

Simulerad antagning med IRT-baserade poäng på högskoleprovet

Per-Erik Lyrén¹

Inga Laukaityte¹

Mattias Wickberg²

Torbjörn Lindquist²

Håkan Pettersson²

¹ Institutionen för tillämpad utbildningsvetenskap, Umeå universitet.

² Universitets- och högskolerådet.

Abstract

This report analyses the potential effects of using Item Response Theory (IRT) instead of Classical Test Theory (CTT) for scoring the SweSAT (*Högskoleprovet*) in university admissions. Currently, SweSAT scores are based on the number of correct answers, whereas IRT offers advantages such as more precise estimates of test-takers' ability and a common scale for item difficulty and ability. To examine the consequences, a simulated admission process for autumn 2024 was conducted in which the standard CTT-based scores were replaced with IRT-based scores calculated using the three-parameter logistic model (3PL).

The simulation included data from 17 test administrations and approximately half a million unique test-takers, as well as application data from over 400,000 individuals. Results show that test scores changed for many applicants – 27% received a higher score and 30% a lower score – but the impact on admissions was minimal. Only 0.6 percent of admitted applicants changed “admission status,” meaning they either lost or gained a place. Differences by gender and age were negligible at the aggregate level, although some programs, such as dentistry and medicine, exhibited somewhat larger relative changes within specific subgroups.

The conclusion is that using IRT-based scoring would hardly affect the overall composition of admitted students. The study also identifies technical and data-related challenges that must be addressed before IRT can be implemented operationally, including the database structure. Future research should simulate the transition from the current scoring method to an IRT-based approach and compare outcomes across different IRT models.

Sammanfattning

Rapporten analyserar vilka effekter som användandet av Item Response Theory (IRT) istället för klassisk testteori (CTT) vid poängberäkning av högskoleprovet kan få för antagningen till högskolan. Högskoleprovet poängsätts i nuläget utifrån antalet rätta svar, medan IRT erbjuder fördelar som mer precisa skattningar av provtagarnas förmåga och en gemensam skala för uppgiftssvårighet och förmåga. För att undersöka konsekvenserna genomfördes en simulerad antagning till höstterminen 2024 där de ordinarie CTT-baserade poängen ersattes med IRT-baserade poäng beräknade med den tre-parametriska logistiska modellen (3PL).

Simuleringen omfattade data från 17 provomgångar och cirka en halv miljon unika provtagare, och ansökningsdata från drygt 400 000 personer. Resultaten visar att provpoängen förändras för många av de sökande – 27 procent fick en högre poäng och 30 procent en lägre poäng – men att effekterna på antagningen är mycket små. Endast 0,6 procent av de antagna byter "antagningsstatus", dvs. de förlorar eller får en plats. Skillnader i kön och ålder är obetydliga på övergripande nivå, även om vissa program, exempelvis tandläkar- och läkarutbildningen, uppvisar något större relativa förändringar i enskilda grupper.

Slutsatsen är att användandet av IRT-baserad poängsättning knappast skulle påverka utbildningarnas sammansättning. Studien identifierar också tekniska och datamässiga utmaningar som måste lösas innan IRT kan användas operativt, bland annat högskoleprovets databasstruktur. Vidare forskning bör bland annat simulera övergången från nuvarande poängsättning till en IRT-baserad sådan och jämföra utfallen från olika IRT-modeller.

Innehållsförteckning

Bakgrund.....	1
<i>Syfte.....</i>	<i>2</i>
<i>Rapportens disposition och författarbidrag</i>	<i>2</i>
Metod	3
<i>IRT-baserade HP-poäng</i>	<i>3</i>
<i>Simulerad antagning</i>	<i>5</i>
Så har simuleringarna gått till.....	5
Urval och urvalsgrupper	5
Effekter på antagningen i stort och på enskilda program	6
Resultat.....	7
<i>Provtagarna.....</i>	<i>7</i>
Poängfördelningar	8
<i>Resultat från simuleringen</i>	<i>12</i>
Sökande med HP-resultat	12
HP-poängen i urvalet hösten 2024	13
Skillnader i antalet antagna mellan SIM och BAS	13
Antagna till program	14
Sammanfattning av simuleringen	19
Diskussion	19
<i>Metod- och datafrågor.....</i>	<i>20</i>
<i>Begränsningar och vidare undersökningar.....</i>	<i>20</i>
Referenser.....	22

Bakgrund

Ända sedan högskoleprovet (HP) infördes på 1970-talet har det som inom psykometrin kallas för klassisk testteori (CTT) fungerat som ramverk för utvecklingen av provet. Under årens lopp har ett annat ramverk, item response theory (IRT), tillämpats i vissa sammanhang, framför allt kopplat till analys av uppgifters egenskaper och kvalitet, så kallad uppgiftsanalys. IRT-modellering kan även användas för att estimeras (skatta) provtagares resultat, ofta kallat förmåga eller *ability*, vilket något förenklat kan benämnas IRT-baserade poäng eller bara IRT-poäng. Fördelar med IRT-modellering jämfört med CTT är exempelvis att (1) det potentiellt ger mer information om uppgifternas kvalitet, (2) uppgifternas svårighetsgrad och provtagarnas förmåga är på samma skala, (3) mätningens precision som standard anges för skattningarna av provtagarnas förmåga hellre än för mätningen i stort, (4) det går att analysera hur väl uppgifter (och provtagare) passar den valda modellen, och (5) det potentiellt ger mer valida skattningar av provdeltagarnas förmåga. Nackdelar finns förstås också, exempelvis att det krävs stora mängder data för tillförlitliga skattningar, och att skattningarna av vissa uppgiftsparametrar överlag är behäftade med stora osäkerheter. Ytterligare ett viktigt skäl till att IRT är av intresse för storskaliga prov är att det är i princip en förutsättning för datoriserade prov med någon form av inbyggd adaptivitet. Detta oavsett om det gäller ”rena” adaptiva prov, där ett beslut om vilken uppgift som kommer närmast görs efter varje uppgift, eller så kallade multistage-prov (MST; se t.ex. Zenisky, Hambleton & Luecht, 2009).

En CTT-analog till IRT-modeller (logistisk regression med observerade poäng som oberoende) används sedan några år tillbaka inom arbetet med högskoleprovets uppgiftsanalys. Det används även ”riktig” IRT-modellering för uppgiftsanalys, huvudsakligen för analyser av hur distraktorer fungerar. Experimentella försök har också gjorts när det gäller att använda IRT i ekvivaleringen av provet, både för det ”gamla” provet (före hösten 2011) och den nuvarande provutformningen. Som nämndes tidigare finns vissa fördelar med IRT-baserade poäng, och det är dessa fördelar som föranleder intresset för IRT-modellering i stort och som utgör det bakomliggande skälet till den föreslagna studien. Datoriseringsperspektivet är intressant men underordnat de andra skälen, givet att det i dagsläget inte föreligger något beslut om att datorisera högskoleprovet.

Förhållandet mellan CTT-poäng (antalet rätta svar) och IRT-poäng skiljer sig något beroende på vilken IRT-modell som används. För 1PL-modellen (Rasch), där den enda modellerade uppgiftsparametern är svårighetsgraden, så är Pearsonkorrelationen mellan CTT-poäng och IRT-poäng så hög som 0,98-0,99 (Fan, 1998; Macdonald & Paunonen, 2002) och rangkorrelationen

är 1,0, dvs. rangordningen utifrån poängen är exakt samma för de två typerna av poäng. Detta är dock under förutsättning att analyserade data inte innehåller saknade värden, vilket inte är fallet med högskoleprovet (där saknade svar räknas som felsvar i rättningen). Det skulle dock i ett operativt läge kunna finnas skillnader i normerat resultat utifrån hur normeringsprocessen går till, särskilt då om IRT skulle användas för ekvivalering av resultaten från olika provomgångar. Med 2PL- och 3PL-modellerna (se beskrivning av 3PL-modellen nedan) skulle skillnaden kunna bli större, eftersom den skattade IRT-poängen för en provtagare i de fallen beror på vilka uppgifter som hen svarat rätt på. Två provtagare som har exakt samma antal rätta svar på samma prov men som svarat rätt på olika uppgifter kan alltså ha något olika skattad IRT-poäng. Korrelationen mellan CTT-poäng och IRT-poäng är hög även i dessa fall, kring 0,97-0,98 (Fan, 1998; Macdonald & Paunonen, 2002). Det finns dock problem med att fokusera på korrelationsmättet eftersom det inte säger något om variationer i variabelförhållandet på olika delar av skalan. Fokuserar vi exempelvis på de högst presterande så kommer korrelationen bland provtagarna på den delen av skalan att vara lägre, på grund av så kallad begränsad variationsbredd (*restriction of range*; se t.ex. Dahlke, Sackett, & Kuncel, 2019; Lyrén, 2008). Under sådana förutsättningar kan det mycket väl finnas skillnader som skulle få konsekvenser i termer av vem som hamnar på "rätt" sida i urvalet till en utbildning och vem som inte gör det.

Syfte

Syftet med studien är att undersöka effekterna av att använda IRT-baserade HP-poäng på utfallet av antagningen till högskolan. Mer specifika frågeställningar av intresse är:

- a) Hur skulle gruppen antagna förändras? Det vill säga, hur stor andel skulle bli antagna som annars inte skulle ha blivit antagna, och vice versa?
- b) Finns det skillnader i effekter mellan olika program/inriktningar?
- c) Förändras gruppen antagna med avseende på kön och ålder?

Rapportens disposition och författarbidrag

Studien består av två huvudsakliga delar: en generell del som beskriver hur de IRT-baserade poängen skattas och förhåller sig till CTT-baserade poäng, och en del som beskriver och redovisar resultat från en simulerad antagning som är baserad på IRT-baserade HP-poäng. PEL står för beskrivning av bakgrund, syfte, generell metodbeskrivning och diskussion, IL står för metodbeskrivning och framtagning av de IRT-baserade poängen och statistik kring dessa, MW och TL står för beskrivning och sammanställning av resultat från simuleringen, och HP står för det tekniska i simuleringen.

