet visar att i

varje fall en del av skillnaderna i provresultat får tillskrivas den differentiella selektionen till provet.

Selektiva bortfall drabbar de flesta empiriska undersökningar. När det gäller den differentiella selektionen mellan personer med olika invandrarstatus ter sig mönstret något olika för männen och kvinnorna. Bland kvinnor födda 1977 är selektionen tämligen svag inom grupp SU, vilket gör att den där är av samma styrka som för grupp IS. Däremot kvarstår för båda könen att selektionen är svagast i grupp SS och starkast i grupp IU. Bland dem födda 1982 har vi en förhållandevis stark selektion bland män från grupp IS och samtidigt en svag selektion bland kvinnor i samma grupp. Detta leder bland männen till att selektionen här når samma styrka som för grupperna SU och IU, medan kvinnornas selektion sjunkit till samma nivå som den för kvinnor i grupp SS.

Även om således selektionseffekterna varierar mellan olika grupper är de genomgående så starka att man måste ta hänsyn till dem vid analys av eventuell bias i högskoleprovet. För att undvika en sådan påverkan bör analyserna baseras också på en tänkt situation där samtliga i populationen genomgått högskoleprovet. Med hjälp av modellering av inkompleta data (s.k. "missing data analysis") är det numera möjligt att skapa en sådan situation (Allison, 1987; Muthén, Kaplan & Hollis, 1987). I princip innebär modellering av inkompleta data att man utifrån tillgängliga uppgifter skattar de uppgifter som saknas i datamatriken. I föreliggande undersökning har hela undersökningsgruppen bakgrundsuppgifter samt fullständiga avgångsbetyg från grundskolan. Däremot saknas helt naturligt högskoleprovsresultat för alla dem som inte genomgått provet. Eftersom det finns ett klart samband mellan avgångsbetygen och resultaten på högskoleprovets generella faktorer är det möjligt att skatta vilka resultat en person som inte genomgått högskoleprovet skulle ha fått om han/hon genomgått provet. Genom att komplettera den ursprungliga datamatriken med dessa skattade värden erhålls en komplett datamatrix. Vi får med andra ord en datamatrix som visar högskoleprovsresultat för alla i populationen. Trots att metoden funnits tillgänglig länge har den inte tillämpats i någon större utsträckning. En anledning till detta är att den kräver en ganska komplex specifikation av modellerna. Tack vare att Jan-Eric Gustafsson och Per Arne Stahl utvecklat ett pre- och postprocess program (för- och efterbearbetningsprogram) har tillgängligheten i hög grad breddats (Gustafsson & Stahl, 2000). Programmet heter STREAMS (Structural equation modeling made simple) och är utformat så att det kan utnyttjas av såväl noviser som erfarna forskare. De statistiska bearbetningarna har genomförts i LISREL 8 (Jöreskog & Sörbom, 1996).

Bortfall är som tidigare påpekats alltid ett stort problem vid empiriska undersökningar. När det gäller validering av urvalstest har man alltid haft problemet att kriterie-måtten saknas för alla dem som inte blir uttagna och det gäller givetvis också för de data som använts i den här studien. Detta medför normalt att de kriteriemått som är kända drabbats av "beskuren variation" vil-

ket i sin tur medför en underskattning av korrelationerna mellan testresultat och kriteriemått (Gross & McGanney, 1987; Gulliksen, 1950; Linn, 1983; Lord & Novick, 1968; Willingham m.fl., 1990). Sedan länge har det funnits korrektionsmetoder för beskuren variation, vilka ger goda estimat av populationens validitet under förutsättning att vissa villkor är uppfyllda (Gross & Gannery, 1987; Holmes, 1990; Linn, 1983; Lord & Novick, 1968). Som diskuteras av Gustafsson och Reuterberg (2000) sker urvalet till högre utbildning i Sverige emellertid enligt ett kompensatoriskt urvalsförfarande vilket ytterligare komplicerar bilden såtillvida att de traditionella korrektionsmetoderna inte är tillämpbara. Dock kan man med hjälp av modellering av inkompleta data numera justera för sådana felfaktorer och erhålla säkra skattningar av validiteten även om urvalet sker enligt en kompensatorisk urvalsmodell (Gustafsson & Reuterberg, 2000; Reuterberg, manus).

Medelvärdesskillnader i delprovens generella faktorer

Först skall gruppskillnader i delprovens generella faktorer studeras. Om man finner skillnader i dessa faktorer som är av samma natur som de skillnader Reuterberg och Hansen (2001) fann beträffande de manifesta delprovresultaten visar detta att gruppskillnaderna i varje fall inte helt kan förklaras av bias på uppgiftsnivå. De generella faktorerna är ju uttryck för den bakomliggande förmåga som delprovet som helhet mäter och skillnader i dessa faktorer svarar normalt för en stor del av de skillnader som gäller för delprovresultaten eftersom de generella faktorerna slår igenom på samtliga uppgifter i provet. Som visats av Reuterberg och Gustafsson (1992) ökar ju en faktors betydelse för totalskillnaderna med kvadraten på det antal uppgifter den påverkar.

Innan gruppskillnaderna studeras på faktornivå finns det skäl att granska de faktiska medelvärdesskillnader som föreligger mellan de fyra grupperna för vart och ett av delproven. Dessa medelvärden redovisas i tabellerna 14 och 15. I tabell 16 sammanfattas skillnadernas storlek med hjälp av intraklasskorrelationer. Korrelationerna ger ett sammanfattande mått på skillnadernas omfattning. I bilaga 1 redovisas motsvarande skillnader separat för manliga och kvinnliga provtagare.

Tabell 14. Delprovsmedelvärden för provtagare med olika invandrarstatus. Födda 1977.

Grupp	ORD	NOG	LÄS	DTK	ELF
SS	23,37	14,51	12,08	11,94	13,05
SU	20,38	12,58	11,00	10,56	13,15
IS	22,30	11,69	11,16	9,83	11,69
IU	19,46	12,55	10,76	10,22	12,88
Totalt	23,17	14,37	12,00	11,83	13,03

Tabell 15. Delprovsmedelvärden för provtagare med olika invandrarstatus. Födda 1982.

Grupp	ORD	NOG	LÄS	DTK	ELF
SS	19,72	11,97	12,80	11,68	11,39
SU	17,68	10,28	11,50	10,05	10,58
IS	17,55	9,62	11,46	9,49	9,79
IU	16,91	9,78	11,18	9,55	10,39
Totalt	19,52	11,80	12,68	11,52	11,31

Tabell 16. Sammanfattning av medelvärdeskillnaderna bland samtliga provtagare. Intraklasskorrelationer.

Födelseår	ORD	NOG	LÄS	DTK	ELF
1977	.120	.112	.082	.117	.043
1982	.105	.134	.109	.144	.068

Liksom Reuterberg och Hansen (2001) fann, är skillnaderna mindre på ELF-provet än på övriga prov och allra minst är de bland dem födda 1977 (tabell 16). Störst skillnader visar däremot NOG och DTK i båda kohorterna och för båda proven gäller att de är något större i den yngre gruppen än i den äldre. Även på LÄS har skillnaderna ökat något över tid, medan de däremot minskat något på ORD, vilket gör att skillnaderna på dessa två prov är av samma storleksordning i den yngre gruppen.

Anledningen till att ELF-provet uppvisar så små skillnader i den äldre kohorten ser vi i tabell 14, där grupp SU har ett något högre medelvärde än grupp SS och grupp IU ett medelvärde som ligger aningen lägre. I övrigt har SS-gruppen genomgående ett klart högre medelvärde än övriga grupper och mellan dessa är skillnaderna i allmänhet små och deras inbördes rangordning varierar från prov till prov. Värt att observera är emellertid att grupp IS genomgående har det lägsta medelvärdet på de två kvantitativa proven NOG och DTK samt att IU ligger lägre än övriga grupper på ORD och LÄS.

Frågan är nu om de skillnader som visats ovan har sitt ursprung i den generella delprovsfaktorn, dvs. den bakomliggande förmåga som delprovet som helhet mäter, eller om de orsakas av sådana mer specifika faktorer som enskilda uppgifter eller grupper av uppgifter mäter. För att identifiera en generell delprovsfaktor för varje prov har provets uppgifter grupperats i tre uppgiftspaket, vilka benämnes med provets namn med ett tillägg av bokstäverna A, B

eller C. De tre uppgiftspaketen för ORD har således beteckningarna ORDA, ORDB respektive ORDC.

Uppdelningen av de enskilda uppgifterna på de olika uppgiftspaketen har skett enligt nedan:

Paket A: uppg. 1, 4, 7, 10 osv.

Paket B: uppg. 2, 5, 8, 11 osv.

Paket C: uppg. 3, 6, 9, 12 osv.

Genom att summera poängen för de uppgifter som ingår i ett och samma paket kommer det totala delprovresultatet att delas upp på tre delskalor, vilka benämns på samma sätt som uppgiftspaketen. De tre delskalor som tillhör samma prov definierar den generella delprovsfaktorn enligt figur 1, nedan.

Analyserna av medelvärdeskillnaderna har sedan genomförts så att de fyra grupper som anger invandrarstatus kodats som "dummyvariabler", vilket innebär att de kan antaga endast värdena 1 eller 0. Härigenom får vi tre nya variabler SUdum, ISdum och IUdum enligt nedanstående tablå (Pedhazur och Pedhazur Schmelkin, 1991, s. 465).

Grupp	SUdum	ISdum	IUdum
SS	0	0	0
SU	1	0	0
IS	0	1	0
IU	0	0	1

Dessa dummyvariabler sätts sedan in i en regressionsekvation. Om vi tar exemplet ORD blir regressionsekvationen följande:

$$\text{ORD} = b_1 * \text{SUdum} + b_2 * \text{ISdum} + b_3 * \text{IUdum} + \text{Konstant}$$

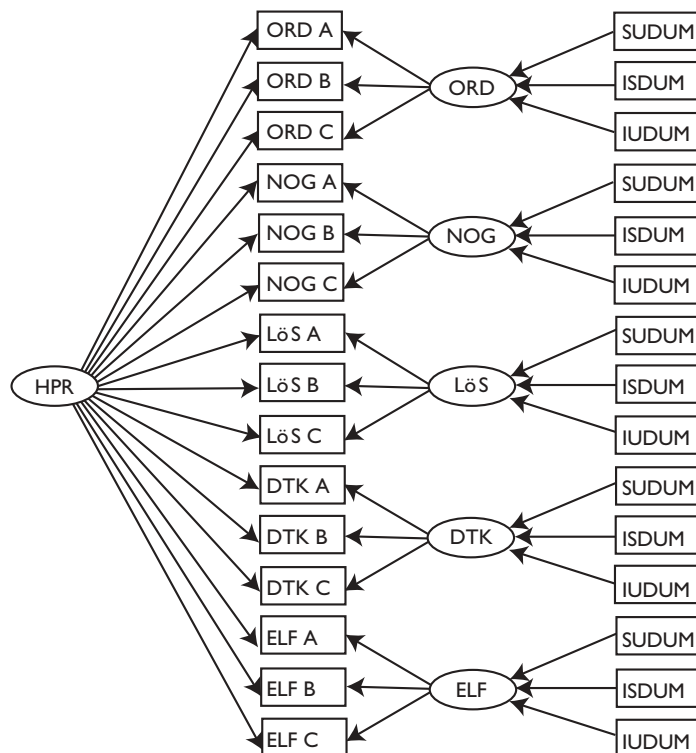
Grupp SS som har värdet 0 i samtliga tre dummyvariabler kommer att utgöra referensgrupp. Härigenom anger den ostandardiserade regressionskoefficienten b_1 medelvärdeskillnaden i den generella delprovsfaktorn mellan grupperna SS och SU, b_2 skillnaden mellan SS och IS samt b_3 skillnaden mellan SS och IU.

