

**MODELLER FÖR JUSTERING AV  
UTPRÖVNINGSDATA FÖR DELPROVEN  
ORD OCH NOG**

**Anders Lexelius  
Christina Jonsson**

Pm nr 188, 2003



ISSN 1100-696X  
ISRN UM-PED-PM--188--SE

## **Abstract**

Since 1977 SweSAT has been used as an instrument for selection in admissions to higher education. Parallel to the development of new versions of tests, extensive research and development work is undertaken, where e.g. “field test value”, i.e. a given test item’s difficulty, is estimated. In order to be able to make generalizations about the entire group of test-takers from the field test value it must be adjusted with consideration to the whole group of test-takers, i.e. not only the field trial group.

The aim of this study has been to compare the result of three models – the constant factor model, the linear regression model and the structural equation model – to adjust the field test values for the subtests WORD (Vocabulary) and DS (Data Sufficiency).

The results show that if no adjustment of field test values is made, (i.e. if values taken directly from the field test groups are used), a systematic error is introduced into estimation of a test average that should be valid for the whole group of test-takers. Furthermore, the differences between the estimations made using the three methods are small. The estimations of the three models are also revealing similar trends in relation to the test-level values for both WORD and DS. In practical terms this means that items that have been tried out in a high-achieving group are adjusted downwards and items tried out in a low-achieving group are adjusted upwards in a corresponding way in accordance with the three models.

## Sammanfattning

Högskoleprovet har sedan 1977 använts som urvalsinstrument för antagning till universitets- och högskoleutbildning. Parallellt med framtagandet av nya provversioner sker ett omfattande forsknings- och utvecklingsarbete, där bl.a. utprövningsvärdet, dvs. uppgiftens svårighetsgrad skattas. För att kunna generalisera uppgifternas utprövningsvärde till samtliga provdeltagare måste detta värde justeras utifrån hur hela provdeltagargruppen, dvs. ej enbart utprövningsgruppen, skulle ha presterat på uppgiften.

Syftet med denna studie har varit att jämföra utfallet av tre modeller, - konstant faktormodell, - linjär regressionsmodell och -strukturell ekvationsmodell, för att justera utprövningsvärden för delproven ORD och NOG.

De resultat som framkommit visar att om inte justering av utprövningsvärden sker, dvs. om värden direkt tagna från utprövningsgruppen används, så införs ett systematiskt fel vid skattning av ett provmedelvärde som ska vara giltigt för hela provdeltagargruppen. Vidare är skillnaden mellan skattningarna som gjorts av de tre modellerna liten. De tre modellernas skattningar går dessutom i samma riktning i förhållande till utprövningsvärden för såväl ORD som NOG. Mer konkret innebär det att uppgifter som prövats ut i en högpresterande grupp justeras nedåt och uppgifter som prövas ut i en lågpresterande grupp justeras på motsvarande sätt uppåt enligt de tre modellerna.

## Inledning

Högskoleprovet har sedan 1977 används som urvalsinstrument för antagning till universitets- och högskoleutbildning i vårt land. Högskoleprovet är ett urvalsprov vars syfte är att rangordna provdeltagarna med avseende på förväntad studieframgång i högskolan. Provet består av 122 flervalssuppgifter fördelade på fem delprov, som har valts med utgångspunkt i relevansen för högskolestudier. Det som mäts i provet är: ordförståelse (ORD), logiskt tänkande (NOG), läsförståelse (LÄS), förmåga att tolka diagram, tabeller och kartor (DTK), och engelsk läsförståelse (ELF).

Ett viktigt led i konstruktionen av högskoleprovet är utprövning av provuppgifter. Varje enskild uppgift prövas ut minst en gång innan den ingår i ett reguljärt prov.