Metod

Frågeställningarna undersöks genom en simulerad antagning där IRT-baserade HP-poäng används istället för befintliga CTT-baserade poäng. Det faktum att giltighetstiden för provresultaten är hela 8 år komplicerar studien avsevärt, framför allt genom att alla provtagares poäng måste vara jämförbara. Detta innebär att för respektive provdel, den kvantitativa (KVANT) och den verbala (VERB), så måste data bestående av $17 \cdot 80 + 2 \cdot 40 = 1420$ uppgifter och i storleksordningen 500 000 provtagare analyseras sammantaget.

IRT-baserade HP-poäng

Proceduren för att få fram IRT-poängen omfattar flera steg:

1. Val av IRT-modell(er)
2. Framtagande av data
3. Kalibrering av uppgifter
4. Skattning av provtagarnas förmåga
5. Skalning och ekvivalering

När det gäller (1) så föll valet ursprungligen på både den tvåparametriska och den treparametriska logistiska IRT-modellen (2PL- och 3PL-modellen). Proceduren därefter i form av (3) och (4) och produktionen av datafiler till antagningen tog emellertid längre tid än förväntat (bland annat p.g.a. de speciallösningar som gjordes vissa provomgångar för att minska effekterna av storskaligt fusk); detta gjorde att vi nöjde oss med en av modellerna, vilket blev 3PL-modellen. När det gäller (5) så valde vi att använda oss av de befintliga poängfördelningarna för respektive provomgång, vilket är rimligt givet att den generella prestationsnivån i provtagargruppen knappast är beroende av vilken specifik modell som används för att poängsätta prestationer på individnivå.

IRT-baserade HP-poäng för varje testadministration under den studerade perioden erhöles alltså med hjälp av 3PL-modellen. Enligt den definieras sannolikheten att personer med en förmåga som är lika med person j 's förmåga att korrekt besvara uppgift i som:

$$P_{ij} = P(X_{ij} = 1 | \theta_j, a_i, b_i, c_i) = c_i + (1 - c_i) \frac{e^{Da_i(\theta_j - b_i)}}{1 + e^{Da_i(\theta_j - b_i)'}}$$

där θ_j är förmågeparametern för person j , D är konstanten som vanligtvis sätts till 1,7, medan b_i är svårighetsparametern, a_i är diskrimineringsparametern och c_i är pseudogissningsparametern för uppgift i . Pseudogissningsparametern motsvarar en nedre asymptot för den logistiska funktionen på så sätt att den indikerar sannolikheten för provtagarna med låg förmåga (låga θ) att svara rätt på den aktuella uppgiften:

$$\lim_{\theta \rightarrow -\infty} P(X_i = 1 | \theta, a_i, b_i, c_i) = c_i.$$

För att uppskatta förmågan θ (eller IRT-baserade HP-poäng) används likelihoodfunktionen:

$$L(\mathbf{u} | \theta_j) = \prod_{i=1}^N P_{ij}^{u_{ij}} Q_{ij}^{1-u_{ij}}$$

där $Q_{ij} = 1 - P_{ij}$, u_{ij} är person j 's poäng (0 eller 1) på uppgift i , och $\mathbf{u}$ är en vektor med 1:or och 0:or som representerar person j 's poäng på uppgifterna 1 till N . Om vi exempelvis har ett prov med fem uppgifter och en vektor som är (1 1 0 0 1) för en viss person så kommer likelihoodfunktionen att beskriva produkten av fem sannolikheter, nämligen att få 1 poäng på uppgifterna 1, 2 och 5 och 0 poäng på uppgifterna 3 och 4. Både likelihoodfunktion och logaritmen av denna (log-likelihoodfunktionen) kan användas för att skatta θ ; i den här studien används log-likelihoodfunktionen.

För att garantera rättvisa och jämförbarhet är alla versioner av provet ekvivalerade. Men som nämnts tidigare valde vi att inte göra en regelrätt ekvivalering av de IRT-baserade HP-poängen, θ , utan att istället skala dem på ett sådant sätt att de skalade θ -värdena har samma fördelning som de ursprungliga skalade värdena för varje provadministration. Det vill säga, för varje provtillfälle så är fördelningen av ursprungliga HP-poäng och fördelningen av IRT-baserade HP-poäng så lika varandra som möjligt.

Det fanns några utmaningar i arbetet med att erhålla de IRT-baserade poängen. För det första: för att erhålla IRT-poäng måste data innehålla poängsatta frågor, dvs. binära variabler med värdena 0 och 1 för alla provtagare. Nödvändiga data finns, men den uppdateras inte efter senare möjliga interventioner begärda av UHR. Till exempel, om UHR beslutar att återkalla vissa delar av ett prov för vissa deltagare, uppdateras endast summapoängen, men inte poängen på enskilda frågor. Det är också omöjligt att veta vilka frågor

som ska ha en ändrad poäng. Detta är ett viktigt problem som bör lösas i framtiden. För det andra: det var mycket komplicerat att skapa filer med både de binärt poängsatta frågorna och bakgrundsinformation som behövs för simuleringen av antagning. Flera tabeller med olika format måste slås samman, och vissa av tabellerna uppdaterades aldrig efter den rutinmässiga proceduren. Alla dessa komplikationer innebär att HP-databasen behöver väsentliga förändringar innan IRT-poängsättning kan användas operativt.

Framtagandet av de IRT-baserade HP-poängen har gjorts med hjälp av R (R Core Team, 2023) och paketet mirt (Chalmers, 2012).

Simulerad antagning

För att kunna analysera konsekvenserna av att använda IRT-baserade HP-poäng i antagningen så är en lämplig metod att med hjälp av en simulerad antagning ändra förutsättningar i proceduren, dvs. att byta ut sökandes ordinarie högskoleprovsresultat och se hur detta påverkar antagningsresultaten. Resultatet av den simulerade antagningen kan därefter jämföras med resultatet från den faktiska antagningen. Metoden innebär alltså att jämföra de personer som hade erbjudits en plats vid simuleringen med de personer som faktiskt blev antagna. Att göra en simulering har även den fördelen att vi kan se vad som händer i hela antagningssystemet om en ändring görs. Det kan till exempel vara så att en person som inte antas på sitt högskoleprovsresultat istället antas på sitt betyg. Simuleringen ger information om utfallet i samtliga urvalsgrupper och därmed den sökandes totala chans att komma in på sökt utbildning med eller utan ändringar i HP-poäng.

Så har simuleringarna gått till

För varje utbildning som studerats i simuleringen av antagningsomgången höstterminen 2024 så har det ursprungliga värdet för urvalsgruppen högskoleprovet bytts ut med den IRT-baserade HP-poängen. För att kunna redovisa hur gruppen antagna skulle påverkas måste vi veta hur det såg ut i antagningen före förändringarna. Inför simuleringen behöver vi därför ett referensvärde, eller "nollvärde". Den faktiska antagningen, kallad BAS, är nollvärdet som resultatet av den simulerade antagningen, SIM, jämförs med.

Urval och urvalsgrupper

Antalet platser på en utbildning kan i vissa fall vara färre än antalet sökande. Om det finns fler behöriga sökande än det finns platser måste det ske ett urval. Hur platserna fördelas till utbildningar på grundnivå är reglerat i 7 kap 13 § högskoleförordningen (1993:100). Enligt 13 § ska de befintliga platserna fördelas mellan sökande med betyg, högskoleprov och, om högskolan så bestämmer, även andra meriter än de båda nämnda. Minst en tredjedel av platserna ska fördelas till sökande på grundval av resultat från

högskoleprovet och minst en tredjedel på grundval av betyg. Hur olika betyg ska meritvärderas regleras i huvudsak i bilaga 3 till högskoleförordningen. Meritvärderingen av betyg är en intrikat process där de olika betygen och deras värden i antagningen anges i form av urvalsgrupp och meritvärde. För sökande med omdöme från folkhögskola är urvalsgruppen BF. Maxvärdet den sökande kan ha i urvalsgruppen är 4,00. Sökande med meriter från gymnasieskolan eller kommunal vuxenutbildning har antingen kompletterat sina betyg och placeras i urvalsgruppen BII eller så har de betyg utan komplettering och placeras i urvalsgruppen BI. Det maximala meritvärdet en sökande med meriter från gymnasieskolan eller kommunal vuxenutbildning kan ha i antagningen är 22,50. Även högskoleprovet är en urvalsgrupp: HP.

Det finns även derivat av de ovanstående urvalsgrupperna. Exempelvis använder Göteborgs universitet högskoleprovet som skiljekriterium vid lika meritvärde till hälften av platserna i betygssurvalet till läkarutbildningen. Eftersom det då är frågan om en avvikelse från den vanliga hanteringen i betygssurvalet så hanteras det genom att placera dessa sökande i en annan urvalsgrupp än BF, BI eller BII. Alla andra urvalsgrupper än BF, BI, BII och HP slås samman i en övrig grupp.

Effekter på antagningen i stort och på enskilda program

Det är inte görligt, och sannolikt inte heller av intresse, att gå igenom sammansättningen av de sökande till varje av de cirka 14 200 utbildningstillfällena som gavs höstterminen 2024 då det till 8 av 10 av dem saknas konkurrens, dvs det finns färre sökande än tillgängliga platser. Saknas konkurrens har inte den sökandes HP-poäng någon betydelse.

Det är istället relevant att fokusera på de utbildningar där det skett ett urval och där HP-poängen har betydelse. Även med en sådan avgränsning blir emellertid antalet utbildningstillfällen, drygt 4 600, för omfattande att studera.

