

**Normering, ekvivalering eller kalibrering av delar
av högskoleprovet**

Christina Stage

Abstract

The SweSAT (Swedish Scholastic Assessment Test) is used for selection to higher education in Sweden. The test consists of five sub-tests measuring different abilities. Test results are valid for five years, and therefore equating of different test versions is an important issue. The equating is made on the total test score by the equi-percentile-equating method. There has been a discussion of whether the predictive validity of the test could be improved by putting different weights on different parts of the test for different educations. This is only possible if results on different parts of the test can be equated separately. In the first part of this report a description is given of an attempt to equate the verbal subtests, and the quantitative subtests separately. The results of this attempt were not very promising, since the quantitative part contained too few items for a reliable equating. The second part of the report describes an attempt to calibrate the separate sub-tests from different test occasions. Even though this calibration did not constitute a proper equating, the results were regarded as being good enough for research purposes.

Introduktion

Högskoleprovet har utgjort urvalsinstrument till högre utbildning sedan våren 1977. Fram till år 1991 dock i begränsad utsträckning, eftersom det bara var sökande som var behöriga enligt den s.k.25:4 regeln som kunde använda det. Sedan hösten 1991 har dock provet varit ett alternativ till gymnasiebetygen för alla sökande till utbildningsplatser där urval måste göras, d.v.s. då antalet sökande överstiger antalet studieplatser.

Sedan våren 1996¹ består högskoleprovet av 122 flervalssuppgifter uppdelade på fem delprov:

- ORD som avser att mäta ordkunskap med 40 uppgifter
- NOG som avser att mäta logisk kvantitativ problemlösning med 22 uppgifter
- LÄS som avser att mäta svensk läsförståelse med fem texter och 20 uppgifter
- DTK som avser att mäta förmågan att tolka diagram, tabeller och kartor med tio figuruppsättningar och 20 uppgifter
- ELF som avser att mäta engelsk läsförståelse med texter av varierande längd och 20 uppgifter

Varje rätt svar ger en råpoäng och det totala antalet råpoäng transformeras till en normerad poäng på en skala från 0.0 till 2.0, där 2.0 är det högsta resultatet

Högskoleprovet är ett generellt prov, som ger ett relativt brett mått på olika aspekter av studieförutsättningar. Inom ramen för Tillträdesutredningen (1985) diskuterades frågan om ett differentiellt prov skulle fungera bättre som urvalsinstrument:

Ett prov liknande det nuvarande, men uppdelat på två delar, den ena inriktad på verbala färdigheter och den andra på matematiska/kvantitativa färdigheter. Delarna ges olika vikt beroende på vilken högskoleutbildning man söker till. (s.173)

¹ De förändringar som har gjorts av provet under årens lopp beskrivs i ”Högskoleprovets utveckling under åren 1977 – 2000” (Stage & Ögren, 2001).

Resultatet av denna diskussion blev dock:

För flertalet utbildningar anser vi att högskoleprovets nuvarande utformning ger en tillfredsställande grund för urvalet. Vi föreslår alltså att högskoleprovet även i fortsättningen ges en utformning som är avsedd att avspegla mera generella studieförutsättningar. (s. 174)

Diskussionen om en differentierad provpoäng har nyligen aktualiserats igen, framför allt efter resultaten från en studie av Svensson, Gustafsson & Reuterberg (2001). I denna studie framkom att sambanden mellan högskoleprovresultat och erhållna poäng, efter ett års studier inom civilingenjörsutbildningen, var 0.19 för de studenter som antogs höstterminen 1993 och 1994, 0.10 för dem som antogs höstterminen 1995 och 0.14 för dem som antogs 1996. Motsvarande samband mellan erhållna poäng efter ett års studier och högskoleprovets kvantitativa delprov (NOG och DTK) var 0.33, 0.28, 0.26 och 0.27 och högskoleprovets verbala delprov (ORD, LÄS och ELF) var 0.16, 0.15, 0.07 och 0.11. Motsvarande resultat för studerande inom juristutbildningen var för totala högskoleprovet 0.24, 0.27, 0.13 och 0.13, för de kvantitativa delproven 0.23, 0.33, 0.26 och 0.24 och för de verbala delproven 0.16, 0.32, 0.21 och 0.22. Dessa olika samband mellan studieresultat och de kvantitativa respektive verbala delproven var anledningen till att frågan om möjligheten att genomföra separata normeringar av två provblock åter aktualiserades.