Till och med våren 1995<sup>1</sup> prövades uppgifter ut på elever i gymnasieskolan. Det fanns vissa problem med dessa rutiner, varav ett hängde samman med motivationsaspekten. Eleverna, som besvarade utprövningsuppgifterna, visste att det var frågan om utprövning av uppgifter och kunde därför inte förväntas ha maximal motivation. En annan invändning var att de grupper som genomförde utprövningen var alltför små och heterogena för att resultaten skulle bli tillförlitliga. Ett återkommande problem var stora bortfall eftersom det var frivilligt att delta i utprövningarna. Dessutom medförde förändringen av gymnasieskolan, från linje till program, svårigheter att få tid för utprövningar på gymnasieskolorna

Från och med våren 1996 görs utprövningen av provuppgifter till kommande högskoleprov i samband med det reguljära provet. De nykonstruerade provuppgifterna placeras ut i ett och samma utprövningsblock. Vid varje provtillfälle prövas det ut ungefär 25 utprövningsprov fördelade på delproven ORD, NOG, LÄS, DTK och ELF.

Varje utprövningsprov ska i största möjligaste mån, till form och innehåll, efterlikna sin motsvarighet i det reguljära provet. Eftersom utprövningsproven av praktiska skäl inte kan fördelas slumpmässigt på provdeltagarna så sker fördelningen efter geografisk spridning på högskoleenhet. Med den geografiska spridningen följer även vissa brister på representativitet och generaliserbarhet vid analys av uppgiftsdata. Det innebär att de 25 utprövningsproven delas upp på samtliga högskoleenheter, dvs. ett utprövningsprov görs av provdeltagarna vid en högskoleenhet, ett annat utprövningsprov görs

---

<sup>1</sup> På grund av förändrade utprövningsrutiner genomfördes ingen utprövning av uppgifter hösten 1995.

vid en annan högskolenhet osv. Eftersom urvalet inte är slumpmässigt utan geografiskt betingat kan prestationen variera mellan högskoleenheterna.

För att utvärdera de uppgifter som ingår i utprövningsblocken finns olika statistiska tekniker att tillgå. Ett ovärderligt hjälpmedel för analys av uppgifter är det itemanalysprogram som är skraddarsytt för högskoleprovet, och som används för att analysera uppgifter, svarsförslag, prov och svarsmönster för olika individgrupperingar. Analysen av utprövningsuppgifterna är även en förutsättning för att proven ska kunna göras parallella mellan olika provtillfällen och för att skillnader mellan grupper ska kunna kontrolleras. Med parallella prov menas att olika provversioner har samma psykometriska egenskaper (t.ex. att de är lika svåra och att de rangordnar provdeltagare på samma sätt). För att ytterligare säkerställa kravet på parallellitet genomförs, en normering av den reguljära provpoängen (råpoäng), som syftar till att jämställa resultaten från olika provtillfällen. Normeringen innebär att provpoängen, som varierar inom intervallet 0-122, överförs till en skala från 0,0 till 2,0 (s.k. normerad poäng), vilket innebär att det är lika lätt eller svårt att erhålla en viss normerad poäng oavsett sammansättningen på provdeltagargruppen även om provets svårighetsgrad skulle fluktueras.

För att kunna generalisera utprövningsdata till samtliga provdeltagare, dvs. till alla provdeltagare vid samtliga högskoleenheter, justeras varje uppgifts utprövningsvärde i paritet med det värde samtliga provdeltagare skulle ha fått om utprövningsuppgiften hade ingått i deras prov. Även de textjusteringar som ibland görs efter utprövningen är något man måste ta hänsyn till när man ska bedöma uppgiftens svårighetsgrad. Om man inte skulle ta hänsyn till prestationsnivån vid högskoleenheten finns det en stor risk att uppgifternas svårighetsgrad missbedöms. Delprovets svårighetsgrad skulle med stor säkerhet, om inte högskoleenheten är representativ för samtliga provdeltagare, över eller underskattas. Problemet är att hitta ett sätt att justera dessa utprövningsvärden så att de på bästa sätt speglar det värde som den totala reguljärgruppen skulle ha fått oberoende av utprövningsgruppens prestation. Problemet med justering av utprövningsvärden har i en tidigare studie undersökts och dokumenterats när utprövningsgruppen var relativt liten och enbart utgjordes av elever i årskurs 3 i den gamla gymnasieskolan, dvs. före 1996. (Henriksson, 1992)

## Syfte

Syftet med denna studie är att jämföra utfallet av tre modeller för att justera utprövningsvärden för delproven ORD<sup>2</sup> och NOG<sup>3</sup>