Av metodologiska skäl analyseras därför endast utfallet för programmen till följande åtta yrken: civilingenjör, förskollärare, jurist, läkare, sjuksköterska, socionom, tandläkare och ämneslärare. Dessa yrkesprogram representerar sammantaget en bredd av vetenskapsområden, de har konkurrens om utbildningsplatserna över tid, och för vart och ett av dem gäller något av följande:

- det har ett relativt stort antal utbildningsplatser,
- det leder till examina inom ett yrke som Arbetsförmedlingen bedömt har goda framtida arbetsmarknadsutsikter.

Det är fullt möjligt att analysera samtliga programtillfällen ht 2024 för dessa åtta utbildningar, men eftersom sökande i regel söker fler än ett programtillfälle för samma utbildning så finns ett problem med dubletter. Till civilingenjör har till exempel 3 av 4 anmälda sökt fler än en utbildning som leder till denna examen (Tabell 1). I resultatredovisningen nettoräknas därför de sökande per examen som utbildningen leder till om inte annat anges.

Tabell 1. Sökande och antal valda alternativ mot examen höstterminen 2024.

Examen	Sökande	Varav		
		I 1:a hand	1 sökt alternativ	≥2 sökta alternativ
Civilingenjör	25 517	16 762	6 446	19 071
Förskollärare	7 434	3 927	5 093	2 341
Jurist/Juris kandidat	15 708	8 242	7 859	7 849
Läkare	9 104	6 848	3 080	6 024
Sjuksköterska	19 755	11 848	7 904	11 851
Socionom	20 432	8 544	10 495	9 937
Tandläkare	3 501	1 535	1 645	1 856
Ämneslärare (avancerad nivå)	13 768	6 841	6 367	7 401
Totalt	115 219	64 547	48 889	66 330

Av tekniska skäl så behåller utländska sökande och sökande med sekretess sitt värde från BAS även i SIM. Av dessa totalt 311 sökande så var det 154 som hade en IRT-baserad HP-poäng som skilde sig åt från den CTT-baserade.

Analyserna av resultaten från simuleringen har gjorts med IBM SPSS Statistics (version 29.0; IBM Corp., 2023).

Resultat

Provtagarna

Nästan 800 000 personer har gjort Högskoleprovet under perioden våren 2016 – våren 2024, se Tabell 2. 99,2 procent av alla provtagare hade ett svenskt personnummer. En del provtagare har gjort provet fler än en gång, vilket innebär att antalet provtagare inte representerar antalet unika provtagare. Därför var endast 59,4 procent av provtagarna med svenskt personnummer unika provtagare under den studerade perioden. Vanligtvis är lite mer än hälften av HP-provtagarna kvinnor. Åldersmässigt är i genomsnitt 59 procent av provtagarna högst 20 år. Den stora majoriteten av provtagare, cirka 80 procent, har 3–4-årigt gymnasium som högsta utbildning.

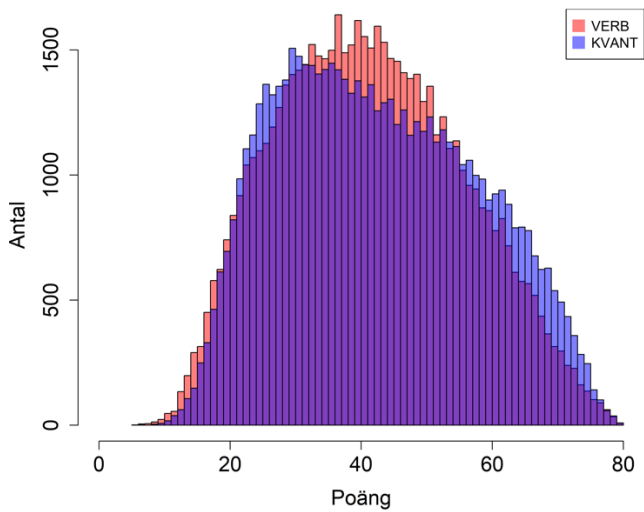
Tabell 2. Fördelning av provtagarnas kön, ålder och utbildning.

Prov	N	Kön	Ålder						UTB MAX					
		M-K	-20	21-24	25-29	30-39	40-	G/R	FHS	G2	G34	≤120	>120	
16:A	75 313	49-51	52	28	12	6	2	1	2	4	78	10	4	
16:B	55 072	48-52	58	23	10	6	2	3	1	3	76	9	5	
17:A	64 962	49-51	54	26	12	6	3	2	2	4	78	9	4	
17:B	46 138	46-54	62	21	9	5	2	2	2	3	79	8	5	
18:A	62 678	49-51	57	25	11	6	2	2	2	3	79	8	4	
18:B	39 246	47-53	62	21	9	6	3	2	2	3	79	8	5	
19:A	55 438	50-50	58	24	10	6	2	2	2	3	79	8	4	
19:B	41 571	47-53	64	19	9	5	2	2	1	3	80	7	4	
20:B	23 288	47-53	69	13	7	7	4	2	2	4	81	5	4	
21:A1	21 871	45-55	51	30	11	6	2	2	1	3	77	10	4	
21:A2	23 556	45-55	52	30	11	6	2	2	2	3	77	10	4	
21:B	28 165	44-56	45	30	13	9	3	2	2	5	70	12	6	
22:A1	29 314	46-54	61	24	8	5	2	2	2	3	80	8	3	
22:A2	28 438	46-54	62	24	8	5	2	2	2	3	80	8	3	
22:B	38 477	45-55	66	20	7	5	2	2	1	3	79	8	4	
23:A	58 990	48-52	63	23	7	5	2	2	1	3	81	7	3	
23:B	42 076	46-54	68	19	6	4	2	2	1	3	81	7	4	
24:A	62 634	48-52	63	25	6	4	2	2	1	3	82	7	3	

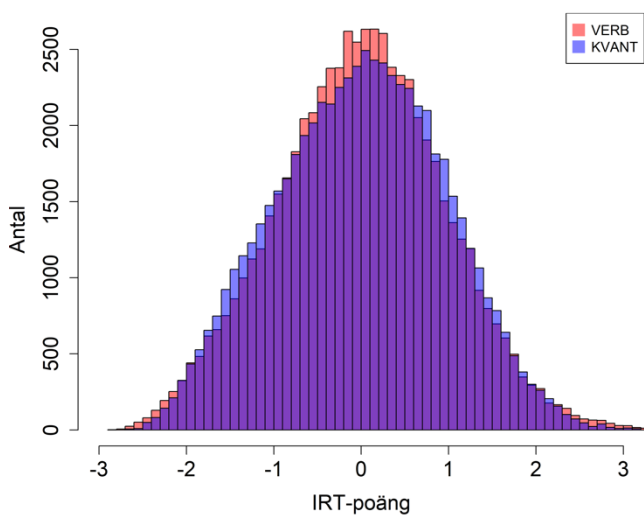
Poängfördelningar

Fördelningar av rå- och IRT-poäng presenteras i Figur 1. Råpoängen (Figur 1 a) är mer snedfördelade än IRT-poängen (Figur 1 b), och fördelningen av den kvantitativa poängen är något plattare än fördelningen av den verbala poängen, särskilt när det gäller råpoängen. Fördelningarna av IRT-poängen ser mer normalfördelade ut, vilket är i linje med modellens antagande om att den latenta förmågan följer en standardnormal fördelning (med medelvärde 0 och standardavvikelse 1).

Figur 2 visar hur fördelningarna av normerade poäng ser ut. Fördelningarna av normerade rå- och IRT-poäng är likadana per design. Även om fördelningarna av normerade rå- och IRT-poäng totalt sett ser likadana ut, finns det vissa skillnader på provtagarnas nivå. En del provtagare får bättre eller sämre resultat när IRT används för poängsättning. Figur 3 visar hur relationen mellan normerade rå- (CTT) och IRT-poäng ser ut för de två provdelarna våren 2024. Det är tydligt att de flesta (ca 60 procent) av provtagarna får samma normerade poäng oavsett poängsättningsmetod. Däremot är spridningen av avvikande poäng bredare på den kvantitativa provdelen jämfört med den verbala provdelen.

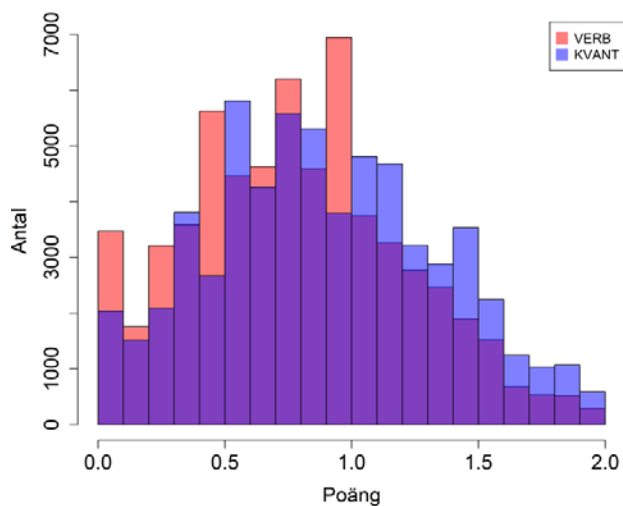


(a)