Förutom de faktiska skillnaderna kommer jag att studera också om dessa skillnader kan förklaras av de betygsskillnader som finns grupperna emellan. Detta sker genom att komplettera den ovanstående regressionsekvationen med den generella betygsfaktorn (Bet). Denna faktor definieras av samtliga betyg som ingår i den generella studieförmågan.

Regressionsekvationen blir i detta fall:

$$\text{ORD} = b_1 * \text{SUdum} + b_2 * \text{ISdum} + b_3 * \text{IUdum} + b_4 * \text{Bet} + \text{Konstant}$$

Den modell som ligger till grund för beräkning av medelvärdesskillnaderna i var och en av de generella delprovsfaktorerna bland provtagare beskrivs i figur 1.



Figur 1. Modell för beräkning av medelvärdesskillnader i de generella delprovsfaktorerna bland provtagare.

Figuren 1 visar att de generella delprovsfaktorerna ORD, NOG osv. definieras av de delskalor som bildats genom att föra samman vissa uppgifter till uppgiftspaket (jfr ovan). På samma sätt definieras den latenta variabeln BETYG av samtliga betyg som ingår i "studieförmåga". För varje delprov har skillnaderna i delprovsfaktorerna beräknats. Detta sker genom att ha med endast de tre dummyvariablerna SUDUM, ISDUM samt IUDUM. Vid beräkningen av skillnaderna i delprovsfaktorerna med kontroll för betygsskillnader har även den latenta variabeln BETYG ingått som oberoende variabel i modellen. Beräkningarna har genomförts för varje delprov separat dels för provtagargruppen, dels för populationen. Det sistnämnda sker liksom tidigare med hjälp av modellering av inkompleta data.

De ostandardiserade regressionskoefficienterna är skalberoende vilket innebär att de ej kan direkt jämföras över de fem delproven. Därför kommer de att divideras med standardavvikelsen för respektive delprovsfaktor i populationen. Härigenom får vi ett mått på skillnaderna "effektstorlek" som är jämförbart över delproven. Dessa mått kallas fortsättningsvis standardise-

rade medelvärdeskillnader. Som påpekas av Bobko, Roth och Bobko (2001) är emellertid dessa mått allmänt sett påverkade av såväl reliabilitetsbrister som varianshomogenitet vilket begränsar deras jämförbarhet. I detta fall är reliabilitetsbrister inga problem eftersom jämförelserna avser latent variabler som är fria från mätfel. Däremot kommer varianshomogeniteten att påverka de jämförelser som sker bland provtagarna eftersom de olika grupperna är olika starkt selekterade. Detta problem får emellertid sin lösning i och med att skillnaderna studeras också på populationsnivå.

När det gäller skillnaderna i de generella delprovsfaktorerna kommer två resultatbilder att redovisas. En gäller de *skillnader som finns bland provtagarna*. Dessa skillnader är så att säga bruttoskillnader i den meningen att de är påverkade av såväl betygskillnader som differentiell selektion till högskoleprovet. Dessa två påverkansfaktorer verkar i princip i olika riktning. Betygsskillnaderna till fördel för grupp SS förstärker skillnaderna i högskoleprovsresultat till fördel för denna grupp framför de övriga, men den differentiella selektionen verkar i motsatt riktning eftersom samtliga övriga gruppers provtagare var strängare selekterade till provet än de provtagare som kommer från grupp SS.

Den andra resultatbilden som skall redovisas är *den för populationen då vi kontrollerat för både betygskillnader och differentiell selektion till provet*. Vid beräkningarna av medelvärdeskillnaderna med kontroll av betyg har också den latent variabeln BETYG ingått i modellen. Den definieras av samtliga de betyg som visats i tabellerna 4 och 5, och på samma sätt som dummyvariablerna påverkar den var och en av generella delprovsfaktorerna.

I princip hade det varit möjligt att redovisa också de resultat som gäller för provtagarna då betygsskillnaderna hålls under kontroll. Denna bild har emellertid så stora likheter med den som skall redovisas för populationen att det inte finns anledning att tynga framställningen med båda dessa resultat.

Tabell 17. Standardiserade medelvärdeskillnader mellan grupp SS och övriga grupper i de generella delprovsfaktorerna bland provtagarna.

Grupp	Födda 1977					Födda 1982				
	ORD	NOG	LÄS	DTK	ELF	ORD	NOG	LÄS	DTK	ELF
SU	-.434	-.380	-.281	-.391	+0.026	-.320	-.423	-.383	-.481	-.195
IS	-.155	-.555	-.265	-.599	-.354	-.350	-.576	-.290	-.641	-.382
IU	-.565	-.385	-.348	-.490	-.046	-.453	-.554	-.491	-.619	-.238

Nästan samtliga medelvärdeskillnader är negativa vilket innebär att grupp SS har ett högre värde än de övriga grupperna. Det enda undantaget från denna tendens är att grupp SU har ett något högre medelvärde än SS på det engelska läsförståelseprovet (ELF) bland dem födda 1977. Skillnaden är dock

ytterst liten vilket gör att slutsatsen blir att det inte föreligger någon skillnad mellan de två grupperna på detta prov. Samma slutsats gäller också för grupp IU på samma prov där skillnaden är så liten som -0.046 .

Den generella slutsatsen av resultaten blir således att på alla prov utom ELF för 77:orna har grupp SS det högsta medelvärdet i de generella delprovsfaktorerna. Detta är den mest väsentliga orsaken till att grupp SS normalt överträffar samtliga övriga grupper i de observerade delprovresultaten. Av detta följer att gruppskillnaderna i observerade provresultat inte på långt när kan förklaras med att vissa enskilda uppgifter har sådana inslag som på ett orättmätigt sätt gynnar grupp SS.

Om man tar genomsnittsvärdet för skillnaderna i tabell 17 som ett övergripande mått på skillnaderna visar detta att grupp SU är den grupp vars prestationer kommer närmast grupp SS, medan grupperna IS och IU presterat lägst i båda årskullarna. På det hela taget är skillnaderna till grupp SS större bland dem födda 1982 än bland 77:orna, vilket är naturligt mot bakgrund av att den differentiella selektionen till provet varit starkast i den äldre åldersgruppen. Som tidigare nämnts tenderar ju den differentiella selektionen att reducera skillnaderna grupperna emellan.

Skillnaderna varierar dock påtagligt mellan de olika delproven. Mest överlägsen är grupp SS på de två kvantitativa proven NOG och DTK, medan ELF är det prov, där de övriga tre grupperna kommer närmast grupp SS och i synnerhet gäller det sistnämnda för dem födda 1977. Även för 82:orna är ELF det prov som ger upphov till minst avvikelser från grupp SS, men skillnaderna går här genomgående till fördel grupp SS. Det faktum att ELF uppvisar minst skillnader beror i första hand på att grupperna SU och IU presterat bra på detta prov. Däremot har IS ett förhållandevis lågt värde.

Värt att observera är också det förhållandet att grupp IU uppvisar de klart lägsta medelvärdena av samtliga grupper på såväl ORD-provet som LÄS-provet. Även grupp IS visar en intressant profil över skillnaderna på de olika delproven. Gruppen har nämligen klart lägre värden på de kvantitativa proven än på de språkliga.

De resultat som kommenterats här gäller som nämnts provtagarna och skillnaderna påverkas därigenom av både selektionseffekter och skillnader i betyg. I tabell 18 redovisas motsvarande skillnader, men nu under antagandet om att samtliga i populationen genomgått högskoleprovet dvs. selektionseffekterna är eliminerade, liksom betygsskillnaderna. Genom att låta betygen ingå som oberoende variabel i regressionskvationen hålls också betygsskillnaderna under kontroll.

Tabell 18. Standardiserade medelvärdeskillnader mellan grupp SS och övriga grupper i de generella delprovsfaktorerna inom populationerna med betyg under kontroll.

Grupp	Födda 1977					Födda 1982				
	ORD	NOG	LÄS	DTK	ELF	ORD	NOG	LÄS	DTK	ELF
SU	-.380	-.324	-.217	-.326	+0.083	-.220	-.324	-.270	-.371	-.098
IS	+0.009	-.386	-.082	-.418	-.188	-.223	-.452	-.261	-.505	-.264
IU	-.463	-.278	-.230	-.371	+0.061	-.371	-.476	-.402	-.526	-.161

Då vi skattar avvikelserna från grupp SS för populationerna och med betygen under kontroll minskar grupp SS överlägsenhet i många fall påtagligt. Fortfarande är dock de allra flesta skillnaderna negativa, vilket visar att grupp SS i de allra flesta fall presterat högre än övriga grupper även då selektionseffekter och betygsskillnader kontrollerats. För 82:orna gäller detta alla jämförelser, men för 77:orna återfinns tre positiva differenser, nämligen för grupperna SU och IU på ELF samt för grupp IS på ORD. I samtliga dessa undantagsfall är skillnaderna dock fortfarande små, varför man kan säga att grupperna SU, IS och IU här presterat i nivå med grupp SS. Samma slutsats kan man också dra för grupp IS på LÄS-provet bland dem födda 1977.

Bland provtagarna fanns en tydlig rangordning mellan de olika grupperna såtillvida att SU var den grupp som låg närmast SS, medan grupperna IS och IU hade lägst prestationer. Denna rangordning gäller också för populationsvärdena, men enbart för dem födda 1982. Bland 77:orna är skillnaderna däremot små mellan grupperna SU, IS och IU, men alla tre grupperna ligger fortfarande klart lägre än SS.

När det gäller avvikelserna från grupp SS på de enskilda delproven är resultatbilden densamma som bland provtagarna. Så är skillnaderna klart störst på de kvantitativa delproven NOG och DTK och minst på ELF. Stora skillnader visar också ORD bland 77:orna.

Även gruppernas resultatprofiler över de olika proven kvarstår i stort sett oförändrade. Så presterar grupp IS klart bättre på de språkliga proven än på de kvantitativa. Som nämnts tidigare ligger gruppens populationsvärden i stort sett i nivå med grupp SS på både ORD och LÄS inom den äldre åldersgruppen. Likaså presterar grupp IU även inom populationerna de lägsta resultaten av alla grupper på ORD och LÄS även om skillnaden till grupp SU är liten på LÄS bland dem födda 1977.

Sammanfattningsvis kan vi konstatera att:

- Grupp SS överträffar de övriga grupperna i samtliga generella delprovsfaktorer, utom den för ELF, bland de provtagare som är födda 1977. Enligt resultaten i tabell 16 kan dock dessa gruppskillnader förklaras av

selektionseffekter och skillnader i betyg i två fall, nämligen då det gäller grupp IS för ORD- och LÄS-faktorerna, där gruppens resultat ligger i nivå med grupp SS.

- ELF-faktorn uppvisar klart lägre gruppskillnader än övriga faktorer och bland 77:orna presterar här grupperna SU och IU i stort sett i nivå med grupp SS, medan IS ligger klart lägre. De största skillnaderna uppvisar däremot NOG- och DTK-faktorerna.
- Grupp SU är den grupp som ligger närmast SS då det gäller resultaten i stort för provtagarna. Men då populationsvärdena studeras med kontroll för selektionseffekter och betygsskillnader är skillnaderna mellan grupperna SU och IU förhållandevis små i den äldre kohorten. IU är dock den grupp som uppnår de lägsta medelvärdena i den yngre kohorten.
- Grupp IS uppvisar en tydlig resultatprofil över de fem delprovsfaktorerna såtillvida att gruppen presterar klart lägre på de kvantitativa faktorerna än på de språkliga. Grupp IU uppvisar tvärtom låga resultat på ORD- och LÄS-faktorerna.

När det gäller den sistnämnda slutsatsen – den om grupp IS:s tydliga profil över de olika delproven – ligger det nära till hands att misstänka att vi har en könsfaktor inblandad. Det är ett väl känt faktum att såväl NOG som DTK är de prov som ger upphov till stora könsskillnader till männens fördel. Av tabell 1 har framgått att grupp IS bland dem födda 1977 är den grupp som har den snedaste könsfördelningen med 59 procent kvinnor i undersökningsgruppen och 65 procent kvinnor bland provtagarna. Även i den yngre kohorten är kvinnorna i majoritet inom grupp IS med 53 procent i undersökningsgruppen och 57 procent bland provtagarna.