## Metod

Högskoleprovets nya utprövningsmodell har medfört att ansträngningar gjorts för att skattningar av utprovade uppgifters svårighetsgrad ska vara reliabla. För att uppnå detta har tre olika modeller använts för skattning av delproven ORD och NOG. De modeller som används för att skatta uppgifternas svårighetsgrad i delproven ORD och NOG är en konstant faktormodell respektive en linjär regressionsmodell. Anledningen till att dessa modeller används har att göra med dessa provs uppgiftsformat. I både ORD och NOG presenteras uppgifterna oberoende av varandra, dvs. sannolikheten att svara rätt på en uppgift har ingenting att göra med hur man lyckas på andra uppgifter. I denna studie har jämförelse gjorts mellan dessa två modeller men också med ytterligare en modell, en s.k. strukturell ekvationsmodell. I proven LÄS, DTK och ELF finns det ett visst beroende mellan text/figur och tillhörande uppgifter. Därför omfattas inte de senare proven av denna studie.

För att studera dessa modeller utformades försöksprov för delproven NOG och ORD. Dessa försöksprov utformades på samma sätt som de utprövningsprov som ges men innehöll samma uppgifter vid samtliga provtillfällen. Försöksprovet NOG innehöll 22 uppgifter medan motsvarande ORD-prov innehöll 15 försöksuppgifter och 25 utprövningsuppgifter. Försöksproven användes vid 7 respektive 9 provtillfällen.

---

<sup>2</sup> **ORD (ordförståelse)** består av 40 uppgifter med fem svarsförslag till varje. Varje uppgift inleds med ett ord eller uttryck i fet stil. Det gäller att avgöra vilket av svarsförslagen som bäst motsvarar innebörden i det givna ordet/uttrycket. Ord av såväl svenskt som främmande ursprung ingår i provet. Provtiden är 15 minuter

<sup>3</sup> **NOG (logiskt tänkande kring matematiska problem)** består av 22 uppgifter, som alla presenterar ett problem följt av två påståenden. Här gäller det att avgöra om den information som påståendena ger är tillräcklig för att problemet ska kunna lösas. Uppgifterna kräver förkunskaper motsvarande kurs matematik A i gymnasieskolan. Uppgiftsformatet är fast, dvs. samtliga uppgifter innehåller samma uppsättning om fem svarsförslag. Provtiden är 50 minuter

## Modeller

De tre modeller som jämförs i denna rapport är således:

- konstant faktormodell
- linjär regressionsmodell
- strukturell ekvationsmodell

**Konstanta faktormodellen**, som är den enklaste av de undersökta modellerna, utnyttjar skillnaden i prestation mellan totalgruppen och utprövningsgruppen på det reguljära delprovet. Genom att bilda differensen mellan totalgruppens och utprövningsgruppens medelvärden på det reguljära provet, dividerat med antalet uppgifter i respektive delprov, erhålles en korrigeringsfaktor

$$pkorr = (M_{V_{totreg}} - M_{V_{utpreg}}) / n$$

Denna korrigeringsfaktor eller konstant adderas sedan uppgift för uppgift till de utprovade uppgifternas lösningsproportioner. Man erhåller följande generella formel för korrigering och skattningen blir då

$$(p_{i=1:n}) = putp_{i=1:n} + pkorr$$

Där  $p$  är den lösningsproportion uppgiften skulle ha haft om samtliga provdeltagare prövats på uppgiften,  $i$  är uppgiftens nummer och  $putp$  är uppgiftens lösningsproportion i utprövningsgruppen.

**Linjära regressionsmodellen**, liksom den tidigare konstanta faktor modellen, utnyttjar det reguljära delprovet, dels för samtliga prövande och dels för utprövningsgruppen, för att finna den faktor med vilken uppgifterna ska korrigeras. I motsats till den konstanta faktormodellen kan korrigeringsfaktorn med den linjära regressionsmodellen vara olika för de uppgifter som ska korrigeras, dvs. korrigeringsfaktorn är ej konstant.