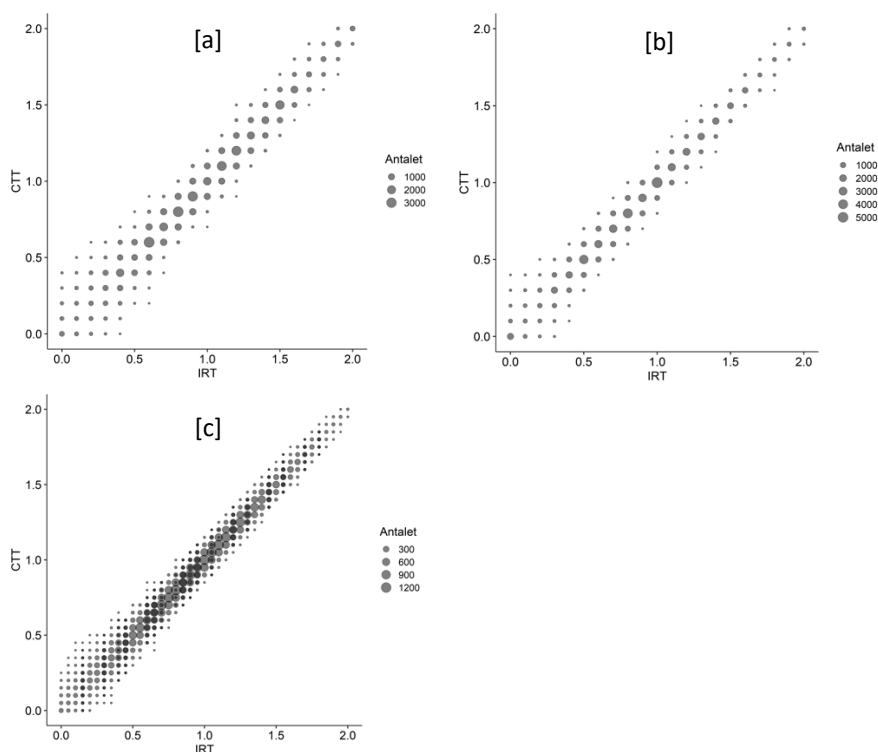


(b)

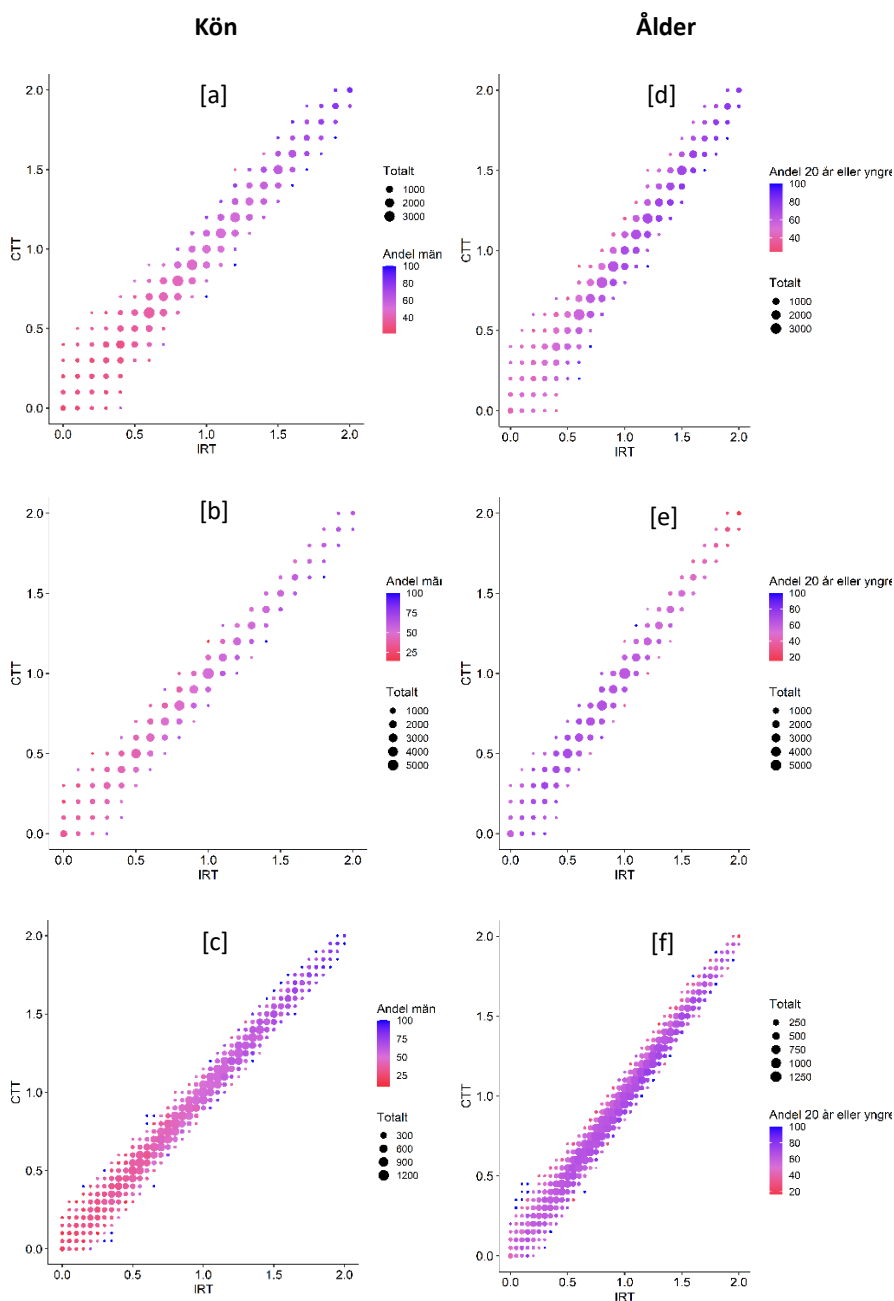
Figur 1. Fördelningar våren 2024 av högskoleprovets kvantitativa och verbala poäng: a) råpoäng (antalet rätta svar), och b) IRT-baserad poäng.



Figur 2. Poängfördelningar av kvantitativa och verbala normerade poäng våren 2024.



Figur 3. Relationen mellan normerade CTT- och IRT-poäng för provet våren 2024: provdelarna kvant [a] och verb [b], samt hela provet [c].



Figur 4. Relationen mellan normerade CTT- och IRT-poäng i relation till kön respektive ålder för provet våren 2024: provdelarna kvant ([a] respektive [d]) och verb ([b] respektive [e]), samt hela provet ([c] respektive [f]).

I Figur 4 (a-c) redovisas hur män respektive kvinnor är procentuellt fördelade på poängskalan. På den kvantitativa provdelen är kvinnorna över-representerade bland de lägsta normerade poängen, medan män är i majoritet bland de högsta. På den verbala provdelen är skillnaden mellan könen betydligt mindre, men även här är den till männens fördel.

Tidigare analyser av provtagares prestationer har visat att det finns ett visst samband mellan ålder och provresultat (se t.ex. Wikström, Laukaityte & Lyrén, 2024). I Figur 4 (d-f) visas relationen mellan normerade CTT- och IRT-poäng i relation till åldersgrupper, där de yngsta provtagarna (20 år eller yngre) jämförs med alla andra (äldre än 20 år). Man kan se att de yngsta provtagarna presterar bättre på den kvantitativa provdelen, medan de äldre presterar bättre på den verbala provdelen.

Resultat från simuleringen

Sökande med HP-resultat

I antagningen hösten 2024 leder den förändrade beräkningen av HP-poäng till att 27 procent får en högre HP-poäng i SIM än i BAS, 30 procent får en lägre HP-poäng medan 43 procent har HP-poäng som inte påverkas alls (tabell 3). Detta gäller oavsett om sökandegruppen som studeras är alla som överhuvudtaget sökte en högskoleutbildning (kolumnen ”Alla sökande”) eller endast de som sökte en utbildning där ett resultat på högskoleprovet är av potentiell betydelse (kolumnen ”Sökande HP-urval”). Här kan också noteras att det totala antalet sökande till högskolan var 402 162, vilket innebär att 41 procent av alla sökande hade ett giltigt resultat på högskoleprovet.

Tabell 3. Utfall av förändrad beräkning av HP-poäng

	Sökande HP-urval		Alla sökande	
	Antal	Andel (%)	Antal	Andel (%)
Högre HP-poäng	32 011	27	44 827	27
Oförändrad HP-poäng	51 120	43	71 284	43
Lägre HP-poäng	34 964	30	49 708	30
Totalt	118 095	100	165 819	100

En förändring i metoden att beräkna HP-poängen kan även leda till att vissa provtagare som i BAS har ett bästa resultat från provtillfälle X, vid omräkningen av HP-poängen i SIM har ett bästa resultat från provtillfälle Y. Vid antagningen till hösten 2024 har drygt hälften av de sökande med ett giltigt högskoleprovsresultat genomfört provet vid fler än ett tillfälle, och det är cirka 9 procent av de sökande med HP-poäng som konkurrerar med

resultatet från annat provtillfälle i SIM än i BAS. 4 av 10 av dessa (eller 3,3 procent av samtliga sökande med HP-poäng) har dock samma HP-poäng i SIM som i BAS, men från olika provtillfällen.

264 sökande i BAS som har ett högskoleprovsresultat finns inte med i SIM i urvalsgrupp HP. För en majoritet av dessa, 146 sökande, är anledningen att det nya HP-värdet blir för lågt för att delta i urvalet, dvs. är under 0,05. Det finns emellertid andra orsaker också. Vissa skillnader mellan BAS och SIM beror på att uttaget från antagningsdatabasen är gjort vid ett senare tillfälle till SIM än till BAS. Ett exempel är en sökande med 1,65 i BAS som har sökt endast en utbildning och direktantagits i SIM.

HP-poängen i urvalet hösten 2024

I Tabell 4 visas sökandegruppens HP-poäng i BAS och i SIM. Sökandegruppen definieras som sökande med HP-poäng som söker till minst en utbildning där urvalsgruppen HP ingår. Överlag är det mycket små skillnader mellan SIM och BAS när det gäller de sökandes HP-poäng. Tabellen visar att det i regel är nödvändigt att gå ner på tusendelar för att notera en skillnad i HP-poäng. Men en liten förskjutning kan noteras till yngre sökande (- 20 år), både kvinnor och män, som får ett bättre resultat i SIM att konkurrera med.