För att kontrollera om det är könsfördelningen som ligger bakom resultatprofilen för grupp IS har medelvärdena för delprovsfaktorerna beräknats separat för män och kvinnor. Resultaten återfinns i Bilaga I. Om vi först ser till 77:orna kan vi där konstatera att profilen är densamma för män och kvinnor. För båda könen gäller nämligen att avvikelserna från grupp SS är störst på just NOG och DTK. Bland 82:orna, där resultatprofilen inte är lika uttalad, uppträder samma profil, men den är tämligen svag för kvinnorna. Av detta kan vi dra slutsatsen att den uttalade resultatprofilen för grupp IS inte helt kan förklaras av den sneda könsfördelningen inom grupp IS.

Vad betyder ålder för invandring för medelvärdena i de generella delprovsfaktorerna?

Enligt den undersökning som genomfördes av Reuterberg och Hansen (2001) finns en tydlig skillnad i provresultat mellan dem som invandrat till Sverige före 10 års ålder och de som kommit senare i livet. Den förstnämnda gruppens högsta normerade provresultat uppgick till 0,82 poäng mot 0,57 för dem som invandrat sent. Mot denna bakgrund är det rimligt att anta att de

förhållandevis låga resultaten på de generella delprovsfaktorerna för grupp IU åtminstone delvis kan vara en effekt av att många i denna grupp kommit till Sverige i 10-årsåldern eller senare.

För att kontrollera vilken betydelse detta har haft för resultaten i föregående avsnitt har grupp IU delats i två grupper, de som kommit innan 10 års ålder – grupp IU1 – och de som kommit senare i livet – grupp IU2. De analyser som skall visa vilken betydelse invandringsålder har, är av samma natur som i föregående avsnitt med den skillnaden att vi nu har fyra gruppers resultat vilka skall jämföras med de för grupp SS.

Tyvär visade det sig att den grupp som invandrat sent bland 82:orna är så liten att den skulle ge alltför osäkra resultat. Analyserna genomförs därför enbart för dem födda 1977. Dock kommer resultaten liksom tidigare att redovisas dels för provtagarna, dels för populationen då såväl selektionseffekter som betygsskillnader hålls under kontroll.

Tabell 19. Standardiserade medelvärdeskillnader mellan grupp SS och övriga fyra grupper i de generella delprovsfaktorerna. Födda 1977.

Grupp	Provtagare					Population				
	ORD	NOG	LÄS	DTK	ELF	ORD	NOG	LÄS	DTK	ELF
SU	-.431	-.380	-.291	-.385	+0.025	-.379	-.318	-.216	-.323	+0.086
IS	-.153	-.556	-.287	-.589	-.346	+0.007	-.384	-.086	-.414	-.185
IU1	-.458	-.328	-.322	-.466	+0.036	-.338	-.221	-.189	-.357	+0.139
IU2	-.760	-.500	-.459	-.505	-.204	-.668	-.387	-.299	-.365	-.073

Grupp IU2 – de som kommit till Sverige i 10-årsåldern eller senare – har genomgående lägre resultat än i grupp IU1 som kommit tidigare, vilket innebär att ålder för invandring är en delförklaring till att grupp IU på det hela taget har låga medelvärden i de generella delprovsfaktorerna. Ser vi till provtagarna presterar IU1 nästan i nivå med eller bättre än SU på alla delprov utom på DTK och detta blir ännu tydligare om vi går till resultaten för populationen. Där ligger nämligen IU1 något högre än SU på alla delprovsfaktorer utom DTK. Däremot ligger den del av grupp IU som invandrat sent i livet (IU2) allra lägst på ORD och LÄS, vilket är naturligt med tanke på att de haft begränsade möjligheter att lära sig svenska språket. De ligger förhållandevis lågt också på NOG och DTK, där deras resultat är tämligen lika de för grupp IS. Detta är i och för sig förklarligt med tanke på att språkproblem hämmar deras resultat också på dessa två prov.

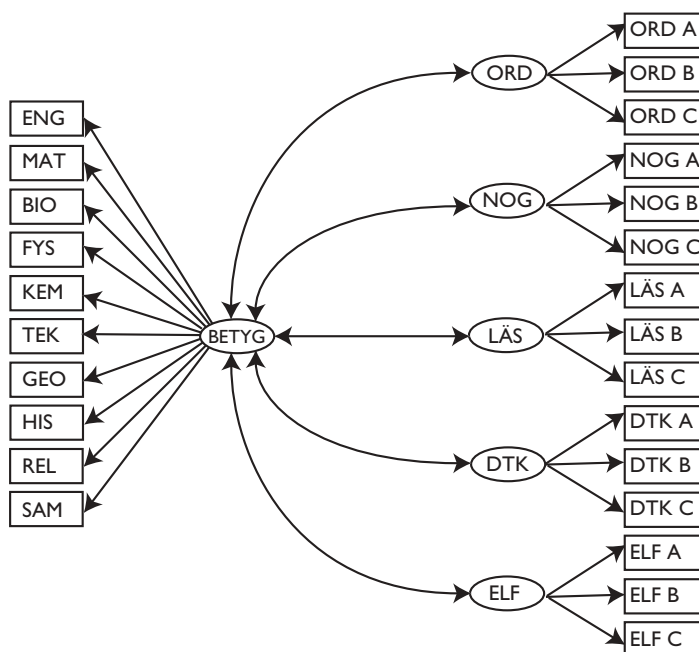
Vi kan således slå fast att medelvärdena på delprovsfaktorerna för grupp IU delvis beror på att gruppen innefattar många som kommit till Sverige relativt sent i livet.

Skillnader i samband mellan delprovets generella faktorer och allmän studieförmåga

I föregående avsnitt kunde vi konstatera att det finns betydande skillnader mellan de olika grupperna vad gäller medelvärdena för de olika delprovets generella faktor. I detta avsnitt är avsikten att granska om sambanden mellan dessa generella faktorer och den generella betygsfaktorn varierar grupperna emellan. Om det finns betydande gruppkillnader i detta avseende är det tecken på att delprovets mätegenskaper skiljer sig från en grupp till en annan. Detta kan vara ett tecken på att de olika delproven har varierande validitet grupperna emellan, vilket i sin tur kan indikera att provet missgynnar vissa kategorier av provtagare.

Analyserna har genomförts så att de tio ämnesbetyg som återfinns i tabellerna 3 och 4 tillsammans konstituerar den generella betygsfaktorn. De olika delprovsfaktorerna har definierats genom de tre skalor som bildats för varje delprov. Därefter har korrelationerna mellan betygsfaktorn och delskalornas faktorer beräknats separat för varje grupp och varje deltest.

Modellen visas i figur 2.



Figur 2. Modell för analys av skillnader i samband mellan den generella betygsfaktorn och BETYG.

Modellen upprepas sedan för var och en av de övriga delproven.

Tabell 20. Korrelationer mellan den generella betygsfaktorn och delskalornas generella faktorer. Födda 1977.

	SS	SU	IS	IU
DTK	.701	.661	.663	.719
NOG	.695	.713	.687	.690
LÄS	.758	.723	.722	.711
ORD	.697	.700	.675	.626
ELF	.703	.682	.693	.673

Tabell 21. Korrelationer mellan den generella betygsfaktorn och delskalornas generella faktorer. Födda 1982.

	SS	SU	IS	IU
DTK	.646	.590	.619	.710
NOG	.645	.594	.723	.645
LÄS	.644	.695	.726	.687
ORD	.642	.724	.637	.695
ELF	.625	.699	.702	.667

Om man tar de genomsnittliga korrelationerna i tabellerna 20 och 21 som övergripande mått på sambandens styrka kan vi konstatera att sambanden på det hela taget är något högre bland 77:or än bland 82:or. Skillnaden är dock obetydlig – trehundra delar. I den äldre kohorten är det LÄS-faktorn som uppvisar högst samband med betyg med korrelationer som genomgående ligger över 0,70. I den yngre gruppen skiljer sig dock LÄS obetydligt från de övriga språkliga delprovsfaktorerna.

En motsvarande jämförelse mellan de olika grupperna visar att sambanden är något högre för grupp SS än för övriga grupper bland 77:orna. I den yngre gruppen är det däremot SS som uppvisar de lägsta sambanden. Man kan därför knappast finna några entydiga tecken på att delprovsfaktorernas samband med betyg varierar på ett systematiskt sätt.

Även om sambandens styrka varierar en del också mellan de olika grupperna på delprovsnivå finns det inte heller här tecken på någon systematisk variation. Det mest avvikande värdet bland 77:orna är det låga sambandet mellan ORD och BETYG för grupp IU. Går man till 82:orna får emellertid motsvarande samband snarast betecknas som högt, varför det inte finns an-

ledning att dra några allmängiltiga slutsatser om att ORD-provets samband med betygen är lågt.

På motsvarande sätt finns det några avvikande samband bland 82:orna. Det gäller DTK som uppvisar ett lågt samband med BETYG inom grupp SU och ett högt samband inom grupp IU. Vidare visar också NOG ett lågt samband inom grupp SU, men ett högt inom grupp IS. Inte heller dessa avvikelser har någon motsvarighet i den andra kohorten, varför de inte bör tillmätas någon större betydelse.

Slutsatsen blir således att det inte finns några tecken på att sambandet mellan delprovets generella faktorer och allmän studieförmåga varierar på ett systematiskt sätt över de olika grupperna.

På jakt efter biasfaktorer

Det faktum att det inte finns några systematiska skillnader i delprovsfaktorernas samband med allmän studieförmåga innebär inte i sig att delproven är fria från bias. Det finns därför anledning att gå vidare och göra mera noggranna analyser. Detta måste emellertid ske på två olika sätt eftersom provens uppbyggnad varierar. ORD, NOG och DTK är ju i huvudsak uppbyggda av fristående uppgifter, medan LÄS och ELF består av grupper av uppgifter som är knutna till en given text. Den senare uppbyggnaden innebär att man förutom en generell faktor eventuellt kan identifiera också mer specifika innehållsfaktorer som återspeglar provtagarens förmåga att förstå det specifika innehåll som texten förmedlar. Dessutom kan man i varje fall teoretiskt tänka sig att det finns biasfaktorer som inte är avgränsade till en viss text.

Innehåller ORD, NOG och DTK biasfaktorer?

Med en biasfaktor menas här att den definieras av ett antal uppgifter som dels uppvisar klara skillnader i lösningsfrekvens till fördel för grupp SS jämfört med övriga grupper, dels att uppgifterna har en gemensam kovarians utöver den som betingas av delprovets generella faktor. Att identifiera enskilda biasuppgifter är i och för sig av intresse, men av klart större intresse är att identifiera biasfaktorer. Anledningen till detta är att biasfaktorerna får en avsevärt större betydelse för gruppskillnaderna i provresultat eftersom de ger större bidrag till provets totala varians. Dessutom ger faktorerna bättre förutsättningar för att tolka och finna förklaringar till en eventuell bias, än vad som är möjligt utifrån enskilda uppgifter vars faktorstruktur är okänd.

Sökandet efter biasfaktorer har i detta fall inneburit följande steg:

- Den hälft av uppgifterna i respektive delprov som uppvisar störst skillnader till fördel för grupp SS har identifierats.
- Med konfirmatorisk faktoranalys har den generella faktor som bildas av samtliga enskilda uppgifter i provet separerats ut, varefter det uppgiftspar som enligt modifieringsindexen har störst gemensam residualvarians fått bilda en residualfaktor (biasfaktor). I detta steg har enbart sådana uppgifter beaktats vilka identifierats i det första steget.
- När residualfaktorn identifierats har ytterligare uppgifter lagts till faktorn. Det gäller här sådana uppgifter som dels identifierats i steg 1, dels uppvisar positiv kovarians med residualfaktorn. En förutsättning för att uppgiften skall föras till biasfaktorn är att den ger ett väsentligt bidrag till biasfaktorn varians.
- Efter det att denna process är avslutad har biasfaktorns bidrag till provets totala varians beräknats och faktorn har signifikansprövats.