För att skatta regressionslinjen används data dels från samtliga provdeltagare och dels från utprövningsgruppen. Modellen får lika många observationer som det finns uppgifter i det prov som skall skattas. Med det reguljära delprovet som utgångspunkt erhålls en lösningsproportion för varje uppgift för samtliga provdeltagare, dvs. variabeln  $y$  anges som förklarande och beroende. Den påverkande variabeln,  $x$  anges som förklarad och oberoende, utgörs av samma reguljära delprov men med den skillnaden att observationerna utgörs av lösningsproportionerna erhållna från utprövningsgruppen.

Regressionsmodellen skrivs som räta linjens ekvation och beskriver sambandet mellan de observerade gruppernas förmåga på det reguljära delprovet.

$$y = \alpha + \beta \cdot x$$

Då utprövningsprovet och det reguljära delprovet till det yttre är identiska men med olika uppgifter hämtade från samma kunskapsområden kan vi anta att proven mäter samma förmåga. Vi använder oss av den linjära regressionsmodellen för att skatta vilken förmåga, uttryckt i lösningsproportion, som krävs för samtliga provdeltagare för att lösa respektive uppgift rätt.

Genom att sätta in de observerade utprövningsvärdena ( $x_i$ ) i den skattade modellen kan skattade värden ( $y_i$ ) för samtliga provdeltagare.

$$y_i = a + b \cdot x_i + e_i; i = 1, \dots, n$$

Feltermen eller residualen ( $e_i$ ) utgör observationens avvikelse från den räta linjen i y-led. Själva skattningen, som utförts med SPSS och med Excel, sker med minsta kvadratmetoden, och fås genom att minimera summan av de kvadrerade feltermerna.

**Strukturella ekvationsmodellen** som är en multivariat metod under vilken den tidigare beskrivna linjära regressionsmodellen kan ses som ett specialfall. Gemensamt för modellerna är att de bygger på relationer mellan de beroende och oberoende variablerna.

Med ny teknik och programvarorna Amos (Arbuckle, 1997) och Streams (Gustafsson & Stahl, 1999) analyseras data med hjälp av strukturella ekvationsmodeller. Där undersöks hur olika bakomliggande latent faktorer relateras till varandra. Relationen mellan dessa faktorer kommer således att vara ett uttryck för de sanna sambanden. Indirekt observeras ett beteende av en latent faktor som i detta fall kan vara förmågan att lösa provuppgifter. Observationerna görs via en mätbar observerbar, manifest variabel, som kan vara provpoäng på ett delprov (Reuterberg & Ohlander, 1999).

Med utgångspunkt i rådata och med hjälp av strukturell ekvationsmodellering görs Maximum-Likelihoodskattningar (ML) av den kompletta kovariansmatrisen, dvs. som om samtliga provdeltagare hade deltagit i utprövningsprovet. Innan processen kommit så långt som till analys av data med strukturella ekvationsmodeller måste skapande av modellen ske.

Ett försök att skapa samband mellan data från det reguljära provet och motsvarande data från utprövningen görs. En fördel med strukturella ekvations-



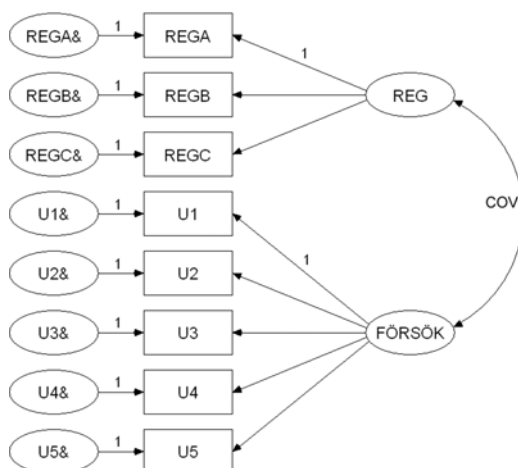
modeller är att test sker med statistiska metoder om data och modell passar ihop. När modellenpassningen är tillfredsställande, dvs. när data passar modellen, vidtar skattningen (ML) av utprövningsdata enligt figuren nedan.