Tabell 4. Medelvärde för HP-poängen för sökandegruppen höstterminen 2024.

	BAS			SIM		
	Kvinnor	Män	Totalt	Kvinnor	Män	Totalt
≤ 20 år	0,85	1,04	0,94	0,85	1,05	0,94
21–24 år	0,82	0,98	0,89	0,82	0,98	0,89
25–29 år	0,80	0,97	0,87	0,80	0,97	0,87
30–39 år	0,81	1,04	0,90	0,80	1,03	0,90
≥ 40 år	0,94	1,24	1,06	0,94	1,23	1,05
Totalt	0,83	1,01	0,91	0,83	1,01	0,91

Att medelvärdet av den HP-poäng som de sökande konkurrerar med är så gott som oförändrat i SIM jämfört med BAS innebär dock inte nödvändigtvis att det inte kommer att ske några förändringar i sammansättningen av antagna till hösten 2024. Antagningen är en komplicerad process där många faktorer har betydelse för den sökandes möjlighet att antas.

Skillnader i antalet antagna mellan SIM och BAS

Det totala antalet antagna i antagningsomgången till hösten 2024 förändras mycket marginellt av en ändring i hur HP-poängen beräknas. Tabell 5 visar att det rör sig om totalt endast 55 färre antagna i SIM än i BAS. Att antalet

antagna inte är exakt detsamma i SIM som i BAS kan hänföras till 'marginal-effekter' utifrån de sökandes antal sökta utbildningar. Tabellen visar också att det sker en omflyttning av 2 017 individer; 1 036 som var antagna förlorar sin plats, 981 som inte tidigare var antagna erbjuds nu en plats på grund av förändrat HP-värde. Resterande 323 103 antagna i BAS förblir antagna i SIM. Det innebär att möjligheten att bli antagen påverkas för cirka 0,6 procent av de antagna.

Tabell 5. Antal antagna och ej antagna i BAS och SIM höstterminen 2024.

		SIM		Totalt
		Ej antagen	Antagen	
BAS	Ej antagen	77 042	981	78 023
	Antagen	1 036	323 103	324 139
	Totalt	78 078	324 084	402 162

Tabell 6 visar fördelningen av de 2 017 individerna efter ålder och kön. Kolumnen "BAS, ej SIM" i tabellen visar de 1 036 antagna i BAS som förlorat sin antagning och som inte har antagits i SIM (till minst en utbildning). Kolumnen "SIM, ej BAS" i tabellen visar de 981 antagna i SIM som inte tidigare varit antagna i BAS men som nu har antagits till (minst) en utbildning i SIM. Det är, relativt sett, något fler yngre kvinnor och något färre kvinnor i åldersgruppen 25–29 år som antagits i SIM, och det är totalt sett något fler kvinnor än män som förlorar sin plats i SIM.

Tabell 6. Antal antagna i BAS och SIM höstterminen 2024 vars möjlighet till antagning till (någon) utbildning påverkas.

	BAS, ej SIM			SIM, ej BAS		
	Kvinnor	Män	Totalt	Kvinnor	Män	Totalt
≤ 20 år	104	137	241	127	129	256
21–24 år	185	195	380	186	174	360
25–29 år	129	65	194	90	59	149
30–39 år	90	46	136	88	59	147
≥ 40 år	66	19	85	47	22	69
Totalt	574	462	1 036	538	443	981

Antagna till program

Det har framgått att den sökandes möjlighet att antas till någon utbildning som denne har sökt näst intill inte påverkas alls av en ändrad HP-poäng. Sökande tenderar emellertid att anmäla sig till fler än en utbildning (se Tabell 1), och det kan därför hända att det är olika utbildningar som den sökande är

antagen till i BAS och SIM. Det är inte osannolikt att det sker en omflyttning av sökande mellan utbildningarna i SIM; ett exempel är en sökande som i BAS var antagen till önskad juristutbildning men som i SIM istället blir antagen till en fristående kurs då HP-poängen inte räckte till för juristutbildningen efter ändringen.

Tabell 7 visar att en förändring av HP-poäng generellt påverkar sammansättningen marginellt även på programnivå. Den förändring som sker mellan könen gäller för kvinnor som sökt till tandläkarutbildningen; där ökar andelen kvinnor med 2 procentenheter från 63 procent till 65 procent. Ingen förändring av betydelse med avseende på kön sker för övriga utbildningar av en ändrad HP-poäng.

Tabell 7. Antal antagna höstterminen 2024 i BAS, samt ändringar i antalet antagna från BAS till SIM, uppdelat efter program (examen), ålder och kön.

Examen	Ålder	BAS (antal)		SIM (förändring)			
		Kvinnor	Män	Kvinnor		Män	
				Ant.	%	Ant.	%
Civilingenjör	20 år och yngre	2 653	5 253	12	0,5	5	0,1
	21–24 år	1 511	2 286	-14	-0,9	-7	-0,3
	25–29 år	141	357	3	2,1	3	0,8
	30–39 år	50	154	3	6,0	-4	-2,6
	40 år och äldre	12	42	0	0,0	-1	-2,4
	Totalt	4 367	8 092	4	0,1	-4	-0,0
Förskollärare	20 år och yngre	583	31	-2	-0,3	0	0,0
	21–24 år	887	59	0	0,0	-1	-1,7
	25–29 år	566	38	4	0,7	0	0,0
	30–39 år	888	52	-1	-0,1	0	0,0
	40 år och äldre	330	20	-1	-0,3	0	0,0
	Totalt	3 254	200	0	0,0	-1	-0,5
Jurist/ Juris kandidat	20 år och yngre	817	331	-4	-0,5	3	0,9
	21–24 år	384	213	-6	-1,6	-4	-1,9
	25–29 år	67	60	2	3,0	1	1,7
	30–39 år	63	63	4	6,3	2	3,2
	40 år och äldre	25	27	1	4,0	1	3,7
	Totalt	1 356	694	-3	-0,2	3	0,4
Läkare	20 år och yngre	434	265	-4	-0,9	-1	-0,4
	21–24 år	235	166	0	0,0	7	4,2
	25–29 år	56	74	-1	-1,8	4	5,4
	30–39 år	54	65	1	1,9	-1	-1,5
	40 år och äldre	17	27	-3	-17,6	-2	-7,4
	Totalt	796	597	-7	-0,9	7	1,2

En förändrad HP-poäng leder till marginellt fler antagna bland de allra yngsta, 20-år och yngre, till civilingenjör-, socionom- och tandläkarutbildningarna. Det omvända gäller för juristutbildningarna där det i SIM antas fler som är 25 år eller äldre. Men som sagt, förändringarna är generellt mycket små.

Tabell 7 (forts.). Antal antagna höstterminen 2024 i BAS samt ändringar i antalet antagna från BAS till SIM, uppdelat efter program (examen), ålder och kön.

Examen	Ålder	BAS (antal)		SIM (förändring)			
		Kvinnor	Män	Kvinnor		Män	
				Ant.	%	Ant.	%
Sjuksköterska	20 år och yngre	1 122	136	-2	-0,2	0	0,0
	21–24 år	1 197	206	7	0,6	1	0,5
	25–29 år	534	125	-8	-1,5	0	0,0
	30–39 år	731	178	5	0,7	0	0,0
	40 år och äldre	457	118	-4	-0,9	1	0,8
	Totalt	4 041	763	-2	-0,0	2	0,3
Socionom	20 år och yngre	924	90	0	0,0	6	6,7
	21–24 år	774	130	-3	-0,4	1	0,8
	25–29 år	305	95	0	0,0	-2	-2,1
	30–39 år	332	101	-5	-1,5	2	2,0
	40 år och äldre	170	45	1	0,6	0	0,0
	Totalt	2 505	461	-7	-0,3	7	1,5
Tandläkare	20 år och yngre	104	63	4	3,8	0	0,0
	21–24 år	51	34	3	5,9	-4	-11,8
	25–29 år	23	11	-1	-4,3	-2	-18,2
	30–39 år	12	4	0	0,0	0	0,0
	40 år och äldre	4	2	0	0,0	0	0,0
	Totalt	194	114	6	3,1	-6	-5,3
Ämneslärare (avancerad nivå)	20 år och yngre	1 132	602	-2	-0,2	-1	-0,2
	21–24 år	1 057	799	-1	-0,1	3	0,4
	25–29 år	477	436	2	0,4	0	0,0
	30–39 år	542	394	0	0,0	3	0,8
	40 år och äldre	502	310	2	0,4	-1	-0,3
	Totalt	3 710	2 541	1	0,0	4	0,2

Det är tydligt att en ändring i hur HP-poängen beräknas nästintill inte har någon påverkan på sammansättning av de antagna med avseende på kön och ålder. Det gäller såväl sett över hela sökandegruppen som på programnivå. Även om en ändring i hur HP-poängen beräknas inte förändrar sammansättningen av antagna på utbildningarna så kan en sådan ändring ha stor betydelse för den enskilde. En antagen i BAS som får en sämre HP-poäng i SIM kan förlora sin plats och ersättas av en person av samma kön i samma ålder.

I ett sådant fall påverkas inte sammansättningen på utbildningen med avseende på ålder och kön, men för de enskilda sökande vars antagningsstatus förändras så är betydelsen så klart stor.

Tabell 8 visar att den som var antagen i BAS i regel även är antagen i SIM till samma utbildning. Den största förändringen sker bland de antagna till tandläkarprogrammet och läkarprogrammet där 6,2 procent respektive 4,1 procent av de som var antagna i BAS inte antas i SIM. Den minsta förändringen sker bland de antagna till förskolläraryrket och ämnesläraryrket där motsvarande siffror är 0,4 respektive 0,8 procent.