Provkonstruktörerna har som målsättning att sammanställa parallella versioner av respektive delprov. Parallellitet har i detta sammanhang två aspekter: dels ska varje delprov anpassas till en modell för hur provet ska se ut med avseende på ämnesområde, innehållskategorier, uppgiftsformat, informationsmängd, informationsvärde och lösningsprocedur. Dels ska enskilda uppgifter och därmed hela delprovet ha definierade svårighetsnivåer, varje uppgift ska diskriminera på ett tillfredsställande sätt och ha balans avseende felaktiga svarsförslag. Detta kan kallas pre-ekvivalering av proven.

Efter provgenomförandet genomförs den egentliga ekvivaleringen av provpoängen. Den ekvivaleringsmetod som används vid normeringen av högskoleprovet är 'equivalent groups equipercentile equating' (se Braun & Holland, 1982 för diskussion av metoden). Syftet med ekvivaleringen är att finna den transformation av råpoäng till normerad poäng som ger en fördelning i den nya provdeltagarpopulationen som så väl som möjligt överensstämmer med fördelningen för tidigare populationer. Metoden förutsätter att vissa jämförbara undergrupper av

provdeltagare har samma resultatfördelning och de resultat som erhålls för dessa grupper definierar ekvivaleringsfunktionen. Det gäller att identifiera de poäng i det nya provet som har samma percentilrangordning som tidigare prov.

Den strategi som tillämpas för att fastställa poänggränserna för den normerade poängen utgörs av en sammanvägning av olika jämförelser. Först studeras den totala provdeltagargruppen avseende sammansättning (kön, ålder och utbildning) och resultat i jämförelse med tidigare provdeltagargrupper, men eftersom sammansättningen av provdeltagargruppen inte får påverka möjligheten att få ett viss normerat provresultat, studeras även utfallet i två referenspopulationer:

Referenspopulation I utväljs ur den totala provdeltagargruppen genom proportionellt stratifierat urval för att ge en jämförbar fördelning av kön, ålder och utbildning. Denna grupp är därmed likadan från provtillfälle till provtillfälle, avseende tre bakgrundsvariabler, som man vet påverkar provresultatet. Gruppen kan därmed antas ha samma förmåga från ett provtillfälle till ett annat. Detta innebär att eventuella skillnader i provutfall mellan olika provtillfällen för denna referenspopulation kan relateras till provet.

Referenspopulation II utgörs av de provdeltagare som vid provtillfället är 18 år och har pågående tre- eller fyraårig gymnasieutbildning. Det innebär att denna referenspopulation har samma ålder och lika lång bakgrundsutbildning vid olika provtillfällen.

Den strategi som tillämpas för normering är således att finna de normerade värden som bäst fördelar provdeltagarna (i samtliga tre grupper) på olika normerade poäng så att överensstämmelsen med tidigare fördelningar blir så god som möjligt. De resultat som kan betraktas som viktigast är de normerade poängerna mellan 1.0 och 2.0, eftersom det är provdeltagare med dessa resultat som konkurrerar om de attraktiva studieplatserna. Av den anledningen påbörjas anpassningen av fördelningarna uppifrån, d.v.s. från den normerade poängen 2.0. Därefter bestäms intervallen så att varje normerad poäng så väl som möjligt motsvarar de fördelningar, som har gällt för tidigare prov, för de två referenspopulationerna och totalpopulationen.

De slutliga gränserna för normerad poäng bestäms genom en sammanjämkning av gränserna för de tre studerade grupperna. För de tio högskoleprov, som ingår i denna studie, har gränserna för den normerade poängen 1.0 varierat från som lägst 68-71 råpoäng (98:B, 99:B) till som högst 77-80 råpoäng (96:A) och för den normerade

poängen 2.0 har kravet varierat från 109 råpoäng (00:B, 99:B, 99:A, 98:B) till 113 råpoäng (96:A).

DEL I

Syfte

Syftet med denna studie är att undersöka utfallet samt effekterna på gruppnivå av att genomföra normering av högskoleprovet uppdelat på ett kvantitativt block bestående av delproven NOG och DTK samt ett verbalt block bestående av delproven ORD, LÄS och ELF.

Metod

Material

Eftersom högskoleprovets sammansättning förändrades år 1996 valdes de tio prov som genomförts efter förändringen för denna blocknormering. Valet av tio prov bestämdes av det faktum att giltighetstiden för provet är fem år. De prov som valdes blev därmed proven fr.o.m. våren 1996 (96A) t.o.m. provet hösten 2000 (00B). Resultaten på dessa prov samt antalet provdeltagare framgår i tabell 1.