Samtliga provdeltagare		Data reguljärt delprov	Data utprövning
		$r_{1,\dots,r_n}$	Skattas av modellen ( $u_{1,\dots,u_n}$ )
	Utprövningsgrupp	$r_{1,\dots,r_n}$	$u_{1,\dots,u_n}$

Mer precist används följande data för att skatta utprövningsdata med hjälp av strukturella ekvationsmodeller för de provdeltagare som *inte* ingick i utprövningsgruppen men som tillhör gruppen samtliga prövande. Observerbara data ( $u_{1,\dots,u_n}$ ) från utprövningen och utprövningsgruppen samt observerbara data ( $r_{1,\dots,r_n}$ ) från det reguljära provet och samtliga prövande används i modellen för att skatta data ( $u_{1,\dots,u_n}$ ) för dem som har resultat på det provet ( $r_{1,\dots,r_n}$ ). Där  $n$  är antalet uppgifter i delprovet.

Efter skattningen har samtliga provdeltagare ”fått utprövningsdata” på den utprövning de inte deltagit i. Modellens uppgift har varit att skatta värden för de provdeltagare som saknar utprövningsvärden. I slutänden betyder det att varje utprövningsuppgift kommer att få ett skattat värde för samtliga provdeltagare.

Figuren nedan visar hur skattningen i praktiken går till. I detta exempel skattas utprövningsvärden ( $U_1,\dots,U_5$ ) för fem uppgifter utifrån det reguljära provet (REG) och ett aktuellt utprövningsprov (FÖRSÖK)



I provet FÖRSÖK finns data för fem uppgifter från aktuell försöksgrupp (manifesta variabler).

Observerbara data finns också från reguljära provet (REG) för samtliga provdeltagare (manifesta variabler).

I bilaga 1 redovisas modellen för delproven NOG och ORD

Sedan bildas REGA, REGB, REGC som summan av var tredje uppgift i respektive delprov.

### Material för studien med NOG-provet

Data är insamlat från sju försöksprov med vardera 22 NOG- uppgifter. Utvärkningsgruppernas storlek framgår av tabellen nedan.

**Tabell 1.** Den reguljära gruppens (N) och försöksgruppens (n) storlek vid respektive provtillfälle.

	Provtillfälle						
	1	2	3	4	5	6	7
Reguljära provet (N)	49594	62537	38429	46381	27738	40228	28505
Försöksprovet (n)	1843	2550	1091	3499	3057	1391	912
Andel (%)	3,7	4,1	2,8	7,5	11,0	3,4	3,2

Av tabell 1 framgår att andelen provdeltagare som genomgått något av de 7 försöksproven NOG varierar mellan 2,8 och 11,0 procent av hela provdeltagargruppen.

### Material för studien med ORD-provet

Data är insamlat från nio försöksprov med vardera 40 uppgifter. Studien har gjorts på 15 av dessa uppgifter som haft samma placering i provet. Utvärkningsgruppernas storlek framgår av tabellen 2.

**Tabell 2.** Den reguljära gruppens (N) och försöksgruppens (n) storlek vid respektive provtillfälle.

	Provtillfälle								
	1	2	3	4	5	6	7	8	9
Reguljära provet (N)	73014	68476	49594	62537	38429	46381	27738	40228	28505
Försöksprovet (n)	2374	3567	1311	11288	3919	1491	1108	2112	1115
Andel (%)	3,2	5,2	2,6	18,1	10,2	3,2	4,0	5,2	3,9

Av tabell 2 framgår att andelen provdeltagare som genomgått något av de 9 försöksproven ORD varierar mellan 2,6 och 10,2 procent av hela provdeltagargruppen.

## Resultat

I tabell 3 kan konstateras att de utprövningsgrupper som deltagit i försöksprovet NOG presterat bättre än samtliga provdeltagare på reguljära NOG vid samtliga provtillfällen med undantag för provtillfälle tre och sju.