Tabell 8. Antal antagna höstterminen 2024 i BAS samt antalet i SIM som inte antas till samma program (examen), uppdelat efter program, ålder och kön.

Examen	Ålder	BAS (antal)		Ej samma examen i SIM			
		Kvinnor	Män	Kvinnor		Män	
				Ant.	%	Ant.	%
Civilingenjör	20 år och yngre	2 653	5 253	22	0,8	74	1,4
	21–24 år	1 511	2 286	27	1,8	63	2,8
	25–29 år	141	357	1	0,7	8	2,2
	30–39 år	50	154	0	0,0	8	5,2
	40 år och äldre	12	42	0	0,0	1	2,4
	Totalt	4 367	8 092	50	1,1	154	1,9
Förskollärare	20 år och yngre	583	31	5	0,9	0	0,0
	21–24 år	887	59	2	0,2	1	1,7
	25–29 år	566	38	0	0,0	0	0,0
	30–39 år	888	52	4	0,5	0	0,0
	40 år och äldre	330	20	1	0,3	0	0,0
	Totalt	3 254	200	12	0,4	1	0,5
Jurist/ Juris kandidat	20 år och yngre	817	331	15	1,8	13	3,9
	21–24 år	384	213	19	4,9	20	9,4
	25–29 år	67	60	3	4,5	1	1,7
	30–39 år	63	63	0	0,0	5	7,9
	40 år och äldre	25	27	1	4,0	0	0,0
	Totalt	1 356	694	38	2,8	39	5,6
Läkare	20 år och yngre	434	265	11	2,5	8	3,0
	21–24 år	235	166	10	4,3	9	5,4
	25–29 år	56	74	5	8,9	2	2,7
	30–39 år	54	65	2	3,7	3	4,6
	40 år och äldre	17	27	3	17,6	4	14,8
	Totalt	796	597	31	3,9	26	4,4

Med ett enda undantag (sjuksköterskeutbildningar) är det män som i större utsträckning påverkas och förlorar sin antagning. Till tandläkarutbildningen påverkas drygt var 10e man och det är främst män i ålder 21–24 år som förlorar en plats.

Tabell 8 (forts.). Antal antagna höstterminen 2024 i BAS samt antalet i SIM som inte antas till samma program (examen), uppdelat efter program, ålder och kön.

Examen	Ålder	BAS (antal)		Ej samma examen i SIM			
		Kvinnor	Män	Kvinnor		Män	
				Ant.	%	Ant.	%
Sjuksköterska	20 år och yngre	1 122	136	20	1,8	2	1,5
	21–24 år	1 197	206	32	2,7	4	1,9
	25–29 år	534	125	27	5,1	3	2,4
	30–39 år	731	178	18	2,5	6	3,4
	40 år och äldre	457	118	9	2,0	2	1,7
	Totalt	4 041	763	106	2,6	17	2,2
Socionom	20 år och yngre	924	90	21	2,3	0	0,0
	21–24 år	774	130	32	4,1	8	6,2
	25–29 år	305	95	15	4,9	4	4,2
	30–39 år	332	101	15	4,5	4	4,0
	40 år och äldre	170	45	2	1,2	0	0,0
	Totalt	2 505	461	85	3,4	16	3,5
Tandläkare	20 år och yngre	104	63	1	1,0	4	6,3
	21–24 år	51	34	1	2,0	7	20,6
	25–29 år	23	11	3	13,0	2	18,2
	30–39 år	12	4	1	8,3	0	0,0
	40 år och äldre	4	2	0	0,0	0	0,0
	Totalt	194	114	6	3,1	13	11,4
Ämneslärare (avancerad nivå)	20 år och yngre	1 132	602	14	1,2	9	1,5
	21–24 år	1 057	799	10	0,9	6	0,8
	25–29 år	477	436	2	0,4	2	0,5
	30–39 år	542	394	2	0,4	2	0,5
	40 år och äldre	502	310	0	0,0	2	0,6
	Totalt	3 710	2 541	28	0,8	21	0,8

För den enskilde så får en förändrad HP-poäng i vissa fall betydelse eftersom de i SIM förlorar den antagning de hade i BAS. Det innebär inte att dessa nödvändigtvis ser sig särskilt ”drabbade” eftersom de istället kan ha blivit antagna till en utbildning de rangordnat högre. Det är ett inte helt osannolikt scenario för sökande till t.ex. socionomutbildningen, men mindre troligt bland sökande till läkarutbildningen eftersom sökande till läkare tenderar att rangordna det utbildningsvalet högst (se Tabell 1). Tre av fyra sökande till

läkarutbildningen har den utbildningen som högst rankad vilket kan jämföras med sökande till socionomutbildningen där endast 4 av 10 rangordnar det utbildningsvalet högst.

Sammanfattning av simuleringen

En förändrad modell tenderar att inte påverka den enskilde provtagarens möjlighet att antas till en utbildning, en antagen i BAS förblir i mycket hög grad antagen till samma utbildning i SIM. Att ändra modell för beräkning av HP-poäng innebär därför inga stora förändringar i utbildningarnas sammansättning med avseende på ålder och kön hösten 2024.

Diskussion

Syftet med denna studie var att undersöka effekterna av att använda IRT-baserade HP-poäng på utfallet av antagningen till högskolan. Resultaten visar att HP-poängen kan förändras avsevärt med IRT-baserad poängsättning, samtidigt som effekterna på antagningen är mycket små.

När det gäller de specifika frågeställningarna kan vi konstatera att det är en mycket liten andel av de sökande vars möjlighet att antas till en utbildning påverkas. Detta gäller både de som i den riktiga antagningen (dvs. BAS) blev antagna till en utbildning men som i den simulerade antagningen (dvs. SIM) inte blir antagna till någon utbildning alls, och de som i BAS inte blev antagna till någon utbildning men som i SIM blir antagna till en utbildning.

När det sedan gäller skillnader mellan olika program så är andelen som inte antas till samma utbildning i SIM som i BAS generellt sett relativt liten, men om andelarna för de olika programmen jämförs så är den största andelen (6,2 procent, tandläkarprogrammet) sexton gånger så stor som den minsta andelen (0,4 procent, förskolläraryrket). Slutsatsen blir således att det finns skillnader i effekter mellan olika program.

Gällande den tredje frågeställningen, om gruppen antagna förändras med avseende på kön och ålder, blir slutsatsen att gruppen antagna förändras mycket lite överlag, och därmed även när det gäller kön och ålder. Betraktar vi olika program/examina kan vi däremot konstatera att det kan bli relativt stora procentuella förändringar i antalet antagna för enskilda grupper baserade på ålder och kön. Det gäller framförallt läkarprogrammet och tandläkarprogrammet, som visserligen är relativt små jämfört med de andra redovisade programmen och därför har små förändringar i absoluta tal. Detta innebär emellertid inte att dessa förändringar nödvändigtvis är obetydliga; det kan hända att dessa program värdesätter varje antagen som bidrar till en mångfald i studentpopulationen.

Metod- och datafrågor

I studien så görs ingen regelrätt ekvivalering av IRT-poängen, utan poängen skalas så att fördelningen matchar den normerade poängfördelningen vid varje provtillfälle. Alternativet hade varit att ekvivalera varje provversion på nytt med utgångspunkt i exempelvis den senaste provversionen, men erfarenheter från försök i normeringen av högskoleprovet har visat att IRT-baserade metoder tenderar att ge resultat som skiljer sig tydligt från CTT-baserade metoder. En fördel med ett sådant tillvägagångssätt skulle vara att det hade varit möjligt att se effekten av IRT-modellering i hela poängsättningsprocessen, men det hade samtidigt ställt andra krav på datainsamlingsdesignen.

Arbetet med studien har också visat på utvecklingsbehov när det gäller högskoleprovsregistret. Registret består av flera olika databaser/tabeller och det är byggt med fokus på provtagarnas övergripande resultat kombinerat med deras personuppgifter (namn, personnummer, svarshäftesnummer). Det är däremot inte anpassat för att på ett enkelt sätt erhålla data på uppgiftsnivå som är rättade (dvs. innehåller information om huruvida provtagaren svarade rätt eller fel på respektive uppgift) och som dessutom innehåller de personuppgifter som är nödvändiga för antagningen. Det är viktigt att ha en datainfrastruktur som är anpassad till och har en beredskap för undersökningar som denna, och det behöver därför diskuteras hur och när en sådan anpassning kan göras.

Begränsningar och vidare undersökningar

En begränsning med studien är att den (av nödvändighet) utgår från att de utbildningsval som de sökande gör inte påverkas av förändrade HP-poäng. Det är emellertid troligt att den utgångspunkten inte skulle hålla i en verklig situation där de sökande har andra HP-poäng, oavsett om det är på grund av IRT-baserad poängsättning eller andra skäl. Tydligt är att såväl antalet sökta utbildningar som de sökandes rangordning av sökta utbildningar kan påverka vem som i slutändan antas till en utbildning, om än på marginalen. Det är inte orimligt att anta att en sökande som ligger precis på gränsen till att antas till en utbildning (utifrån de meriter som krävts historiskt sett) skulle göra andra val om hens meriter genom en annan poängberäkningsmodell försämrades flera steg på skalan. Och omvänt; en sökande som ligger ganska långt steg från att antas till en utbildning av intresse skulle kanske göra andra val om hens meriter låg närmare det som tidigare krävts för att bli antagen. Det är dock svårt att sia om vilka effekter detta skulle få på antagningen, men utifrån att effekterna på resultatfördelningen av att använda IRT vid beräkningen av resultat är relativt begränsad så torde påverkan på ansökningarna följaktligen också vara relativt begränsad. Och det på övergripande nivå, för enskilda individer kan såklart påverkan vara större.