Eftersom bias mig veterligt inte studerats på detta sätt tidigare – i varje fall inte i Sverige – har samtliga steg genomförts oavsett hur väl en modell med enbart en generell faktor är anpassad till data. Det har emellertid visat sig att om en sådan enfaktormodell visar mycket god anpassning blir signifikansprövningen av biasfaktorerna osäker. Så kan t.ex. en biasfaktor som förklarar mindre än en procent av provets totalvarians ge upphov till t-värden som närmar sig det kritiska värdet för 5 procents-nivån. Av detta skäl redovisas signifikansprövningen av biasfaktorerna samt deras bidrag till provets totalvarians enbart i de fall då RMSEA är 0,025 eller högre.

Tabell 22. Enfaktormodellens anpassning (RMSEA), biasfaktorernas bidrag till totalvariansen samt signifikansprövning av biasfaktorernas bidrag (t-värde).

Prov	Grupp	RMSEA	Bidrag till totalvarians (%)	t-värde
NOG-77	SU	0,030	1,5	2,07
	IS	0,023		
	IU	0,026	1,4	1,72
DTK-77	SU	0,026	0,9	1,25
	IS	0,028	1,7	4,50
	IU	0,026	1,1	3,71
ORD-77	SU	0,014		
	IS	0,028	0,8	1,06
	IU	0,023		
NOG-82	SU	0,031	1,6	0,27
	IS	0,036	3,0	1,16
	IU	0,018		
DTK-82	SU	0,025	1,2	0,81
	IS	0,000		
	IU	0,009		
ORD-82	SU	0,015		
	IS	0,013		
	IU	0,017		

Ett RMSEA-värde på 0,05 eller lägre brukar tolkas som en god modell Anpassning (Steiger, 1990). Som framgår av tabell 22 är samtliga värden avse-

värt lägre, vilket visar att enfaktormodellen är fullt tillräcklig för att förklara de inbördes sambanden mellan provuppgifterna. Detta innebär i sin tur att det inte finns något utrymme för biasfaktorer av betydelse. Detta är fullt i överensstämmelse med att de biasfaktorer som ändå skapats visat sig förklara en mycket liten andel av provens totalvarians. Den högsta andelen uppvisar grupp IS på NOG-provet 1982 med 3 procent. Samtidigt visar emellertid t-värdet att biasfaktorn inte ger ett signifikant bidrag till totalvariansen.

Dessa resultat föranleder således slutsatsen att det inte går att spåra några grupper av uppgifter som tillsammans gynnar grupp SS och som dessutom mäter någon specifik förmåga av betydelse för provresultatet vid sidan av den som mäts av den generella delprovsfaktorn.

Finns det textfaktorer i ELF och LÄS som gynnar grupp SS?

Som tidigare nämnts skiljer sig ELF och LÄS från de övriga delproven såtillvida att de innehåller ett antal längre texter med ett antal frågor knutna till varje text. För LÄS-provets del gäller att varje version av provet består av fem texter med fyra frågor knutna till var och en av dem. ELF har tre längre texter med mellan tre och sex frågor knutna till varje text. Därtill kommer ett antal korta texter med normalt endast en fråga.

Anknytningen av flera frågor ger upphov till lokala beroenden vilket medför en höjning av kovarianserna mellan de frågor som refererar till en och samma text (Wainer & Thissen, 1996). Om dessa texter har ett innehåll som gynnar en viss grupp kommer detta att få ett större genomslag i totalresultatet därför att de tenderar att samvariera. Det är således av intresse att först studera om texternas innehåll kan ge upphov till bias i de totala provresultaten. Syftet med proven är ju inte att mäta provtagarnas förmåga att förstå innehållet i en specifik text, så om vi finner tydliga gruppskillnader vad gäller textfaktorerna rör det sig om bias och inte om valida gruppskillnader.

För att undersöka eventuell bias i textfaktorerna har först en generell delprovsfaktor identifierats. Härigenom isoleras de gruppskillnader som orsakas av den generella förmågan att läsa en svensk eller engelsk text. Därefter har textfaktorer skapats utifrån de återstående kovarianserna mellan de uppgifter som är knutna till respektive text. Dessa residualfaktorer avser då innehållet i texten och är opåverkade av den generella delprovsfaktorn.

En förutsättning att residualfaktorerna skall få något genomslag i totalresultaten är dock att de inte är alltför svaga. Endast sådana residualfaktorer som har en signifikant varians har därför studerats. De övriga har betraktats som så svaga att de inte får någon betydelse för gruppskillnaderna i de totala provresultaten.

Första steget i analysen är således att bestämma textfaktorernas bidrag till delprovets varians. Resultatet av denna analys framgår av tabellerna 23–26.

Faktorerna har fått en beteckning som anknyter till respektive texts rubrik, därav de något kryptiska faktorbeteckningarna.

Tabell 23. Den generella faktorns och textfaktorernas bidrag till delprovets totala varians inom respektive grupp. LÄS-77. (I procent.)

Faktor	SS	SU	IS	IU
Läs	65,6	65,3	64,8	63,9
Betyg	–	–	2,1	0,0
Liv	0,1	1,0	0,1	0,3
Celiak	0,1	0,5	–	0,3
Metaf	0,4	–	1,0	0,4
Eutro	1,6	1,1	1,2	2,2
Mätfel	32,4	32,9	32,2	32,8
Totalt	100,0	100,0	100,0	100,0

Anm. "–" anger att variansen är negativ.

Tabell 24. Den generella faktorns och textfaktorernas bidrag till delprovets totala varians inom respektive grupp. LÄS-82. (I procent.)

Faktor	SS	SU	IS	IU
Läs	63,4	64,3	52,0	66,9
Brev	0,3	0,0	0,9	0,4
Missbruk	1,5	0,4	5,6	0,6
Tid	0,1	–	–	0,9
Kluven	0,4	0,5	–	0,8
Bli	2,3	2,7	0,7	–
Mätfel	32,0	32,6	40,9	31,6
Totalt	100,0	100,0	100,0	100,0

Den generella faktorn ger genomgående signifikanta bidrag till den totala delprovsvariansen, vilket är naturligt med tanke på att den innefattar samtliga 20 provuppgifter. Av residualfaktorerna är det flera som också ger signifikanta bidrag i grupp SS. Ser man till deras procentuella bidrag är dock dessa genomgående små. Att de ändå blir statistiskt signifikanta hänger naturligtvis samman med att grupp SS innefattar så många individer. I de övriga grupperna, som består av mycket färre individer, återfinns endast ett fåtal signifikanta textfaktorer och det är enbart en som bidrar med mer än 5 procent av den totala variansen, nämligen textfaktorn "Missbruk" i LÄS-82 för grupp IS.

För att närmare granska medelvärdeskillnaden mellan grupperna SS och IS på denna faktor har en tvågruppsanalys genomförts varvid alla parame-

trar utom faktormedelvärdena satts lika för de två grupperna. När det gäller "Missbruk" hade grupp IS ett något högre medelvärde än grupp SS, men skillnaden var inte signifikant ($t = 0,76$). Eftersom denna textfaktor ger det klart största bidraget till totalvariansen, men trots detta ett icke-signifikant bidrag, finns det knappast någon anledning att närmare granska övriga textfaktorer med avseende på bias.

Den slutsats som dessa analyser leder till är således att det inte finns några tecken på att texterna i LÄS-77 och LÄS-82 haft ett innehåll som på något betydelsefullt sätt gynnat grupp SS framför de övriga grupperna.

ELF-proven har analyserats på samma sätt som LÄS-proven och variansuppdelningarna för ELF gav följande resultat.

Tabell 25. Den generella faktorns (Elf) och textfaktorernas bidrag till delprovets totala varians inom respektive grupp. ELF-77. (I procent.)

Faktor	SS	SU	IS	IU
Elf	74,0	75,6	74,6	76,6
Woman	1,1	0,8	0,4	0,4
Sun	–	–	–	0,2
Lean	0,1	0,1	0,2	0,8
Mätfel	24,9	23,9	25,0	22,0
Totalt	100,0	100,0	100,0	100,0

Tabell 26. Den generella faktorns (Elf) och textfaktorernas bidrag till delprovets totala varians inom respektive grupp. ELF-82. (I procent.)

Faktor	SS	SU	IS	IU
Elf	77,2	75,6	61,0	76,1
Woman	0,6	0,8	1,4	0,2
Sun	0,6	0,7	–	0,1
Lean	0,6	1,2	4,2	0,7
Mätfel	21,0	21,7	33,3	22,9
Totalt	100,0	100,0	100,0	100,0

Liksom för LÄS förklarar den generella faktorn en mycket stor del av provets totala varians. När det gäller ELF är denna andel i de flesta grupper omkring 75 procent av den totala variansen, vilket är en större andel än vad vi fann för

LÄS. Denna skillnad mellan de två proven har utretts i ett annat sammanhang av Reuterberg (2002).

Ser vi till textfaktorerna är deras andel av den totala variansen låg även på ELF-provet, vilket innebär att eventuella gruppskillnader på dessa faktorer får ringa genomslagskraft på provets totala varians. Det finns därför ingen anledning att ytterligare fördjupa analyserna av detta, så **den slutsats man måste dra på grundval av resultaten i detta avsnitt är att det inte finns några tecken på att gruppskillnaderna i provresultat på något markant sätt påverkas av det innehåll, som texterna i LÄS och ELF har.**

Det faktum att det inte varit möjligt att identifiera några biasfaktorer innebär inte att delproven är helt fria från uppgifter som missgynnar vissa provtagare. Som visats i tabellerna 23–26 svarar mätfelen för en tämligen ansenlig del av delprovets totalvariens. I begreppet ”mätfel” ingår två komponenter, nämligen rena slumpfel orsakade av t.ex. gissning samt specifika inslag i uppgiften. Med specifika inslag menas här sådana inslag som är mer eller mindre unika för uppgiften. Dessa specifika inslag kan naturligtvis vara av sådan art att de gynnar vissa provtagare framför andra. Om det finns sådana inslag är de antingen svaga i den meningen att de förklarar en liten del av uppgiftens totala varians eller, också är de av sådan art att de är begränsade till en enstaka eller några få uppgifter. Om så inte vore fallet skulle de ha upptäckts i den konfirmatoriska faktoranalysen.

Den generella högskoleprovsfaktorns roll för gruppkillnaderna

Att högskoleprovet mäter en generell faktor som slår igenom på samtliga delprov har tidigare visats av Gustafsson et al. (1992). De data som låg till grund för denna undersökning var emellertid hämtade från ett högskoleprov och en provtagargrupp med en väsentligt annorlunda sammansättning än vad som gäller för de data som ingår i föreliggande undersökning. Det kan därför vara befogat att först studera de interkorrelationer som föreligger mellan de olika delprovsfaktorerna.

Tabell 27. Interkorrelationer mellan de generella delprovsfaktorerna bland provtagarna.

	Födda 1977				Födda 1982			
	NOG	LÄS	DTK	ELF	NOG	LÄS	DTK	ELF
ORD	.525	.813	.546	.834	.511	.795	.550	.848
NOG		.611	.864	.578		.586	.868	.590
LÄS			.621	.814			.610	.829
DTK				.606				.634

Tabell 28. Interkorrelationer mellan de generella delprovsfaktorerna i populationerna.