Genom att använda ojusterade värden, dvs. uppgiftsvärden direkt tagna från försöksgruppen, finns det en risk att man inför ett systematiskt fel vid skattning av ett provmedelvärde som ska vara giltigt för totalgruppen. Detta eftersom både försöksgruppen och totalgruppen i allmänhet skiljer sig åt i prestation. Som framgår i tabell 3 är det endast vid provtillfälle 3 som försöksgruppen och totalgruppen har likvärdig prestation på reguljära NOG-provet. Extremt blir fallet om vi inriktar oss på dem som deltog vid provtillfälle 6 då försöksgruppen presterade mycket bättre än totalgruppen. I det senare fallet skulle en ojusterad skattning, med uppgiftsdata enbart hämtad från försöksgruppen, leda till ett relativt stort fel vid skattning av totalgruppens prestation. Försöksprovets svårighetsgrad skulle med andra ord underskattas vilket skulle få som följd att provet skulle vara svårt eller kanske rentav mycket svårt för totalgruppen. Genom att justera utprövningsvärdena så att hänsyn tas till prestationsnivån i den försöksgrupp där uppgifterna är utprovade så erhålls en bättre skattning av samtliga provdeltagares prestation på uppgifterna

På motsvarande sätt kan konstateras utifrån tabell 4 att de utprövningsgrupper som deltagit i försöksprovet ORD presterat bättre än samtliga provdeltagare på reguljära ORD-provet vid tre provtillfällen nämligen provtillfälle fyra, fem och sju. Skillnaden mellan försöksgrupp och reguljärgrupp var då mer än en poäng, med den största skillnaden 1,68 poäng vid provtillfälle sju.

I tabell 3 och tabell 4 framkommer att samtliga använda modeller ”trycker ihop” de skattade medelvärdena i förhållande till motsvarande utprovningvärden. Uppgifter som prövats ut i en högpresterande grupp justeras nedåt och uppgifter som prövats ut i en lågpresterande grupp justeras på motsvarande sätt uppåt. Man kan även konstatera att de faktiska utprovningvärdena skiljer sig från de värden som skattats av de tre modellerna. Skillnaden mellan de skattningar som gjorts med konstanta faktormodellen, linjära regressions- och strukturella ekvationsmodellen är liten. De tre modellernas skattningar går dessutom i samma riktning i förhållande till utprovningvärden, detta gäller både för NOG och ORD.

Detta åskådliggörs även av figur 2 och figur 3

**Tabell 3.** Reguljära värden totalt och för försöksgruppen på delprovet NOG och värden på försöksprovet för försöksgruppen, skattade värden för samtliga provdeltagare med konstant faktormodell, linjär och strukturell modell. Medelvärden (M), standardavvikelse (s) och medelproportion (p\*) av 22 uppgifter.

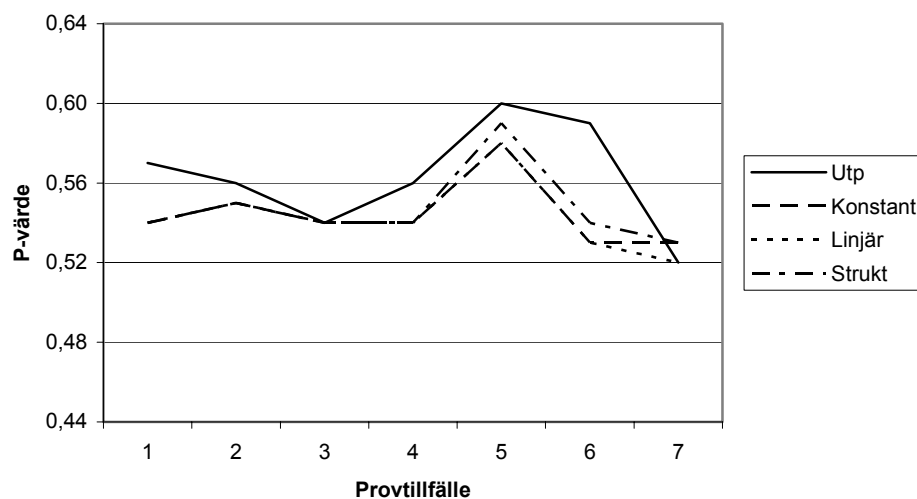
Prov	Reguljära provet NOG (22 uppgifter)									Försöksprovet NOG (22 uppgifter)								
	Totalgruppen			Försöksgruppen			Diff mellan grupperna			Försöksgrupp			Konstant faktormodell		Linjär		Strukturell	
	M	s	p	M	s	p	M	p	M	s	p	M	p	M	p	M	p	
1	11,24	4,10	0,51	11,80	3,82	0,54	-0,56	-0,03	12,49	4,77	0,57	11,93	0,54	11,87	0,54	11,94	0,54	
2	11,83	4,52	0,54	12,06	4,36	0,55	-0,23	-0,01	12,29	4,83	0,56	12,06	0,55	12,03	0,55	12,08	0,55	
3	11,80	3,87	0,54	11,80	3,82	0,54	0,00	0,00	11,97	4,79	0,54	11,98	0,54	11,98	0,54	11,95	0,54	
4	11,15	4,10	0,51	11,57	4,06	0,53	-0,42	-0,02	12,29	5,06	0,56	11,87	0,54	11,81	0,54	11,90	0,54	
5	11,22	4,21	0,51	11,47	4,11	0,52	-0,25	-0,01	13,11	4,99	0,60	12,86	0,58	12,82	0,58	12,88	0,59	
6	11,17	4,28	0,51	12,46	4,10	0,57	-1,29	-0,06	12,96	4,89	0,59	11,67	0,53	11,62	0,53	11,78	0,54	
7	10,46	3,95	0,48	10,20	3,88	0,46	0,26	0,01	11,36	5,01	0,52	11,62	0,53	11,52	0,52	11,62	0,53	