Tabell 1. Provresultatens medelvärden (M) och spridningar (s) fr.o.m. våren 1996 (96A) t.o.m. hösten 2000 (00B), samt antal provdeltagare vid varje provtillfälle (N).

	96A	96B	97A	97B	98A	98B	99A	99B	00A	00B
M	76.2	75.4	70.6	72.6	72.2	67.7	68.9	67.7	69.0	68.3
s	17.5	17.1	18.0	17.3	17.5	17.6	17.2	17.6	18.4	17.2
N	82264	56067	82673	60905	75352	55088	68476	49594	62496	38429

Som framgår av tabell 1 har provet successivt blivit svårare under den aktuella tidsperioden, något som varit avsiktligt eftersom de första versionerna efter omstruktureringen 1996 var något för lätta. Idealt bör medelvärdet ligga mellan 67 – 69 råpoäng.

Högskoleprovet är avsett att vara ett generellt prov, där samtliga delprov, ska mäta en speciell aspekt av studieförmåga. I tabell 2 visas de

fem delprovens genomsnittliga interkorrelationer för de tio aktuella proven.

Tabell 2. Delprovens interkorrelationer. I diagonalen (fetstil) anges delprovens reliabilitetskoefficienter (α) och i den nedre triangeln (kursivstil) anges korrelationskoefficienterna korrigerade för reliabilitetsbrister.

	ORD	NOG	LÄS	DTK	ELF
ORD	.84	.29	.59	.34	.58
NOG	.36	.76	.45	.65	.45
LÄS	.64	.62	.68	.47	.61
DTK	.44	.88	.67	.72	.46
ELF	.71	.58	.82	.60	.80

Som framgår av tabell 2 föreligger ett högt samband mellan delproven NOG och DTK. 77 procent av variansen i dessa två delprov är gemensam och dessa två delprov kan därför sägas mäta en gemensam förmåga, som har kallats kvantitativ. Delprovet ORD har, som framgår av tabellen, de högsta sambanden med delproven LÄS och ELF. 41 procent av variansen är gemensam med delprovet LÄS och 50 procent med delprovet ELF. Delproven LÄS och ELF har också tämligen höga inbördes samband, 67 procent av variansen är gemensam för dessa två delprov. Delproven LÄS och ELF har dock tydliga samband även med delproven NOG och DTK. Delprovet LÄS har 38 procent av variansen gemensam med delprovet NOG och 45 procent med delprovet DTK. Delprovet ELF har 34 procent gemensam varians med delprovet NOG och 36 procent med delprovet DTK. Trots det har i detta sammanhang delproven ORD, LÄS och ELF tillsammans fått utgöra ett block som har benämnts verbalt.

Genomförande

Normering

Eftersom ett högskoleprovsresultat är giltigt i fem år är det viktigt att resultaten från olika tillfällen är jämförbara. Normeringen av högskoleprovet har som främsta syfte att göra poängen jämförbar från ett provtillfälle till ett annat, d.v.s. det ska vara precis lika lätt eller svårt

att erhålla en viss normerad poäng oavsett vid vilket provtillfälle man deltagit.

I denna studie genomfördes normeringen av de kvantitativa och de verbala provblocken så att för summapoängen av NOG och DTK respektive summapoängen av ORD, LÄS och ELF genomfördes en lika percentil-ekvivalering, men endast för totalgruppen provdeltagare. I de fall där det förelåg tveksamheter avseende gränssättning, vilket framför allt gällde det kvantitativa blocket, gjordes en anpassning till den normering som tidigare gjorts för totalprovet.

Resultat

Det kvantitativa provblocket kom att bestå av 42 uppgifter och det verbala av 60 uppgifter. 42 uppgifter är ett tämligen lågt antal, vilket medför att det är många individer som har samma råpoäng. Detta i sin tur medför att den normerade poängen blir mer ojämnt fördelad än vad som gäller för det verbala provblocket och än mer än vad som gäller för totalprovet.

I tabell 3 visas resultat för totalgrupperna provdeltagare vid de tio provtillfällena.

Tabell 3. Medelvärden och spridningar samt korrelationer med totalpoängen på respektive prov för det kvantitativa (rad 1-3) respektive det verbala (rad 4-6) provblocket.