I rapportens metoddelen nämns skälen för varför 3PL-modellen valdes till beräkningen av de IRT-baserade HP-poängen. Här kan det vara värt att påpeka två saker. För det första att det hade varit tämligen meningslöst i det här sammanhanget att studera 1PL-modellen, eftersom 1PL-baserade poäng och råpoäng har exakt samma rangordning (givet att det inte förekommer saknade data, vilket inte är fallet med högskoleprovet). Det skulle emellertid finnas ett värde i att undersöka 2PL-modellen, men det valdes alltså bort i den här studien av tids- och resursskäl. För det andra så kan noteras att även om IRT och specifikt 3PL-modellen används i analyser av uppgifternas kvalitet så används inte parametervärdena explicit för beslut om en uppgift ska tas med i ett reguljärt prov eller inte. Om så däremot skulle göras, exempelvis att uppgifter med ett c -värde över ett visst tröskelvärde skulle diskvalificeras, så skulle det kunna påverka förhållandet mellan IRT- och CTT-baserade poäng. Ju lägre tröskelvärde desto starkare samband mellan de två poängen, vilket skulle bli särskilt tydligt i den lägre delen på poängskalan. Vilket i sin tur torde ge mindre effekter på antagningen.

I den här studien kan det simulerade tillståndet sägas visa hur antagningen ser ut när *alla* giltiga HP-resultat beräknats utifrån en IRT-baserad poängsättning. Detta skulle då representera en tidpunkt minst åtta år efter införandet av en sådan poängsättning, om man utgår från den nuvarande giltighetstiden på åtta år och att HP-resultaten inte ändras retroaktivt. Det kan även vara av intresse att simulera ett tillstånd som visar hur antagningen ser ut vid införandet av en ny poängsättningsmodell, vilket skulle kunna göras genom att endast ändra HP-poängen från den senaste provomgången. Utfallet från en sådan studie skulle kanske inte visa några påtagliga effekter av poängsättningen, men de skulle troligen vara större än i den här studien. Effekterna i en verklig situation beror dock även på hur stor andel av de giltiga resultaten som kommer från den senaste provomgången. Med en förändrad poängsättning så följer troligen en osäkerhet kring värdet och konkurrenskraften i redan befintliga resultat. Om poängsättningen skulle ändras så är det därför rimligt att förvänta sig fler provtagare, vilket också skulle kunna påverka effekten av poängsättningen.

Referenser

- Chalmers, R. P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, 48(6), 1-29. <https://doi.org/10.18637/jss.v048.i06>.
- Dahlke, J. A., Sackett, P. R., & Kuncel, N. R. (2019). Effects of range restriction and criterion contamination on differential validity of the SAT by race/ethnicity and sex. *Journal of Applied Psychology*, 104(6), 814–831. <https://doi.org/10.1037/apl0000382>
- Fan, X. (1998). Item Response Theory and Classical Test Theory: An Empirical Comparison of their Item/Person Statistics. *Educational and Psychological Measurement*, 58(3), 357-381. <https://doi.org/10.1177/0013164498058003001>
- IBM Corp. (2023). IBM SPSS Statistics for Windows (Version 29.0) [Computer software]. Armonk, NY: IBM Corp.
- Lyrén, P.-E. (2008). Prediction of Academic Performance by Means of the Swedish Scholastic Assessment Test. *Scandinavian Journal of Educational Research*, 52(6), 565–581. <https://doi.org/10.1080/00313830802497158>
- Macdonald, P., & Paunonen, S. V. (2002). A Monte Carlo Comparison of Item and Person Statistics Based on Item Response Theory versus Classical Test Theory. *Educational and Psychological Measurement*, 62(6), 921-943. <https://doi.org/10.1177/0013164402238082>
- R Core Team (2023). R: A language and environment for statistical computing. Wien, Österrike: R Foundation for Statistical Computing. <https://www.R-project.org/>
- Wikström, C., Laukityte, I., & Lyrén, P.-E. (2024). *Högskoleprovet våren och hösten 2023. Provtagargruppens sammansättning och resultat* (BVM 73:2024). Umeå: Umeå universitet, Institutionen för tillämpad utbildningsvetenskap, Beteendevetenskapliga mätningar
- Zenisky, A., Hambleton, R.K., Luecht, R.M. (2009). Multistage Testing: Issues, Designs, and Research. In: van der Linden, W., Glas, C. (eds.) *Elements of Adaptive Testing*. Statistics for Social and Behavioral Sciences. Springer, New York, NY. https://doi.org/10.1007/978-0-387-85461-8_18

RAPPORTER FRÅN INSTITUTIONEN FÖR BETEENDEVETENSKAPLIGA MÄTNINGAR

Utgivna rapporter i föregående serie:

<http://www.umu.se/edmeas/publikationer/index.html>

2004

- BVM nr 1. LÄRARENKÄT OM DE NATIONELLA PROVEN I MATEMATIK. Kurs B, C och D hösten 2003. Maria Ericsson, Björn Sigurdsson
- BVM nr 2. ORDFÖRSTÅELSE. En litteraturstudie med anknytning till högskoleprovets ORD-prov. Sandra Scott
- BVM nr 3. SJÄLVVÄRDERING SOM METOD FÖR ATT MÄTA MÅLUPPFYLLELSE VIA PROV. Anna Sundström
- BVM nr 4. ATT MÄTA SKRIVFÖRMÅGA. En forskningspresentation om provformat, reliabilitet, validitet samt sociala aspekter. Marit Sigurdson
- BVM nr 5. KLASSISK OCH MODERN TESTTEORI. Analys av det teoretiska och det praktiska körkortsprovet. Marie Wiberg
- BVM nr 6. UTBYTESKOMPLETTERINGAR BLAND DEM SOM AVSLUTADE GYMNASIET 1997–2001. Differenser mellan avgångsbetyg från gymnasiet och betyg som har kompletterats efter den ordinarie gymnasieskolan. Kent Löfgren
- BVM nr 7. HÖGSKOLEPROVET VÅREN OCH HÖSTEN 2004. Provdeltagargruppens sammansättning och resultat. Christina Stage, Gunilla Ögren
- BVM nr 8. HÖGSKOLEPROVET OCH DE MÅLRELATERADE BETYGEN. En studie av de första eleverna med de nya gymnasiebetygen. Anders Lexelius

2005

- BVM nr 9. LÄRARES SKATTNINGAR AV SINA ELEVERS PROVRESULTAT. Gunilla Näsström
- BVM nr 10. LÄRARENKÄT OM DE NATIONELLA PROVEN I MATEMATIK. Kurs B, C och D våren 2004. Maria Ericsson, Björn Sigurdsson
- BVM nr 11. SOCIALGRUPPSSKILLNADER I RESULTAT PÅ HÖGSKOLEPROVET. Christina Stage
- BVM nr 12. ÄR DET SVÅRARE ATT DELA MED FYRA ÄN MED TVÅ NÄR MAN LÄSER MATTE C? En jämförelse av svårighetsgrad mellan olika versioner av matematikuppgifter i Nationella kursprov. Ewa Bergqvist, Anna Lind

- BVM nr 13. DEN SVENSKA FÖRARPRÖVNINGENS RESULTAT. Sambandet mellan kunskapsprovet och körprovet för underkända och godkända provtagare. Anna Sundström, Marie Wiberg
- BVM nr 14. DATORBASERADE PROV – egenskaper, möjligheter och begränsningar. Christina Wikström
- BVM nr 15. PRESTATIONSSKILLNADER MELLAN FLICKOR OCH POJKAR I NO. En studie av uppgiftsformatets betydelse i TIMSS 2003. Niklas Eriksson
- BVM nr 16. FLICKOR, POJKAR, FYSIK OCH MATEMATIK. Skillnader i inställning mellan hög- och lågpresterande i TIMSS 1995. Lena Adolfsson
- BVM nr 17. HÖGSKOLEPROVET VÅREN OCH HÖSTEN 2005. Provdeltagargruppens sammansättning och resultat. Christina Stage, Gunilla Ögren

2006

- BVM nr 18. MODELL FÖR BESKRIVNING AV KURSPLANEN FÖR DEN SVENSKA FÖRARUTBILDNINGEN: EN LITTERATURSTUDIE. Tova Stenlund
- BVM nr 19. VIKTNING AV DELPROVEN I HÖGSKOLEPROVET. Christina Stage, Mats Hamrén, Christina Jonsson
- BVM nr 20. TIMSS FIXPUNKTER. En analys av vad elever med olika resultat i TIMSS 2003 vet och kan göra. Peter Nyström
- BVM nr 21. MODELLPRÖVNING. Empirisk prövning av teoretiska modeller för beskrivning av kursplan för förarutbildning. Tova Stenlund
- BVM nr 22. UTVECKLING AV INSTRUMENT FÖR ATT MÄTA BEGREPPET UPPLEVD FÖRARKOMPETENS. Anna Sundström
- BVM nr 23. HUR HÖGSKOLEINSTITUTIONER OCH ENHETER KAN GRANSKA SINA KURSER UR JÄMSTÄLLDHETS- OCH GENUS-PERSPEKTIV. Nuläget och framåtblick med praktiska förslag. Kent Löfgren
- BVM nr 24. JÄMFÖRELSE AV INNEHÅLL I DEN GAMLA OCH DEN NYA KURSPLANEN FÖR DEN SVENSKA FÖRARUTBILDNINGEN. Tova Stenlund, Widar Henriksson, Anna Sundström
- BVM nr 25. HÖGSKOLEPROVET VÅREN OCH HÖSTEN 2006. Provdeltagargruppens sammansättning och resultat. Christina Stage, Gunilla Ögren