**Tabell 4.** Reguljära värden totalt och för försöksgruppen på delprovet ORD för 40 uppgifter och värden på försöksprovet för försöksgruppen, skattade värden för samtliga provdeltagare med konstant faktormodell, linjär och strukturell modell för 15 uppgifter. Medelvärden (M), standardavvikelse (s) och medelproportion (p\*)

Prov	Reguljära provet ORD (40 uppgifter)									Försöksprovet ORD (15 uppgifter)								
	Totalgruppen			Försöksgruppen			Diff mellan grupperna			Försöksgrupp			Konstant faktormodell		Linjär		Strukturell	
	M	s	p	M	s	p	M	p	M	s	p	M	p	M	p	M	p	
1	23,06	6,56	0,58	22,85	6,55	0,57	0,21	0,01	8,56	2,97	0,57	8,63	0,58	8,65	0,58	8,63	0,58	
2	22,50	7,09	0,56	22,44	6,90	0,56	0,06	0,00	8,78	2,98	0,59	8,80	0,59	8,80	0,59	8,81	0,59	
3	22,27	7,09	0,56	21,22	7,02	0,53	1,05	0,03	8,49	2,89	0,57	8,88	0,60	8,88	0,59	8,82	0,59	
4	21,79	7,41	0,54	22,64	7,51	0,57	-0,85	-0,03	9,81	2,97	0,65	9,49	0,63	9,35	0,62	9,53	0,64	
5	21,82	7,06	0,55	23,10	7,24	0,58	-1,28	-0,03	9,68	2,93	0,65	9,20	0,62	9,15	0,61	9,28	0,62	
6	22,04	7,15	0,55	21,32	6,95	0,53	0,72	0,02	8,37	2,89	0,56	8,64	0,58	8,61	0,57	8,56	0,57	
7	21,68	7,42	0,54	23,35	7,83	0,58	-1,67	-0,04	9,68	3,12	0,65	9,05	0,61	9,08	0,61	9,14	0,61	
8	21,79	7,20	0,54	21,36	7,09	0,53	0,43	0,01	8,84	2,98	0,59	9,00	0,60	8,98	0,60	8,98	0,60	
9	21,63	7,50	0,54	20,31	6,92	0,51	1,32	0,03	8,13	2,91	0,54	8,63	0,57	8,57	0,57	8,53	0,57	