	96A	96B	97A	97B	98A	98B	99A	99B	00A	00B
M_k	25.5	25.4	23.2	24.4	24.1	22.2	21.6	21.8	23.4	22.4
s_k	7.8	7.1	7.6	7.3	7.5	7.6	7.0	7.0	7.5	6.9
r_{tk}	.80	.78	.80	.76	.78	.79	.76	.78	.77	.76
M_v	50.7	49.9	47.4	48.2	48.2	45.6	47.3	45.9	45.6	45.9
s_v	12.2	12.4	12.8	12.5	12.5	12.5	12.7	12.9	13.4	12.8
r_{tv}	.92	.93	.93	.93	.93	.93	.93	.94	.93	.94

Som framgår av tabell 3 återspeglas skillnaderna i svårighetsnivå mellan de olika provtillfällena (som framgick i tabell 1) tämligen jämnt även på de två blocken av delprov. Sambanden mellan totalpoängen och det verbala blocket är tämligen höga, vilket är ett förväntat utfall med tanke på att det verbala blocket utgör 66 procent av provet.

Reliabiliteten på totalprovet har för samtliga aktuella versioner varit $r = .92$, reliabiliteten på det verbala blocket har varierat mellan $r = .90$ och $r = .92$ och på det kvantitativa blocket mellan $r = .82$ och $r = .87$.

I tabell 4 visas resultaten för män respektive kvinnor på de två provblocken.

Tabell 4. Resultatmedelvärden för män respektive kvinnor på de två provblocken och totalt för proven 96A – 00B samt skillnaden (d) i standardavvikelseenheter.

Prov	Tot			Verb			Kvan		
	M	Kv	d	M	Kv	d	M	Kv	d
96A	79.6	73.2	.36	51.4	50.1	.11	28.2	23.2	.64
96B	79.0	72.5	.38	51.0	49.1	.15	28.1	23.4	.66
97A	74.2	67.6	.37	48.3	46.6	.13	25.9	21.0	.78
97B	75.6	70.2	.31	48.8	47.8	.08	26.8	22.4	.60
98A	75.3	69.7	.32	48.5	47.9	.05	26.8	21.8	.67
98B	70.8	65.2	.32	46.0	45.2	.06	24.9	20.0	.64
99A	72.1	66.3	.34	48.0	46.8	.09	24.1	19.5	.66
99B	70.8	65.2	.32	46.8	45.2	.12	24.0	20.0	.57
00A	72.0	66.4	.30	46.1	45.2	.07	25.9	21.2	.63
00B	70.9	66.3	.27	46.5	45.4	.08	24.4	20.9	.51

Som framgår av tabell 4 blev utfallet, inte oväntat, att skillnaden mellan män och kvinnor är stor på det kvantitativa blocket, medan den är försumbar på det verbala blocket. På totalprovet varierar effektstorleken mellan 0.27 och 0.38 med genomsnittet 0.33 och typvärdet 0.32. På det verbala blocket är variationen mellan 0.05 och 0.15 med genomsnittet 0.09 och typvärdet 0.08. På det kvantitativa blocket varierar skillnaden mellan .51 och .78 standardavvikelseenheter med genomsnittet 0.64.

I tabell 5 visas hur stor andel av männen respektive av kvinnorna som fått den normerade poängen 2.0 vid respektive provtillfälle. I tabell 5 framgår också hur stor andel av männen respektive kvinnorna som skulle ha fått den normerade poängen 2.0 på det verbala respektive det kvantitativa blocket om normeringen hade genomförts separat för de två blocken.

Tabell 5. Andel av männen respektive kvinnorna som erhöll den normerade poängen 2.0 vid de tio aktuella provtillfällena. Samt andel män respektive kvinnor som skulle ha fått den normerade poängen 2.0 vid separat normering av det verbala och kvantitativa blocket.

Prov	Totalt			Verbalt			Kvantitativt		
	M	Kv	Tot	M	Kv	Tot	M	Kv	Tot
96A	0.9	0.4	0.6	0.7	0.7	0.7	1.3	0.2	0.7
96B	1.0	0.3	0.6	0.4	0.4	0.6	1.6	0.2	0.8
97A	0.8	0.3	0.5	0.4	0.4	0.4	1.4	0.2	0.7
97B	0.9	0.3	0.6	0.6	0.5	0.5	1.6	0.3	0.9
98A	0.7	0.4	0.5	0.4	0.5	0.5	0.5	0.1	0.3*
98B	0.9	0.4	0.6	0.6	0.6	0.6	1.0	0.2	0.6
99A	1.0	0.3	0.6	0.7	0.7	0.7	1.7	0.2	0.6
99B	0.8	0.3	0.6	0.6	0.5	0.6	0.8	0.2	0.5
00A	1.2	0.5	0.8	0.6	0.6	0.6	1.7	0.2	0.9**
00B	0.9	0.4	0.6	0.6	0.6	0.6	1.1	0.2	0.6

* Denna andel skulle ha blivit 0.8 om gränsen hade sänkts en råpoäng, här gjordes en anpassning till totalnormeringen. Jämför sid 8.