	Födda 1977				Födda 1982			
	NOG	LÄS	DTK	ELF	NOG	LÄS	DTK	ELF
ORD	.652	.865	.667	.877	.621	.841	.651	.880
NOG		.725	.902	.692		.680	.898	.679
LÄS			.733	.866			.700	.866
DTK				.712				.714

Samtliga delprov korrelerar relativt högt positivt vilket visar att det finns en tämligen stark gemensam faktor. I synnerhet gäller detta då korrelationerna studeras på populationsnivå där den "restriction of range" som följer av selektionen till provet inte inverkar. Även om de inbördes sambanden är höga lämnar de ett visst utrymme för inslag av specifika faktorer i vart och ett av delproven. Detta innebär att man kan dela upp variansen i vart och ett av del-

proven på en del som förklaras av den generella högskoleprovsfaktorn och en del som är att hänföra till det specifika inslaget i varje delprov.

En intressant fråga blir då hur den generella högskoleprovsfaktorn påverkar de gruppskillnader som tidigare påvisats i vart och ett av delproven. Innan vi går till denna analys kan det vara av intresse att visa hur stor del av varianserna i delproven som påverkas av den generella högskoleprovsfaktorn.

I vilken grad den generella högskoleprovsfaktorn (HPR) påverkar delprovresultaten framgår av tabell 29. Som värdena i tabellen visar är det HPR som svarar för huvuddelen av variansen i de språkliga delproven, medan HPR och delprovsfaktorerna har ungefär lika stark påverkan på resultaten på de två kvantitativa delproven NOG och DTK.

Tabell 29. Andel delprovsvarians som förklaras av den generella högskoleprovsfaktorn. (I procent.)

	Födda 77		Födda 82	
	Provtagare	Population	Provtagare	Population
ORD	73,7	79,0	75,3	80,3
NOG	50,4	60,5	47,8	58,7
LÄS	81,5	87,0	79,0	83,7
DTK	55,1	69,1	52,7	63,4
ELF	80,8	83,9	86,7	87,4

Som väntat utifrån de inbördes korrelationerna mellan delprovsfaktorerna i tabellerna 27 och 28 utövar HPR en starkare påverkan på delproven då denna påverkan studeras på populationsnivå jämfört med på provtagarnivå. Selektionseffekterna till högskoleprovet medför alltså att vi något underskattar den betydelse som HPR har på gruppskillnaderna i delprovresultat om analyserna baseras på enbart dem som genomgått högskoleprovet.

Av tabell 29 framgår vidare att HPR har en starkare inverkan på de språkliga delprovsfaktorerna, ORD, LÄS och ELF, än på de kvantitativt inriktade faktorerna. Om man ser till populationernas resultat svarar HPR för 80 procent av den totala variansen eller däröver, medan motsvarande andel på de kvantitativa proven är mellan 60 och 70 procent.

Medelvärdesskillnaderna i den generella HPR-faktorn redovisas i tabell 30. De värden som gäller för provtagare är bruttoskillnaderna, medan de för populationerna är de skillnader som återstår då hänsyn tagits till såväl selektionseffekter som skillnader i betyg.

Tabell 30. Gruppsskillnader i den generella högskoleprovsfaktorn (HPR).

Grupp	Födda 1977		Födda 1982	
	Provtagare	Population	Provtagare	Population
SU	-.282	-.232	-.327	-.229
IS	-.347	-.181	-.445	-.324
IU	-.362	-.253	-.433	-.359

Som tabellen visar är samtliga differenser negativa vilket innebär att grupp SS ligger högre än samtliga övriga grupper på den generella högskoleprovsfaktorn. Ser vi till bruttoskillnaderna dvs. de skillnader som föreligger oavsett betygsskillnader och oavsett differentiell selektion till högskoleprovet är det grupp SU som ligger närmast SS. Lägst ligger således grupperna IS och IU och mellan dem är bruttoskillnaderna små. När det gäller dem födda 1977 medför kontrollen av betyg och selektionseffekter en tydlig förändring för grupp IS. Då är det nämligen denna grupp som kommer närmast grupp SS och detta är ett fullt logiskt resultat mot bakgrund av att IS har ett klart lägre betygsmedelvärde än grupp SU såväl bland provtagare som i population bland dem födda 1977 (jfr tabell 9).

Även bland 82:orna kommer grupp SU närmast SS och detta gäller oavsett om vi kontrollerat för selektionseffekter och betyg eller inte. Den inbördes skillnaden mellan grupperna IS och IU är däremot små i båda fallen.

Den generella slutsats man kan dra på grundval av resultaten i tabell 30 är således att de gruppsskillnader som framträder i de observerade resultaten på högskoleprovet till väsentlig del har sin grund i skillnader i den generella faktor som hela provet mäter. Om man beaktar den starka påverkan som HPR utövar på delprovresultaten innebär detta bland annat att det torde vara ytterst svårt att på något avgörande sätt ändra på gruppsskillnaderna i provresultat genom att göra specifika ingrepp i delproven och i ännu högre grad genom ingrepp i enstaka provuppgifter.

De skillnader som påvisats i HPR svarar naturligtvis för en stor del av skillnaderna i delprovresultat, men vid sidan av HPR påverkas naturligtvis också dessa skillnader av de gruppsskillnader som föreligger i delprovsfaktorerna. I tabellerna 31 och 32 redovisas dessa skillnader i delprovsfaktorerna. Det är här viktigt att observera att dessa skillnader har beräknats så att de är opåverkade av de skillnader som finns i HPR. Man kan säga att resultaten i tabellerna 31 och 32 anger det specifika bidraget av delprovsfaktorerna till gruppsskillnaderna i delprovresultat.

Tabell 31. Delprovsfaktorernas specifika bidrag till gruppskillnader i delprovresultat. Födda 1977.

	Provtagare					Population				
	ORD	NOG	LÄS	DTK	ELF	ORD	NOG	LÄS	DTK	ELF
SU	-.365	-.325	-.185	-.328	+0.087	-.383	-.315	-.208	-.306	+0.089
IS	-.146	-.542	-.242	-.576	-.325	+0.002	-.382	-.082	-.398	-.193
IU	-.506	-.340	-.272	-.434	+0.009	-.465	-.264	-.215	-.343	+0.071

Tabell 32. Delprovsfaktorernas specifika bidrag till gruppskillnader i delprovresultat. Födda 1982.

	Provtagare					Population				
	ORD	NOG	LÄS	DTK	ELF	ORD	NOG	LÄS	DTK	ELF
SU	-.350	-.460	-.466	-.500	-.182	-.238	-.346	-.272	-.368	-.121
IS	-.261	-.532	-.347	-.577	-.245	-.189	-.440	-.232	-.467	-.234
IU	-.448	-.571	-.551	-.611	-.185	-.369	-.491	-.400	-.509	-.167

En jämförelse mellan resultaten i tabellerna 31 och 32 med dem i tabellerna 17 och 18 visar på mycket stora likheter. Detta innebär att kontrollen av HPR inte på något avgörande sätt påverkar gruppskillnaderna i de generella delprovsfaktorerna.

Slutsatsen blir därför att det finns gruppskillnader också i HPR, men att dessa skillnader inte förändrar de skillnader som tidigare påvisats på delprovsnivå. Detta innebär att orsaken till gruppskillnaderna i högskoleprovresultat inte är att finna i specifika egenskaper hos enskilda uppgifter, utan de ligger på de mer generella förmågor som mäts av delproven och den förmåga som hela provet mäter.

Sammanfattande diskussion

Föreliggande undersökning utgör en uppföljning av en tidigare studie, genomförd av Reuterberg och Hansen (2001), vilken publicerats i Högskoleverkets rapportserie med titeln *Vilken betydelse har utländsk bakgrund för resultatet på högskoleprovet?* I den undersökningen jämfördes fyra grupper med olika invandrarstatus med avseende på bland annat resultat på högskoleprovets olika delprov. De grupper som jämfördes var:

Grupp SS	personer födda i Sverige med minst en förälder född i Sverige
Grupp SU	personer födda i Sverige med båda föräldrarna födda utomlands
Grupp IS	personer födda utomlands med båda föräldrarna födda i Sverige
Grupp IU	personer födda utomlands med minst en förälder född utomlands

Bland de resultat som närmast berör föreliggande undersökning kan nämnas att grupp SS överträffar samtliga andra grupper både vad gäller den högsta uppnådda normerade totalpoängen på provet och på de resultat som gäller de enskilda delproven. Av de enskilda delproven var det DTK som uppvisade störst skillnader och ELF minst. Det fanns inga tecken på att skillnaderna förändrats på ett systematiskt sätt under den tid som studerades (1991–1996). Ålder för invandring till Sverige hade stor betydelse för resultaten. De som kommit till Sverige före 10 års ålder uppnådde klart högre resultat jämfört med dem som kommit senare i livet.

När man ser resultat från Reuterbergs och Hansens undersökning ligger det nära till hands att tolka dem som uttryck för bias. Med bias menas att en grupp på ett orättmätigt sätt gynnas av provet framför andra grupper. Så skedde t.ex. på 90-talet då könsskillnaderna på provet stod i fokus. I den allmänna debatten framfördes ibland att högskoleprovet var ”ett prov som gjordes av män för män”.

Ett stort misstag som många gjorde vid denna tid var emellertid att man bortsåg från det faktum att högskoleprovet är ett frivilligt prov och att de som kom till provet utgjorde ett positivt gallrat urval ur samtliga i sin åldersklass. Det är emellertid inte enbart så att provtagargruppen är positivt gallrad. Den är dessutom olika strängt gallrad från olika grupper i populationen. Bland annat har Mäkitalo och Reuterberg (1996) och Reuterberg (1998) visat att graden av selektion till provet skiljer sig mellan män och kvinnor och mellan provtagare från olika sociala grupper. När det gäller föreliggande undersökning har resultaten visat att det föreligger också en betydande differentiell

selektion till provet mellan grupper med olika invandrarstatus. Den starkaste selektionen sker från grupp IU och den näst starkaste från SU. Grupperna SS och IS är följaktligen de grupper varifrån den minst stränga selektionen sker. Dessa selektionsmekanismer gör att det inte finns någon anledning till att förvänta sig lika provresultat för provtagare med olika invandrarstatus.

Den tekniska termen för orättmätigt gynnande av en viss grupp framför andra är bias. Enligt definitionen innebär bias att två grupper som är lika bra i den förmåga eller de förmågor som provet avser mäta uppnår olika resultat på provet. Bias skall således vara orsakat av för provet irrelevanta faktorer.

En central fråga när det gäller bias är om denna skall studeras på uppgiftsnivå ("item bias") eller på provnivå ("test bias"). De biasundersökningar som hittills genomförts har vanligen handlat om bias på uppgiftsnivå, vilket innebär att man jämfört olika gruppers lösningsfrekvenser på enskilda uppgifter och där grupperna jämförts med avseende på den totala poängen på provet.

Detta förfaringsätt har enligt min mening flera påtagliga nackdelar. En sådan är att om det finns en viss bias i det totala provresultatet kommer man att upptäcka enbart sådana uppgifter som har en särskilt hög grad av bias. En annan är att man inte får något svar på i vad mån biasuppgifterna gör att olika grupper får olika totalresultat på provet. Det har också visat sig vara svårt att utifrån enskilda uppgifter sluta sig till orsakerna till bias (Boldt, 1983). Två uppgifter kan ha många yttre kännetecken gemensamma utan att mäta samma bakomliggande förmåga. Likaså kan två till synes helt olika uppgifter vara uttryck för en och samma bakomliggande förmåga.

Roznowski och Reith (1999) förespråkar att bias skall studeras på testnivå eftersom det är det totala testresultatet som ligger till grund för beslut och inte resultat på enskilda uppgifter. Att studera bias på testnivå har också den fördelen att man med modern statistisk metodik lätt kan granska om de uppgifter som visar bias utgör indikatorer på en och samma bakomliggande specifik förmåga, vid sidan av den generella förmåga som hela testet mäter, eller om de mäter helt olika specifika förmågor. Om det visar sig att ett antal biasuppgifter mäter samma specifika förmåga, är det dels lättare att spåra orsaken till bias, dels får detta en avsevärt större effekt på det totala provresultatet jämfört med om uppgifterna är okorrelerade med varandra. Denna effekt ökar nämligen med kvadraten på antalet uppgifter som mäter den specifika förmågan (Reuterberg & Gustafsson, 1992).