2007

- BVM nr 26. UTVÄRDERING AV INSTRUMENT FÖR UPPLEVD TEORETISK OCH PRAKTISK KOMPETENS. Ett försök med en ny förarprovsmodell. Anna Sundström
- BVM nr 27. BEGREPPET KOMPLETTERING I ETT EXAMINATIONSPERSPEKTIV. Ett försök med det praktiska körkortsprovet. Tova Stenlund, Widar Henriksson, Charlotte Wahl, Bengt Holmberg
- BVM nr 28. NY KURSPLAN OCH NYA FORMER FÖR EXAMINATION. Jämförelse av förarprovet för körkort B före och efter förändringen. Marie Wiberg
- BVM nr 29. SEKVENTIELLA EFFEKTER I HÖGSKOLEPROVET AVSEENDE DELPROVET NOG. Anders Lexelius
- BVM nr 30. SAMSTÄMMIGHET MELLAN MÅL OCH PROV. En studie av det svenska förarutbildningssystemet. Tova Stenlund, Widar Henriksson, Anna Sundström
- BVM nr 31. HÖGSKOLEPROVET VÅREN OCH HÖSTEN 2007. Provdeltagargruppens sammansättning och resultat. Christina Stage, Gunilla Ögren
- BVM nr 32. SVENSKA SKOLANS LÄRARE OCH UNDERVISNING I MATEMATIK OCH NO I ETT INTERNATIONELLT PERSPEKTIV. Några resultat från TIMSS 2003. Susanne Alger

2008

- BVM nr 33. UTVECKLING OCH UTVÄRDERING AV INSTRUMENT INFÖR EN STUDIE AV LIKVÄRDIGHETEN I KÖRPROVET. Susanne Alger, Widar Henriksson, Anna Sundström
- BVM nr 34. HÖGSKOLEPROVET VÅREN OCH HÖSTEN 2008. Provdeltagargruppens sammansättning och resultat. Christina Stage, Gunilla Ögren
- BVM nr 35. URVALSPROV UR ETT SVENSKT OCH INTERNATIONELLT PERSPEKTIV. Christina Wikström

2009

- BVM nr 36. RAMVERK FÖR NATIONELLA ÄMNESPROV I BIOLOGI, FYSIK OCH KEMI I GRUNDSKOLANS ÅRSKURS 9. Peter Nyström, Maria Åström
- BVM nr 37. KUNSKAP, INTRESSE OCH STUDIEFRAMGÅNG. En litteraturgranskning. Ingela Eriksson, Widar Henriksson

- BVM nr 38. ELEVERS SYN PÅ NATIONELLA PROV. Resultat från elevenkät vid genomförandet av nationella ämnesprov i biologi, fysik och kemi i grundskolans årskurs 9 våren 2009. Maria Åström, Hanna Eklöf
- BVM nr 39. LIKVÄRDIGHETEN I KÖRPROVET. En studie av samstämmighet i bedömningen. Susanne Alger, Widar Henriksson, Anna Sundström
- BVM nr 40. HÖGSKOLEPROVET VÅREN OCH HÖSTEN 2009. Provdeltagargruppens sammansättning och resultat. Christina Stage, Gunilla Ögren

2010

- BVM nr 41. SAMMANHÅLLET PROV. Utvärdering av effekter av ett sammanhållet teoretiskt och praktiskt förarprov. Susanne Alger, Widar Henriksson, Anna Wänglund
- BVM nr 42. ETT NYTT HÖGSKOLEPROV. Bakgrund och konsekvenser. Christina Stage, Gunilla Ögren
- BVM nr 43. HÖGSKOLEPROVET VÅREN OCH HÖSTEN 2010. Provdeltagargruppens sammansättning och resultat. Christina Stage, Gunilla Ögren

2011

- BVM nr 44. ELEVERS SYN PÅ NATIONELLA PROV. Resultat från elevenkät vid genomförandet av nationella ämnesprov i biologi, fysik och kemi i grundskolans årskurs 9 våren 2010. Maria Åström, Hanna Eklöf
- BVM nr 45. PROVTAGARES UPPFATTNING OM FÖRARPROVET. Enkätundersökning före och efter införandet av sammanhållet förarprov. Susanne Alger, Anna Sundström
- BVM nr 46. ATT MÄTA ALLMÄNHETENS KUNSKAPER OM TRAFIKSÄKERHET. Utveckling och utprovning av instrument. Anna Sundström, Per-Erik Lyrén, Susanne Alger
- BVM nr 47. SAMSTÄMMIGHET MELLAN KURSPLAN OCH PROV. En studie av kunskapsprovet för körkortsbehörighet AM. Gunilla Näsström, Susanne Alger, Anna Sundström
- BVM nr 48. UTVÄRDERING AV KUNSKAPSPROVET FÖR KÖRKORTS-BEHÖRIGHET AM. Gunilla Näsström, Susanne Alger, Anna Sundström
- BVM nr 49. HÖGSKOLEPROVET VÅREN 2010. Provdeltagargruppens sammansättning och resultat. Christina Stage, Gunilla Ögren

BVM nr 50. BESKRIVNING AV DATA OM DET PRAKTISKA FÖRARPROVET 1998–2005 OCH ANALYS AV TILLFÖRLITLIGHETEN. Susanne Alger, Anna Sundström

BVM nr 51. BESKRIVNING AV DATA OM DET PRAKTISKA FÖRARPROVET 2007–2008 OCH ANALYS AV TILLFÖRLITLIGHETEN. Susanne Alger, Anna Sundström

2012

BVM nr 52. DET NYA HÖGSKOLEPROVET. Provdeltagargruppens sammansättning och resultat hösten 2011. Christina Stage, Gunilla Ögren

BVM nr 53. PROV FÖR FÖRARBEVIS FÖR MOPED KLASS II. Susanne Alger, Widar Henriksson, Hanna Eklöf

BVM nr 54. PROV FÖR FÖRARBEVIS FÖR MOPED KLASS II. Enkätstudie avseende provkonstruktion och provgenomförande. Susanne Alger

BVM nr 55. SÄMRE ÅR FÖR ÅR? En studie av förändringen av andelen godkända på körprov för körkortsbehörighet B. Susanne Alger, Hanna Eklöf

2013

BVM nr 56. HÖGSKOLEPROVET VÅREN OCH HÖSTEN 2012. Provdeltagargruppens sammansättning och resultat. Christina Stage, Gunilla Ögren

BVM nr 57. DET PRAKTISKA KÖRPROVET 2009–2011. Analys av datas tillförlitlighet samt provresultat och trender. Susanne Alger, Hanna Eklöf

BVM nr 58. NYA KURS PROV I BIOLOGI OCH KEMI FÖR GYMNASIESKOLAN. Bakgrund och resultat 2011. Gunnel Grelsson, Göran Bergqvist, Niklas Jacobsson

BVM nr 59. HÖGSKOLEPROVET VÅREN OCH HÖSTEN 2013. Provdeltagargruppens sammansättning och resultat. Gunilla Ögren

2014

BVM nr 60. UPPFÖLJNING AV SAMMANHÅLLET FÖRARPROV. Utvärdering av effekter av ett sammanhållet teoretiskt och praktiskt förarprov mellan 2007 och 2013. Susanne Alger, Hanna Eklöf

BVM nr 61. HÖGSKOLEPROVET VÅREN OCH HÖSTEN 2014. Provdeltagargruppens sammansättning och resultat. Gunilla Ögren

2016

- BVM nr 62. TRENDER 1998–2015 AVSEENDE FÖRÄNDRINGEN AV ANDELEN GODKÄNDA PÅ KÖRPROV FÖR KÖRKORTSBEHÖRIGHET B. Susanne Alger, Hanna Eklöf
- BVM nr 63. HÖGSKOLEPROVET VÅREN OCH HÖSTEN 2015. Provtagargruppens sammansättning och resultat. Per-Erik Lyrén
- BVM nr 64. JÄMFÖRELSE MELLAN PROVORTER. En studie av körprov för behörighet B. Susanne Alger, Hanna Eklöf

2017

- BVM nr 65. HÖGSKOLEPROVET VÅREN OCH HÖSTEN 2016. Provtagargruppens sammansättning och resultat. Per-Erik Lyrén
- BVM nr 66. SAMMA PROV – ANDRA PROVTAGARE. En studie av förändringar i körprovsresultat och provtagargruppens sammansättning. Susanne Alger, Hanna Eklöf

2018

- BVM nr 67. HÖGSKOLEPROVET VÅREN OCH HÖSTEN 2017. Provtagargruppens sammansättning och resultat. Per-Erik Lyrén

2019

- BVM nr 68. HÖGSKOLEPROVET VÅREN OCH HÖSTEN 2018. Provtagargruppens sammansättning och resultat. Per-Erik Lyrén, Inga Laukaityte

2021

- BVM nr 69. HÖGSKOLEPROVET VÅREN OCH HÖSTEN 2019. Provtagargruppens sammansättning och resultat. Christina Wikström, Inga Laukaityte

2022

- BVM nr 70. HÖGSKOLEPROVET HÖSTEN 2020. Provtagargruppens sammansättning och resultat. Christina Wikström, Inga Laukaityte

2023

- BVM nr 71. HÖGSKOLEPROVET VÅREN OCH HÖSTEN 2021. Provtagargruppens sammansättning och resultat. Christina Wikström, Inga Laukaityte, Per-Erik Lyrén

BVM nr 72. HÖGSKOLEPROVET VÅREN OCH HÖSTEN 2022. Provtagar-
gruppens sammansättning och resultat. Christina Wikström, Inga
Laukaityte, Per-Erik Lyrén

2024

BVM nr 73. HÖGSKOLEPROVET VÅREN OCH HÖSTEN 2023. Provtagar-
gruppens sammansättning och resultat. Christina Wikström, Inga
Laukaityte, Per-Erik Lyrén

2025

BVM nr 74. HÖGSKOLEPROVET VÅREN OCH HÖSTEN 2024. Provtagar-
gruppens sammansättning och resultat. Christina Wikström, Inga
Laukaityte, Per-Erik Lyrén