** Denna andel skulle ha blivit 0.4 om gränsen hade höjts en råpoäng, här gjordes en anpassning till totalnormeringen. Jämför sid 8.

Av resultaten i tabell 5 framgår att vid separat normering av det verbala blocket är det ungefär lika stora andelar av männen och av kvinnorna som får det högsta resultatet 2.0. På det kvantitativa blocket däremot är det en avsevärd skillnad mellan hur stor andel av männen respektive kvinnorna som får 2.0. (För beskrivning av sammansättningen av grupperna män och kvinnor se t.ex. Stage, & Ögren, 2001)

I tabell 6 visas hur stor andel av provdeltagarna vid varje provtillfälle som var 20 år eller yngre. I tabell 6 framgår även hur stor andel av de provdeltagare som var 20 år eller yngre, som fick den normerade poängen 2.0 vid varje provtillfälle. Slutligen hur stor andel av denna åldersgrupp som skulle ha fått den normerade poängen 2.0 vid separat normering av ett verbalt och ett kvantitativt block.

Tabell 6. Andel av provdeltagarna med normerad poäng 2.0 på totalprovet och det verbala respektive det kvantitativa blocket som var 20 år eller yngre.

Prov	% _≤ 20 år	Provet totalt	Verbalt	Kvantitativt
96:A	44	0.5	0.4	0.8
96:B	52	0.4	0.2	1.0
97:A	44	0.4	0.2	0.9
97:B	54	0.4	0.2	1.0
98:A	45	0.3	0.1	0.4
98:B	51	0.3	0.1	0.7
99:A	43	0.4	0.2	0.8
99:B	49	0.3	0.1	0.6
00:A	42	0.4	0.1	1.0
00:B	51	0.3	0.1	0.8

Av tabell 6 framgår att andelen unga provdeltagare har varierat från som lägst 42 procent av samtliga provdeltagare till som högst 52 procent. Andelen ungdomar bland provdeltagarna med resultatet 2.0 på totalprovet har uppvisat en sjunkande trend, från drygt 30 procent till och med hösten 1997 har andelen sjunkit till c:a 20 procent vid sekelskiftet. Andelen ungdomar bland provdeltagare med resultatet 2.0 har som lägst utgjort 0.3 procent och som högst 0.5 procent. På det verbala blocket utgjorde ungdomsgruppen 0.4 procent våren 1996, men har sedan dess varierat mellan 0.1 och 0.2 procent av provdeltagarna med resultatet 2.0. Inte heller på det kvantitativa blocket har variationen varit särskilt stor, som lägst har ungdomsgruppen utgjort 0.4 procent vid ett provtillfälle i övrigt har de varierat mellan 0.8 och en procent. Sammantaget gäller att andelen ungdomar med resultatet 2.0 skulle bli högre på det kvantitativa blocket, men lägre på det verbala blocket om normeringen gjordes separat.

Diskussion

Utfallet av detta försök till normering på blocknivå gav inget stöd för separat normering av delar av högskoleprovet. En viktig slutsats är att de 42 uppgifter som, med nuvarande utformning av provet, utgör det kvantitativa blocket är alltför få för normering med lika percentil-

metoden. Det visar sig i de ojämna steg som erhålls i normerad poäng, som kan ses i den sista kolumnen i tabell 5.

Antalet uppgifter i det kvantitativa blocket är även för lågt för att kunna ge en acceptabel reliabilitet.

Utfallet på gruppnivå visar att män skulle ha en fördel vid separat normering av det kvantitativa blocket. Andelen av männen som skulle få resultatet 2.0 skulle vara mellan fyra och åtta gånger så stor som andelen av kvinnorna som skulle få samma resultat. Vid normering av totalprovet har det som mest varit tre gånger så många av männen som av kvinnorna som fått den högsta poängen. Vid separat normering av det verbala blocket skulle å andra sidan andelen män och kvinnor med resultatet 2.0 vara ungefär desamma.

Ungdomsgruppen, d.v.s. provdeltagare som är 20 år eller yngre, skulle öka något på den högsta poängnivån vid separat normering av det kvantitativa blocket, men minska något på samma poängnivå vid separat normering av det verbala blocket.