Frågan om bias på uppgiftsnivå eller bias på testnivå är inte bara en fråga av teoretiskt intresse. Den får också stora praktiska konsekvenser. Om man studerar bias på uppgiftsnivå leder resultaten normalt till att man finner ett antal uppgifter som uppvisar större gruppskillnader än andra. Den naturliga konsekvensen blir då att man också letar efter orsakerna på uppgiftsnivå och utifrån detta inriktar sig på att modifiera uppgifterna på ett sådant sätt att man tror sig minska gruppskillnaderna. Om man däremot studerar bias på testnivå är den första frågan man ställer sig huruvida gruppskillnaderna har

sitt ursprung i den generella förmåga som hela testet mäter. Om det visar sig att så är fallet, är modifieringar av enskilda uppgifter en otillräcklig åtgärd eftersom orsakerna då inte ligger i specifika inslag i uppgifterna, utan i de bakomliggande faktorer som provet mäter. I stället måste man då ställa sig frågan om dessa faktorer har ett så högt prognostiskt värde att de är outhärliga som prognosvariabler. Skulle så inte vara fallet bör naturligtvis hela testet eller delar av det bytas ut mot något annat test som är likvärdigt vad gäller prognosvärde, men som ger upphov till mindre bias.

Att studera bias på testnivå utesluter emellertid inte att man också studerar eventuell bias i enskilda uppgifter. Framför allt bör man då inrikta arbetet på att granska om det finns grupper av uppgifter som dels visar bias, dels uppvisar ett inbördes samband som inte kan förklaras av den generella faktorn. Om båda dessa villkor är uppfyllda mäter testet förutom den generella faktorn också en specifik faktor, vars relevans för prognosvärdet måste övervägas. Om den specifika faktorn betraktas som viktig för det som testet är avsett att mäta handlar det enligt Coles och Moss (1989) definition inte om bias. Bedöms däremot faktorn som mer eller mindre irrelevant för kriteriet föreligger bias och man bör då överväga att byta ut de uppgifter som utgör indikatorer på den specifika faktorn. Detta bör emellertid inte ske med automatik, utan först bör man göra en totalbedömning och granska huruvida de nackdelar som följer av bias är av större betydelse än de positiva bidrag som faktorn ger till det totala testresultatets mätegenskaper. För att göra en sådan bedömning krävs att man har genomfört en ingående analys av hela provets mätegenskaper. Detta innebär att det mest relevanta sättet att studera bias är att göra det på testnivå, och detta har varit utgångspunkt för föreliggande studie.

Ett problem med att studera bias på testnivå är dock att man måste finna en extern kontrollvariabel med vars hjälp man kan jämföra de grupper som skall jämföras med avseende på den förmåga som testet är avsett att mäta. Idealt skulle naturligtvis framgång i högskolestudierna utgjort den mest lämpliga kontroll- och kriterievariabeln i denna undersökning, men denna variabel lider av den bristen att den inte är jämförbar mellan personer som genomgått olika högskoleutbildningar. Dessutom har vi i undersökningen inte haft tillgång till några säkra framgångsmått för dem födda 1982. Vad som då ligger närmast till hands är att välja gymnasiebetygen som mått på studieförmåga. Tyvärr lider dessa betyg av den nackdelen att inte heller de är jämförbara över olika utbildningar eftersom varje gymnasieutbildning har sin egen referensgrupp vid betygsättningen. Därför har jag fått välja avgångsbetygen från grundskolan som kontrollvariabel. Den stora fördelen med grundskolebetygen är att de finns tillgängliga för nästan samtliga personer som ingår i undersökningen och de är dessutom jämförbara mellan olika personer, eftersom de insamlats vid en tidpunkt då alla har samma utbildningsbakgrund.

En förutsättning för att nå jämförbara resultat mellan de olika grupperna är dock att kontrollvariabeln (grundskolebetygen) har samma mätegenskaper i de olika grupperna. Vid en kontroll av detta visade det sig att betyget

i svenska hade en klart lägre vikt i den generella betygsfaktorn för personer tillhörande grupperna SU och IU jämfört med dem tillhörande grupperna SS och IS, varför detta betyg fick uteslutas. Därmed kom kontrollvariabeln att bestå av den generella betygsfaktor som definieras av betygen i engelska, matematik, biologi, fysik, kemi, teknik, geografi, historia, religion och samhällskunskap.

Ett annat stort problem oavsett om man studerar bias på uppgiftsnivå eller bias på testnivå är den differentiella selektionen till högskoleprovet. Som nämnts ovan är grupperna IU och SU strängare selekterade till provet än grupperna SS och IS. Som diskuterats av Reuterberg (1996) leder denna differentiella selektion till att de gruppkillnader man finner i provresultat bland provtagarna kan skilja sig väsentligt från dem man skulle ha funnit i hela i populationen. Med andra ord kan provtagarnas resultat ge en snedvriden bild av hur provet faktiskt fungerar för de olika grupperna. Ett sätt att lösa detta problem är att skatta hur resultaten skulle se sig om samtliga i populationen hade genomgått högskoleprovet. En sådan skattning kan numera göras med hjälp av en statistisk metodik som kallas "modellering av inkompleta data" eller "missing data analysis" (Allison, 1987; Muthén, Kaplan & Hollis, 1987).

De resultat som redovisats tidigare i rapporten avser dels de som gäller för provtagarna, dels de skattade resultat som avser populationen. I den fortsatta diskussionen kommer jag att begränsa mig till de resultat som gäller för populationen, då betygen kontrolleras. Detta är nämligen de resultat som ger den mest relevanta bilden av bias i högskoleprovets delprov.

De data som ligger till grund för analyserna är hämtade från VALUTA-projektets stora longitudinella databas. Två årskullar har valts ut, nämligen de som är födda 1977 respektive 1982. Den äldre gruppen omfattar 93 200 personer varav 19 000 har genomgått högskoleprovet under våren 1996, det år då personerna fyllde 19 år. I den yngre gruppen ingår sammanlagt 89 600 personer och av dem har 10 800 genomgått högskoleprovet våren 2001.

De två grupperna skiljer sig åt i flera väsentliga avseenden. Den äldre gruppen har fått grundskolebetyg enligt den gamla femgradiga betygsskalan, medan den yngre fått de nya kriterierelaterade betygen vilka uttrycks i fyra skalsteg. En annan väsentlig skillnad är att den yngre gruppen haft möjlighet att i efterhand komplettera sina gymnasiebetyg, medan så inte är fallet för den äldre. Det sistnämnda torde vara en bidragande orsak till att andelen provtagare är betydligt lägre i den yngre gruppen.

Vi har således inte bara en differentiell selektion mellan personer med olika invandrabakgrund, utan vi måste dessutom räkna med en differentiell selektion mellan de två årskullarna. Båda dessa förhållanden utgör starka skäl för att genomföra analyserna med hjälp av modellering av inkompleta data och skatta de skillnader som föreligger på populationsnivå. Vidare har skillnaderna studerats med kontroll av betygsskillnaderna. Det senare motiveras av att bias enligt definitionen avser skillnader mellan grupper som är jämställda vad avser den förmåga som testet är avsett att mäta.

Som nämnts ovan är den första fråga man ställer sig när det gäller bias på testnivå om det finns några skillnader mellan de olika grupperna i den generella förmåga som testet mäter. Överfört till denna undersökning innebär detta att vi först granskar om det finns några skillnader mellan personer med olika invandrarstatus i de generella delprovsfaktorerna. Skillnaderna har beräknats så att medelvärdet i delprovets generella faktor för var och en av grupperna har jämförts med det medelvärde som gäller för grupp SS. Dessa skillnader har tidigare redovisats i tabell 18, men för att underlätta läsningen upprepas de här.

Tabell 33. Standardiserade medelvärdesskillnader mellan grupp SS och övriga grupper i de generella delprovsfaktorerna inom populationerna med betyg under kontroll.

Grupp	Födda 1977					Födda 1982				
	ORD	NOG	LÄS	DTK	ELF	ORD	NOG	LÄS	DTK	ELF
SU	-.380	-.324	-.217	-.326	+.083	-.220	-.324	-.270	-.371	-.098
IS	+.009	-.386	-.082	-.418	-.188	-.223	-.452	-.261	-.505	-.264
IU	-.463	-.278	-.230	-.371	+.061	-.371	-.476	-.402	-.526	-.161

Bland dem födda 1982 är samtliga skillnader negativa vilket innebär att grupp SS har ett högre medelvärde i samtliga delprovsfaktorer jämfört med övriga grupper². Även bland 77:orna går de flesta jämförelser till fördel för grupp SS, men i tre fall går skillnaderna i motsatt riktning. Ett av dem gäller ORD, där grupp IS har ett i princip lika högt medelvärde som grupp SS. Vidare har grupperna SU och IU nått ett något högre medelvärde än grupp SS på ELF (engelsk läsförståelse).

Det är knappast någon tillfällighet att ELF är den faktor som svarar för två av de tre positiva differenserna i tabell 31. ELF-provet är nämligen det prov som uppvisade minst skillnader i totalpoäng bland provtagarna i den undersökning som genomfördes av Reuterberg och Hansen (2001) och detta gäller också i föreliggande undersökning (jfr tabell 16). En trolig förklaring till detta är att många personer i grupperna SU och IU haft minst lika goda möjligheter att lära sig engelska som de tillhörande grupp SS, samtidigt som ELF är det enda prov som inte refererar till svenska förhållanden. De största gruppskillnaderna uppvisar däremot NOG- och DTK-faktorerna, vilket också rimmar väl med Reuterbergs och Hansens resultat.

2. 1 provpoäng motsvarar t.ex. en medelvärdesskillnad på PTK-provet på -0,4 av cirka 2 poäng. Detta kan verka vara en liten skillnad, men summerar man skillnaderna för alla (fem) delproven blir resultatet 10 poäng, vilket motsvarar en sänkning av högskoleprovsresultatet från 1,6 till 1,4.

Grupp SU är den grupp som på det hela taget ligger närmast SS, medan grupp IU får de lägsta resultaten. Att grupp SU får ett högre värde totalt sett än grupp IU är helt naturligt med tanke på att de som tillhör den förstnämnda gruppen normalt vistats i Sverige hela sitt liv, medan de tillhörande grupp IU flyttat in till Sverige. För att kontrollera vilken betydelse som ålder för inflyttning har delades grupp IU i två undergrupper: de som invandrat till Sverige före 10 års ålder och de som invandrat senare i livet. Denna kontroll visade att de låga resultaten för grupp IU delvis kan förklaras av att en relativt stor andel av dem kommit till Sverige senare i livet. De som kommit före 10 års ålder presterar nämligen resultat som ligger fullt i nivå med grupp SU medan de som kommit senare presterar lågt på nästan samtliga delprov.

Grupp IS torde vara den grupp vars uppväxtförhållanden mest liknar dem för grupp SS. Trots detta presterar de klart sämre resultat än grupp SS. Bland 77:orna är deras genomsnittliga prestationer i nivå med dem tillhörande grupp SU, men bland 82:orna ligger de klart lägre. Förklaringen till detta är att de uppvisar en mycket markerad profil över de olika delproven med förhållandevis goda prestationer på ORD och LÄS, men svaga prestationer på de kvantitativa proven NOG och DTK. Med tanke på att kvinnorna utgör en klar majoritet i gruppen och att kvinnor presterar klart lägre än män på de två kvantitativa proven föreföll den sneda könsfördelningen vara den mest troliga orsaken till de jämförelsevis låga resultaten på de kvantitativa delproven. En kontroll visade emellertid att den tydliga profilen inte är en effekt enbart av den sneda könsfördelningen. Även andra faktorer måste vara bidragande, men vilka dessa är kan inte närmare utredas med de data som finns tillgängliga.