Högskoleverket (2002) har genomfört en simulerad antagning till jurist- och civilingenjörsutbildningarna med den separata normering av ett kvantitativt och ett verbalt block, som här genomförts. Utfallet när endast det verbala blocket användes vid antagning till juristutbildningen var att förändringarna i den antagna gruppen var mycket små. Till civilingenjörsutbildningen skulle 16 procent lägre andel kvinnor ha antagits i provgruppen, om antagningen skulle ha baserats på endast det kvantitativa blocket. Å andra sidan skulle i detta hypotetiska fall andelen antagna kvinnor i betyggruppen ha ökat så att nettoresultatet blev att endast två procent fler män skulle ha blivit antagna. En slutsats från denna simuleringsstudie var:

...om man vill åstadkomma tydliga effekter på den antagna studentpopulationens sammansättning, t.ex. vid antagning till vissa ingenjörsutbildningar, behövs mer kraftfulla redskap än vad en separat normering identisk med eller liknande den som redovisats ovan kan åstadkomma.. (p 22)

DEL II

Erfarenheterna från försöket till normering på blocknivå är att, det måste betraktas som omöjligt att genomföra en normering på delprovsnivå. Det finns emellertid starka önskemål framför allt i forskningssyfte att kunna jämföra även delproven mellan olika provtillfällen och olika år. Avsikten med del II av denna studie är därför att genomföra en form av kalibrering av delproven.

Syfte

Syftet med den andra delen av denna rapport är att beskriva en modell för kalibrering av högskoleprovets delprov och utfallet av denna kalibrering

Metod

Eftersom två prov genomförs varje år har för närvarande 14 prov genomförts, sedan våren 1996 då högskoleprovet senast ändrades. Det är dessa 14 prov som har ingått i denna kalibreringsstudie.

Vid kalibreringen har referenspopulation I använts (se sid. 5). Denna population är jämförbar mellan olika provtillfällen avseende såväl kön, som utbildning och ålder och bör därmed ha ungefär samma prestationsnivå vid olika provtillfällen. I tabell 7 redovisas medelvärdet för referenspopulation 1 på samtliga delprov fr.o.m. våren 1996 (96A) t.o.m. hösten 2002 (02B).

Tabell 7. Delprovsmedelvärden för referenspopulation I på proven 96A – 02B.

Prov	ORD	NOG	LÄS	DTK	ELF	Tot
96A	28.0	12.5	12.0	11.0	12.4	76.0
96B	26.3	12.3	12.9	11.0	12.2	74.8
07A	24.5	11.7	12.4	9.5	10.8	68.8
97B	24.8	11.6	12.2	10.2	12.8	71.6
98A	25.7	11.1	11.4	10.9	12.7	71.6
98B	25.1	10.2	10.5	9.6	10.7	66.2
99A	25.1	8.7	12.0	10.8	11.5	68.1
99B	24.5	10.1	11.2	9.7	11.6	67.1
00A	24.3	10.6	11.5	10.8	11.0	68.3
00B	25.0	10.4	12.2	9.7	10.9	68.2
01A	24.6	9.8	12.6	10.6	10.6	68.2
01B	24.4	10.0	11.5	10.3	10.7	66.8
02A	24.2	9.9	11.9	10.7	10.5	67.1
02B	24.0	9.4	12.0	9.4	10.7	65.5
M	25.0	10.6	11.9	10.3	11.4	69.2
s	1.06	1.10	0.62	0.61	0.84	3.17
R	4.0	3.8	2.4	1.7	2.3	10.5

Som framgår av tabell 7 varierar delprovsmedelvärdernas spridningar mellan 0.61 och 1.06 och för totalpoängen är spridningen 3.17. Det är således tämligen stor variation.

Förutsatt att referenspopulation 1 verkligen är jämförbar mellan de olika provtillfällena bör i tabell 7 beskrivna skillnader kunna hänföras till svårighetskillnader mellan provversionerna.²

I tabell 8 redovisas avvikelser från totalmedelvärdet för varje delprov vid varje provtillfälle.

² Efter den genomgripande förändring av högskoleprovet som genomförts inför provet våren 1996, visade sig de första proven bli alltför lätta. Något som avsiktligt, men successivt korrigerades i senare provversioner. Fr.o.m. hösten 1998 har ambitionen varit att delproven ska ha samma svårighetsgrad.

Tabell 8. Avvikelser från totalmedelvärdet för samtliga delprov 96A – 02 A för referenspopulation I.

	ORD diff	NOG diff	LÄS diff	DTK diff	ELF diff
96:A	+3.0	+1.9	+0.1	+0.7	+1.0
96:B	+1.3	+1.7	+1.0	+0.8	+0.8
97:A	-0.5	+1.1	+0.5	-0.8	-0.6
97:B	-0.2	+1.0	+0.3	-0.1	+1.4
98:A	+0.7	+0.5	-0.5	+0.6	+1.3
98:B	+0.1	-0.4	-1.4	-0.6	-0.7
99:A	+0.1	-1.9	+0.1	+0.5	+0.1
99:B	-0.5	-0.5	-0.7	-0.6	+0.2
00:A	-0.7	0.0	-0.4	+0.5	-0.4
00:B	0.0	-0.2	+0.3	-0.6	-0.5
01:A	-0.4	-0.8	+0.7	+0.3	-0.8
01:B	-0.6	-0.6	-0.4	0.0	-0.7
02:A	-0.8	-0.7	0.0	+0.4	-0.9
02:B	-1.0	-1.2	+0.1	-0.9	-0.7

Kalibreringen genomfördes genom att varje delprovsmedelvärde korrigerades för avvikelserna från totalmedelvärdet. På individnivå kommer denna kalibrering naturligtvis bara att vara korrekt för dem som har resultat omkring medelvärdet, men eftersom kalibreringen endast är avsedd att användas i forskningssyfte kan detta dock anses acceptabelt.

Resultat och diskussion

För att kontrollera hur kalibreringen hade fungerat, prövades den på ett tioprocentigt stickprov från var och en av de 14 provdeltagargrupperna. För dessa stickprov beräknades det kalibrerade medelvärdet på varje delprov (tabell 9). I tabell 9 framgår även det ursprungliga medelvärdet för stickproven samt medelvärdet av de kalibrerade medelvärdena.

Tabell 9. Utfall, för tio-procentiga stickprov ur respektive totalpopulation, av den korrektionsfaktor som bestämdes utifrån referenspopulation I. *Antalet i varje stickprov anges med kursiv stil.*

Prov		ORD	NOG	LÄS	DTK	ELF	Tot
	M_{urspr}^3	22.6	11.8	12.2	11.1	12.2	69.8
	S_{urspr}	1.05	1.12	.69	.57	.84	3.10
	<i>Antal</i>						
96A	<i>8 155</i>	22.3	11.9	12.0	11.0	12.0	69.2
96B	<i>5 629</i>	22.6	11.8	12.2	11.2	12.1	69.9
97A	<i>8 201</i>	23.6	11.7	12.2	11.1	12.1	70.8
97B	<i>6 154</i>	22.3	12.1	12.2	11.5	12.2	70.2
98A	<i>7 475</i>	22.5	11.9	12.0	11.2	12.3	70.0
98B	<i>5 462</i>	23.0	12.1	12.2	11.2	12.5	71.1
99A	<i>6 802</i>	22.3	11.8	12.1	11.2	12.4	69.8
99B	<i>4 931</i>	22.9	11.7	11.7	11.2	12.1	69.6
00A	<i>6 175</i>	22.6	11.9	12.3	11.0	12.3	70.1
00B	<i>3 787</i>	21.9	11.8	12.2	11.3	12.0	69.5
01A	<i>4 594</i>	22.6	11.9	12.2	11.0	12.2	69.9
01B	<i>2 721</i>	22.2	11.8	12.1	11.1	12.1	69.2
02A	<i>4 025</i>	22.7	11.8	12.0	11.0	12.3	69.8
02B	<i>2 859</i>	22.6	11.6	12.4	11.2	12.3	70.0
	M_{kal}^3	22.6	11.9	12.1	11.2	12.2	69.9
	S_{kal}	.41	.15	.31	.36	.15	.53

Som framgår av tabell 9 överensstämmer medelvärdena före och efter kalibreringen tämligen väl. Den högsta avvikelser är en tiondel. Samtidigt har spridningarna av delprovsmedelvärdena minskat avsevärt för samtliga delprov efter kalibreringen.

På individnivå medför kalibreringen att vissa värden kan bli negativa och andra kan bli högre än den maximala poängen på respektive delprov. Eftersom ingen provdeltagare kan ha minusresultat och ingen heller kan ha högre resultat än antalet möjliga korrekta svar måste

³ Ovägda medelvärden

dessa värden korrigeras. I tabell 10 visas på vilka delprov dessa problem uppstår.