Förutom IS uppvisar också IU en tydlig profil över delproven med låga resultat på främst ORD och LÄS. Till stor del kan detta förklaras av att gruppen innefattar ett stort antal personer som kommit till Sverige i 10-årsåldern eller senare. De har naturligtvis haft klart sämre möjligheter än andra att lära sig det svenska språket, vilket får det starkaste genomslaget i ORD och LÄS. Språksvårigheterna, tillsammans med de mer begränsade erfarenheterna av svenska företeelser, får naturligtvis också till följd att de fått sämre förutsättningar att prestera väl även på NOG och DTK. På ELF, där svenska språket och kännedom om svenska förhållanden inte är en förutsättning på samma sätt, presterar de däremot ett resultat som överträffar grupp IS.

Om man ser till den totala bild som skillnaderna i delprovfaktorerna uppvisar, är det uppenbart att dessa svarar för en stor del av gruppskillnaderna i provresultat. Detta hindrar emellertid inte att det också kan finnas specifika biasfaktorer i delproven vilka verkar förstärkande på bias. För att undersöka detta valdes ur ORD, NOG och DTK ut de uppgifter som uppvisade störst gruppskillnader i lösningsfrekvens. Därefter undersöktes med konfirmatorisk faktoranalys huruvida det var möjligt att identifiera några sådana mer specifika biasfaktorer som gav ett väsentligt tillskott till den totala delprovsvariansen. Några sådana faktorer kunde emellertid inte spåras, varför slutsatsen blir

att dessa delprov inte innehåller några specifika biasfaktorer av betydelse för delprovskillnaderna i resultat.

LÄS och ELF, som är uppbyggda kring ett begränsat antal texter med kluster av uppgifter kopplade till varje text, analyserades med avseende på eventuell bias i dessa textfaktorer. Inte heller dessa analyser visade att texterna hade ett innehåll som på ett orättmätigt sätt gynnade vissa av grupperna.

Den första slutsatsen man kan dra utifrån denna undersöknings resultat är att *orsaken till gruppskillnaderna i totalpoäng får sökas på delprovnivå eller i hela högskoleprovets utformning*. Det handlar således om skillnader i de mer generella faktorerna.

Gruppskillnaderna kan givetvis påverkas också av mer specifika delprovfaktorer såsom t.ex. texternas innehåll i ELF och LÄS, vissa ordkategorier i ORD som låneord eller äldre svenska ord och uttryck, figurernas och tabellernas utformning i DTK eller den matematiska strategi som leder fram till lösningen av uppgifterna i NOG. Det finns emellertid *inga tecken på att denna typ av specifika delprovfaktorer spelar någon större roll för gruppskillnaderna i totalresultat*.

Däremot kan det finnas bias i enskilda uppgifter. Huruvida detta är fallet har inte varit möjligt att kontrollera med den statistiska metodik som tillämpats här. Bias i enskilda uppgifter har emellertid liten betydelse för totalresultatet såvida uppgifterna inte är klart positivt korrelerade inbördes. Dvs. att de utgör indikatorer på en specifik delprovfaktor, vilken i så fall skulle visa sig i den konfirmatoriska faktoranalysen. Med tanke på att inga specifika biasfaktorer kunnat spåras, och att de modeller som prövats uppvisar mycket god överensstämmelse med de empiriska data finns inget väsentligt utrymme för en sådan inverkan. Slutsatsen blir därför att *eventuell bias i de enskilda uppgifterna spelar en mycket liten roll för skillnaderna i totalresultat*.

Det faktum att gruppskillnaderna huvudsakligen kan hänföras till de generella delprovfaktorerna får viktiga praktiska konsekvenser. Det visar nämligen *att man inte kan åstadkomma några större förändringar av gruppskillnaderna i provresultat genom att modifiera enstaka uppgifter. För att uppnå en väsentlig förändring av gruppskillnaderna i det totala högskoleprovresultatet krävs att man helt eller delvis ersätter något eller några av de befintliga delproven med delprov som mäter andra bakomliggande förmågor. Samtidigt måste man se till att dessa nya prov har en prognostisk validitet som är av samma styrkeordning som de nu befintliga och som dessutom ger lägre gruppskillnader i resultat*.

Man frågar sig då vad dessa mer generella faktorer innefattar. När det gäller skillnaderna i den för hela provet generella faktorn kan de exempelvis hänga samman med hur hela testsituationen är utformad, att samtliga delprov bygger på flervalsuppgifter, att huvudparten av proven ges på svenska och att många uppgifter refererar till svenska förhållanden. Vad som talar för att detta är viktiga orsaker till skillnaderna är, att gruppskillnaderna är klart minst på ELF-provet – ett prov som ges helt på engelska och vars frågor normalt inte avser svenska förhållanden.

Undersökningen har vidare visat att grupp IU får låga resultat särskilt på de språkliga proven, ORD och LÄS, samt att det inom denna grupp finns ett klart samband mellan den tid man vistats i Sverige och provresultat. De som kommit till Sverige före 10 års ålder klarar nämligen provet klart bättre än de som invandrat senare i livet. Kunskaper och färdigheter i svenska förefaller således vara en väsentligt bidragande orsak till skillnaderna i provresultat. Likaså är erfarenheter av typiskt svenska förhållanden en ytterligare bidragande faktor till skillnaderna på de två kvantitativa proven NOG och DTK.

En relevant fråga i detta sammanhang är om högskoleprovet nödvändigtvis bör ha ett så tungt språkligt inslag. Av de fem delproven är ju tre – ORD, LÄS och ELF – rent verbala prov som dessutom visat sig vara starkt korrelerade inbördes (Reuterberg, 2002). Med tanke på det stora antalet uppgifter i ORD förstärks den språkliga faktorns betydelse för totalresultatet mer än vad antalet prov ger anledning att förvänta. Om man anser att provet har en alltför stark verbal betoning handlar det om bias och då skulle man överväga om något av dessa prov kan bytas ut mot ett prov som är mindre verbalt laddat, om en reduktion av antalet uppgifter i ORD får några negativa effekter på hela provets mätegenskaper, eller om det skulle vara möjligt att införa en annan typ av uppgifter som är mindre verbalt betonade. Som exempel på sådana provtyper kan nämnas figurprov eller symbolprov. Är det däremot så att den starka verbala laddningen bedöms var viktig för provets prognostiska validitet skall några sådana åtgärder inte vidtas. De skillnader som framträder i provresultaten betingas ju då av en förmåga som är viktig för möjligheterna att genomföra högskolestudierna på ett framgångsrikt sätt, och skillnaderna är då att betrakta som valida. Dock är det en bedömningsfråga vad man totalt sett vinner eller förlorar på en minskning av de verbala inslagen i provet.

Nu är det ju inte enbart de verbala proven som ger gruppskillnader till fördel för dem som är födda i Sverige och har minst en svenskfödd förälder, utan även NOG och DTK uppvisar stora skillnader i samma riktning. Detta är i och för sig inte förvånande med tanke på att såväl den givna informationen i uppgifterna som frågorna ges på svenska. När det gäller DTK refererar uppgifterna också ofta till typiskt svenska förhållanden.

Om man ser till de korrelationer som delprovsfaktorerna uppvisar med betyg finns det inget delprov som avviker från de övriga vad gäller att prognostisera studieförmåga, i varje fall inte så som denna förmåga definierats här. Bland 77:orna ligger de genomsnittliga sambanden för samtliga delprov strax under 0,70 och bland 82:orna något över 0,65. Korrelationen kan anta värden mellan 0 och 1, där 0 betyder inget samband och 1 betyder perfekt samband. Man får således gå på andra överväganden än prognosvärdet om något av delproven skall bytas ut.

Frågan om att ersätta något eller några av de nu befintliga delproven med prov av annan art är en fråga som för övrigt nyligen aktualiserats i den utvärdering av högskoleprovet som Högskoleverket lät en internationell expertgrupp genomföra. Expertgruppens rekommendationer finns presenterade i

rapporten ”The Swedish national aptitude test: a 25-year testing program. Current status and future development”. (Högskoleverket, 2002)

Som korrelationerna visat har det gamla femgradiga betygssystemet ett något bättre prognosvärde än det nya fyrgradiga. Denna skillnad bör följas upp och närmare analyseras eftersom den kan få stor betydelse för betygens respektive provresultatens roll i det framtida urvalssystemet till högre utbildning.

De slutsatser som dragits här vad gäller eventuell bias i högskoleprovet bygger på antagandet att grundskolebetygen utgör en god indikator på personernas förutsättningar att bedriva framgångsrika studier även på högskolenivå. Sambandet mellan medelbetyget från gymnasieskolan och medelbetyget från grundskolan uppgår till 0,76. Detta tyder på att grundskolebetygen utgör en god ersättning för gymnasiebetygen. Dock kan man inte bortse ifrån det faktum att det kan finnas bias i grundskolebetygen. Om detta är fallet finns det utrymme för bias också i högskoleprovet utan att denna bias visar sig vid analyserna. Finns det bias i högskoleprovet som är av samma natur som de i betygen kommer nämligen denna bias i betygen att medföra en underskattning av eventuell bias i provresultaten.

Det enda sättet att komma ifrån detta problem är att finna en mer relevant kontrollvariabel och det handlar då om ett direkt mått på framgång i högskolestudierna. Vilket framgångsmått som skall väljas är emellertid en fråga som inte har ett självklart svar, och som inte skall behandlas här. Några säkra framgångsmått fanns som tidigare nämnts inte tillgängliga då denna undersökning startade, men uppgifterna är under bearbetning inom VALUTA-projektet och kommer att finnas tillgängliga inom en snar framtid. Med dessa uppgifter tillgängliga förfogar vi över ett helt unikt datamaterial för studier av den prognostiska validiteten hos urvalstest.

Det finns således all anledning att replikera den undersökning som här presenterats, men med ett mer relevant kriteriemått – t.ex. poängproduktion – och med utnyttjande av modern statistisk metodik. *Den metod som här använts – bias på testnivå analyserad med hjälp av strukturell ekvationsmodellering – har nämligen visat sig vara lämplig.* Till metodens fördelar hör att:

1. man utgår från hela provets faktorstruktur,
2. man enkelt kan dela upp provets totala varians i olika komponenter vars ursprung kan identifieras,
3. man arbetar med sanna värden, vilket innebär att man undviker påverkan av mätfel,
4. man får veta hur väl analysmodellen stämmer med de empiriska data samt
5. det är möjligt att skatta resultat för hela populationen varigenom man undviker en påverkan av selektionen och differentiell selektion till provet.
6. men analyserar man varje uppgifts mätegenskaper i det sammanhang som uppgiften skall fungera – det vill säga tillsammans med övriga uppgifter i provet – vilket är den allra största fördelen med metoden.

Bilaga I

Tabell 1.1. Standardiserade medelvärdeskillnader mellan respektive grupp och grupp SS i de generella delprovsfaktorerna. Jämförelser mellan manliga och kvinnliga provtagare. Födda 1977.

Grupp	Män					Kvinnor				
	ORD	NOG	LÄS	DTK	ELF	ORD	NOG	LÄS	DTK	ELF
SU	-.438	-.437	-.475	-.571	+.107	-.480	-.597	-.572	-.772	+.001
IS	-.190	-.625	-.285	-.798	-.575	-.134	-.749	-.628	-1.057	-.524
IU	-.669	-.574	-.613	-1.013	+.108	-.564	-.583	-.694	-.808	-.085

Tabell 1.2. Standardiserade medelvärdeskillnader mellan respektive grupp och grupp SS i de generella delprovsfaktorerna. Jämförelser mellan män och kvinnor i populationen med kontroll för betyg och selektionseffekter. Födda 1977.