Tabell 11. Delprov där korrigeringar medför ”omöjliga” provresultat med respektive korrigeringsmodell

	ORD (40)	NOG (22)	LÄS (20)	DTK (20)	ELF (20)
96A		-1.9		-0.7	-0.1
96B		-0.7			
97A	40.6	-0.1		20.8	20.6
97B	40.2	-1.0		20.1	
98A			20.5		-0.3
98B		22.4	21.4	20.6	20.6
99A		23.9		-0.5	-0.1
99B	40.6	22.5	20.6	20.6	
00A	40.7		20.4		20.4
00B	40.1	22.2		20.6	20.5
01A	40.4	22.7			20.8
01B	40.7	22.6	20.4		20.7
02A	40.8	22.7			20.9
02B	41	22.2		20.9	20.9

Sammanfattande kommentar

Normering i egentlig mening kan, som nämndes i inledningen, inte genomföras på delprovsnivå, eftersom delproven innehåller alltför få uppgifter. Eftersom det i forskningssyfte kan vara av intresse att jämföra provdeltagare mellan olika provtillfällen även på delprovsnivå, har en form av kalibrering genomförts av samtliga delprov fr.o.m. våren 1996 t.o.m. hösten 2002. Anledningen till att våren 1996 utgjorde ändpunkt bakåt i tiden är att högskoleprovet dessförinnan hade en helt annan sammansättning vilket gör en kalibrering omöjlig för proven före detta tillfälle. (Det enda delprov som inte förändrades vid detta tillfälle var delprovet DTK).

Kalibreringen gjordes genom att utnyttja en av de referenspopulationer som används vid normeringen av högskoleprovet, nämligen

gen referenspopulation I. Referenspopulation I utväljs via stratifierat urval, så att den vid varje provtillfälle har samma sammansättning avseende kön, ålder och utbildning. För denna grupp fanns sedan tidigare medelvärden på samtliga delprov.

För varje delprov beräknades medelvärdet för referenspopulation I på de 14 provomgångar som kalibreringen gällde. Därefter beräknades varje enskilt delprovs avvikelse från detta medelvärde och skillnaden användes som korrektionsfaktor. Den största korrektion som gjordes var på delprovet ORD våren 1996 som sänktes med 3 poäng (se fotnot 1, s. 4); hösten samma år var sänkningen 1.3 poäng. Delproven NOG våren och hösten 1996 sänktes med 1.9 respektive 1.7 poäng, våren 1997 var sänkningen 1.1 poäng, och våren 1999 höjdes samma delprov med 1.9 poäng. Samtliga korrigeringar i övrigt var en poäng eller mindre uppåt eller nedåt.

Utfallet av kalibreringen kontrollerades genom att varje delprovs korrektionsfaktor användes för att beräkna nya medelvärden på tio-procentiga slumpmässiga stickprov ur var och en av de 14 provdeltagarpopulationerna. Kalibreringen verkade ha fungerat tillfredsställande såtillvida att de kalibrerade medelvärdena överensstämde med stickprovets ursprungliga medelvärden, så när som på en tiondel för delproven NOG, LÄS, ELF och DTK. För samtliga delprov har även spridningen på medelvärdena minskat avsevärt, vilket visar att syftet att få delproven mer jämförbara har uppnåtts.

Referenser

Braun, H. & Holland, P. (1982). *Observed-Score Test Equating: A Mathematical Analysis of Some ETS Equating Procedures*. I Holland, P. & Rubin, D. (Red.) *Test Equating*. New York: Academic Press.

Holland, P. & Rubin, D. (1982). *Test Equating*. New York: Academic Press.

Högskoleverket (2002). Högskoleprovet: Effekter på antagningen av uppdelning i verbal och kvantitativ del. Högskoleverkets skriftserie 2002:25R. Stockholm: Högskoleverket.

SOU 1985:57. Tillträde till högskolan. Betänkande av tillträdesutredningen. Stockholm: Utbildningsdepartementet.

Stage, C. & Hamrén, M. (2002). Normering av högskoleprovet våren 2002. Umeå: Opublicerat manuskript, Umeå: Enheten för pedagogiska mätningar, Umeå universitet.

Stage, C. & Hamrén, M. (2002). Normering av högskoleprovet hösten 2002. Umeå: Opublicerat manuskript, Umeå: Enheten för pedagogiska mätningar, Umeå universitet.

Stage, C. & Ögren, G. (2001). *Högskoleprovets utveckling under åren 1977-2000. Provets sammansättning och provdeltagargruppens sammansättning och resultat*. Pm nr 169. Umeå: Enheten för pedagogiska mätningar, Umeå universitet.

Stage, C. & Ögren, G. (2003). *Högskoleprovet våren och hösten 2002. Provdeltagargruppens sammansättning och resultat*. Pm nr 180. Umeå: Enheten för pedagogiska mätningar, Umeå universitet.

Svensson, A., Gustafsson, J-E & Reuterberg, S-E. (2001). *Högskoleprovets prognosvärde*. Högskoleverkets rapportserie 2001:19R. Stockholm: Högskoleverket.