Grupp	Män					Kvinnor				
	ORD	NOG	LÄS	DTK	ELF	ORD	NOG	LÄS	DTK	ELF
SU	-.411	-.399	-.419	-.511	+.152	-.390	-.469	-.379	-.594	+.154
IS	-.049	-.440	-.021	-.540	-.368	+.076	-.457	-.200	-.672	-.180
IU	-.549	-.415	-.378	-.788	+.071	-.446	-.414	-.448	-.578	+.114

Tabell 1.3. Standardiserade medelvärdeskillnader mellan respektive grupp och grupp SS i de generella delprovsfaktorerna. Jämförelser mellan manliga och kvinnliga provtagare. Födda 1982.

Grupp	Män					Kvinnor				
	ORD	NOG	LÄS	DTK	ELF	ORD	NOG	LÄS	DTK	ELF
SU	-.521	-.540	-.350	-.490	-.335	-.565	-.731	-.562	-.605	-.288
IS	-.603	-.918	-.383	-.846	-.724	-.576	-.718	-.538	-.571	-.458
IU	-.887	-.890	-.652	-.839	-.566	-.649	-.786	-.540	-.589	-.213

Tabell 1.4. Standardiserade medelvärdeskillnader mellan respektive grupp och grupp SS i de generella delprovsfaktorerna. Jämförelser mellan män och kvinnor i populationen med kontroll för betyg och selektionseffekter. Födda 1982.

Grupp	Män					Kvinnor				
	ORD	NOG	LÄS	DTK	ELF	ORD	NOG	LÄS	DTK	ELF
SU	-.369	-.404	-.231	-.369	-.198	-.351	-.537	-.389	-.447	-.093
IS	-.503	-.883	-.308	-.754	-.642	-.230	-.410	-.280	-.335	-.150
IU	-.767	-.757	-.530	-.711	-.434	-.497	-.652	-.419	-.481	-.081

Bilaga 2

Tabell 2.1. Delprovsmedelvärden i relation till invandrarstatus och kön. Födda 1977.

Grupp	Män					Kvinnor				
	ORD	NOG	LÄS	DTK	ELF	ORD	NOG	LÄS	DTK	ELF
SS	23,91	16,09	11,98	13,14	13,65	22,93	13,24	12,15	10,97	12,57
SU	21,20	14,49	11,00	11,96	13,91	19,74	11,09	11,00	9,47	12,57
IS	22,67	13,82	11,50	11,50	12,29	22,09	10,53	10,97	8,91	11,36
IU	19,74	14,00	10,73	11,05	13,40	19,20	11,15	10,78	9,41	12,37
Samtliga	23,71	15,96	11,91	13,04	13,63	22,74	13,09	12,07	10,86	12,54

Tabell 2.2. Delprovsmedelvärden i relation till invandrarstatus och kön. Födda 1982.

Grupp	Män					Kvinnor				
	ORD	NOG	LÄS	DTK	ELF	ORD	NOG	LÄS	DTK	ELF
SS	19,89	12,99	12,86	12,68	12,14	19,56	11,05	12,75	10,78	10,72
SU	17,95	11,48	11,88	11,19	11,25	17,43	9,14	11,15	8,96	9,95
IS	17,67	10,31	11,72	10,10	10,19	17,45	9,10	11,26	9,03	9,49
IU	16,64	10,46	11,09	10,11	10,63	17,19	9,06	11,27	8,97	10,14
Samtliga	19,68	12,82	12,75	12,51	12,03	19,38	10,89	12,62	10,63	10,65

Tabell 2.3. Sammanfattande mått på medelvärdeskillnaderna bland provtagarna. Intraklasskorrelationer.

Födelse- år	Män					Kvinnor				
	ORD	NOG	LÄS	DTK	ELF	ORD	NOG	LÄS	DTK	ELF
1977	.128	.105	.075	.121	.042	.116	.124	.088	.122	.041
1982	.116	.146	.104	.164	.088	.097	.137	.117	.141	.054

Referenser

- Allison, P. D. (1987). Estimation of linear models with incomplete data. I C. Clogg (Red.), *Sociological Methodology, 1987*, 71–103. San Francisco: Jossey Bass.
- Bobko, P., Roth, P. L. & Bobko, C. (2001). Correcting the effect size of d for range restriction and unreliability. *Organizational Research Methods, 4*(1), 46–61.
- Boldt, R. (1983). *Status of research on item content and differential performance on tests used in higher education*. (Research Report RR-83-3). Princeton, NJ: Educational Testing Service.
- Carlton, S. & Harris A. (1992). *Characteristics associated with differential item functioning on the Scholastic Aptitude Test: Gender and majority/minority group comparisons*. (Research Report RR-92-64). Princeton, NJ: Educational Testing Service.
- Cole, N. & Moss, P. A. (1989). Bias in test use. I R. L. Linn (Red.), *Educational measurement*. (3:e upplagan). New York och London: Macmillan.
- Freedle, R. & Kostin, I. (1997). Predicting black and white differential item functioning in verbal analogy performance. *Intelligence, 24*(3), 417–444.
- Gross, A. L. & McGanney, M. L. (1987). The restriction of range problem and nonignorable selection processes. *Journal of Applied Psychology, 72*(4), 604–610.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: John Wiley & Sons Inc.
- Gustafsson, J.-E. (1988). Hierarchical models of individual differences and cognitive abilities. I R. J. Sternberg (Red.), *Advances in the psychology of human intelligence*. Vol. 4, 35–71. Hillsdale, NJ: Erlbaum.
- Gustafsson, J.-E., Andersson, A. & Hansen, M. (2000). Prestationer och prestationsskillnader I 1990-talets skola. I J. Palme (Red.), *Välfärd och skola*. Stockholm: Norstedts Tryckeri AB.
- Gustafsson, J.-E. & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research, 28*, 407–434.

Gustafsson, J.-E. & Reuterberg, S.-E. (2000). Metodproblem vid studier av högskoleprovets prognosförmåga – och deras lösning! *Pedagogisk Forskning i Sverige*, 5(4), 273–283.

Gustafsson, J.-E. & Stahl, P. A. (2000). *STREAMS. User's guide. Version 2.5 for Windows 95/98/NT/2000*. Mölndal, Sverige: Multivariate Ware.

Gustafsson, J.-E., Wedman, I. & Westerlund, A. (1992). The dimensionality of the Swedish Scholastic Aptitude Test. *Scandinavian Journal of Educational Research*, 36(1), 21–39.

Holmes, D. J. (1990). The robustness of the usual correction for restriction of range due to explicit selection. *Psychometrika*, 55, 19–32.

Humphreys, L. G. (1986). An Analysis and Evaluation of Test and Item Bias in the Prediction Context. *Journal of Applied Psychology*, 71(2), 327–333.

Högskoleverket (1999). *Forsknings- och utvecklingsarbete om högskoleprovet*. (Inbjudan 1999-01-25). Stockholm: Högskoleverket.

Högskoleverket (2002). The Swedish national aptitude test: A 25-year testing program. Current status and future development. *Högskoleverkets rapportserie 2002:22R*. Stockholm: Högskoleverket.

Jöreskog, K. G. & Sörbom, D. (1996). *LISREL 8 User's Reference Guide (2nd ed.)*. Chicago: Scientific Software.

Linn, R. L. (1983). Predictive bias as an artifact of selection procedures. I H. Wainer & S. Messick (Red.), *Principles of Modern Psychological Measurement. A festschrift for Frederic M. Lord*. Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

Lord, F. M. & Novick, M. R. (1968). *Statistical analysis with missing data*. New York: Wiley.

Marculides, G. A. & Hersberger, S. L. (1997). *Multivariate Statistical Methods. A First Course*. N J: Mahwah, Lawrence Erlbaum Associate Publishers.

Maruyama, G. M. (1998). *Basics of Structural Equation Modeling*. Cal: Thousand Oaks, SAGE Publications.

Muthén, B. O., Kaplan, D. & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, 45, 19–30.

- Mäkitalo, Å. & Reuterberg, S.-E. (1996). *Who takes the Swedish Scholastic Aptitude Test? A study of differential selection to the SweSAT in relation to gender and ability*. (Reports from the Department of education and educational research, 1996:03). Göteborg: Göteborgs universitet, institutionen för pedagogik.
- Pedhazur, E. J. & Pedhazur Schmelkin, L. (1991). *Measurement, Design, and Analysis: An Integrated Approach*. Hillsdale, N.J.: Lawrence Erlbaum.
- Reuterberg, S.-E. (1996). Selektionsmekanismernas betydelse för gruppskillnader på högskoleprovet. I *Högskoleprovet genom elva forskares ögon*. Högskoleverkets rapportserie 1996:22 R. Stockholm: Högskoleverket.
- Reuterberg, S.-E. (1998). On differential selection in the Swedish Scholastic Aptitude Test. *Scandinavian Journal of Educational Research*, 42, 81–97.
- Reuterberg, S.-E. (1999). Textinnehåll och könsskillnader i delprovet engelsk läsförståelse (ELF). I *Fokus på högskoleprovet*. (Högskoleverkets rapportserie 1999:6S, 36–42). Stockholm: Högskoleverket.
- Reuterberg, S.-E. (2002). *Tre studier kring de språkliga delproven i högskoleprovet* (Rapport 2002:14). Göteborg: Göteborgs universitet, institutionen för pedagogik och didaktik.
- Reuterberg, S.-E. (2003). *Correcting validity coefficients for restriction of range by missing data modeling*. Manus.
- Reuterberg, S.-E. & Gustafsson, J.-E. (1992). Confirmatory factor analysis and reliability: Testing measurement model assumptions. *Educational and Psychological measurement*, (52), 795–811.
- Reuterberg, S.-E. & Hansen, M. (2001). *Vilken betydelse har utländsk bakgrund för resultatet på högskoleprovet?* Högskoleverkets rapportserie 2001:3 R. Stockholm: Högskoleverket.
- Roznowski, M. & Reith, J. (1999). Examining the measurement quality of tests containing differentially functioning items: Do biased items result in poor measurement? *Educational and Psychological Measurement*, 59(2), 248–269.
- Shepard, L., Camilli, G. & Averill, M. (1981). Comparisons of Procedures for Detecting Test-Item Bias with Both Internal and External Ability Criteria. *Journal of Educational Statistics*, 6(4), 317–375.

Schmitt, A. & Bleistein, C. A. (1987). *Factors affecting differential item functioning for black examinees on the Scholastic Aptitude Test analogy items*. (RR-87-23). Princeton, NJ: Educational Testing Service.

Schmitt, A. P., & Dorans, N. (1990). Differential item functioning for minority examinees on the SAT. *Journal of Educational Measurement*, 27(1), 67–81.

Stage, C. (1985). *Gruppskillnader i provresultat. Uppgiftsinnehållets betydelse för resultatskillnader mellan män och kvinnor på prov i ordkunskap och allmänorientering*. Umeå: Umeå universitet, pedagogiska institutionen.

Stage, C. & Ögren, G. (2001). *Högskoleprovets utveckling under åren 1977–2000. Provets sammansättning och provdeltagargruppens sammansättning och resultat*.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173–180.

Svensson, A. (2002). *Den sociala snedrekryteringen till högskolan – när och hur uppstår den?* (Rapport 2002:10). Göteborg: Göteborgs universitet, institutionen för pedagogik och didaktik.

Svensson, A., Gustafsson, J.-E. & Reuterberg, S.-E. (2001). *Högskoleprovets prognosvärde. Samband mellan provresultat och framgång första studieåret vid civilingenjör-, jurist- och grundskolläraryr utbildningarna*. Högskoleverkets rapportserie 2001:19 R. Stockholm: Högskoleverket.

Wainer, H. & Thissen, D. (1996). How is Reliability Related to the Quality of Test Scores? What is the Effect of Local Dependence on Reliability? *Educational Measurement: Issues and Practice*, 15, 22–29.

Willingham, W. W., Lewis, C., Morgan, R. & Ramist, L. (1990). *Predicting college grades. An analysis of institutional trends over two decades*. Princeton, NJ.: Educational Testing Service.