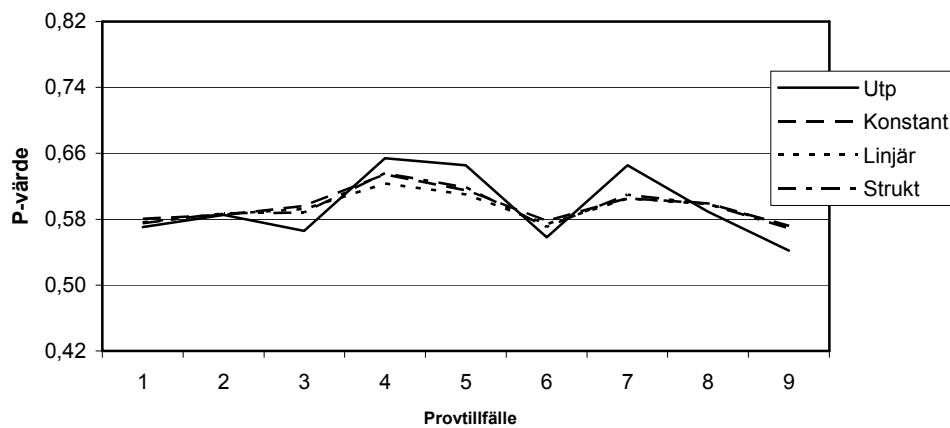
\*) Medelproportionen är avrundade värden.

### NOG-provet



**Figur 2.** Medelproportion av 22 NOG-uppgifter. Utprövningvärde (Utp), skattat värde med konstanta faktormodellen (Konstant), skattat värde med linjär regression (Linjär) och skattat värde med strukturell ekvationsmodell (Strukt).

### ORD-provet



**Figur 3.** Medelproportion av 15 ORD-uppgifter. Utprövningvärde (Utp), skattade värden med konstanta faktormodellen (Konstant), skattat värde med linjär regression (Linjär) och skattat värde med strukturell ekvationsmodell (Strukt).

Av figur 2 och figur 3 framgår, liksom av tabell 3 och tabell 4, att modellerna ganska väl överensstämmer vad gäller skattningarna. Vidare indikerar differensen mellan utprövningsvärdena och skattade värden från modellerna att vi skulle erhålla över/underskattningar av provet svårighetsgrad om inte hänsyn togs till kunskapsnivån på den grupp som deltog i utprövningen.

Om vi sedan studerar sambanden mellan de reguljära NOG-proven och motsvarande utprövningsprov för dem som har observerbara data i båda variablerna vid respektive provtillfälle finner vi korrelationer som varierar mellan 0,75 och 0,79 för de sju provtillfällena. Om man däremot korrelerar den generella latent faktorn i det reguljära provet med motsvarande generella latent faktor i utprövningsprovet, som genererats via modellen, så får man en rejäl ökning av korrelationerna med cirka 0,20 enheter till mellan 0,97 och 1,00. Om man däremot korrigerar, sambandet mellan de båda latent faktorerna för reliabilitetsbrister enligt klassisk testteori, för attenuation så erhålls korrelationsvärden mellan 0,96 och 1,00. (Predhazur & Predhazur, 1991, sid 114) Man kan med andra ord konstatera att de båda metoderna för beräkning av korrelationer med korrektion för reliabilitetsbrister ger likvärdiga resultat.

För ORD erhålls liknande resultat, med något lägre men fortfarande höga samband, mellan de reguljära ORD-proven och utprövningsproven för dem som har observerbara data på bägge proven. Korrelationerna varierar mellan 0,67 och 0,75. Om man korrelerar den generella latent faktorn i det reguljära ORD-provet med motsvarande generella latent faktor i utprövningsprovet så får man även här som för delprovet NOG en rejäl ökning av korrelationerna med värden över 0,20 enheter till mellan 0,90 och 0,97. För ORD-provet erhålls korrelationer mellan 0,87 och 0,97 vid korrektion för attenuation. Även här, liksom för NOG-provet, är skillnaden mellan de båda metodernas resultat av korrelationsvärden väldigt liten. Orsaken till att ORD har något lägre korrelationer än NOG förklaras förmodligen av att antalet uppgifter är mindre i försöksprovet för ORD än i försöksprovet för NOG. Både det reguljära NOG-provet och motsvarande försöksprov innehåller 22 uppgifter vardera medan det reguljära ORD-provet innehåller 40 uppgifter och motsvarande försöksprov 15 uppgifter. Att en latent faktor som bygger på innehållet i 40 ORD uppgifter skulle ha ett nästan 100-procentigt samband med ett prov med 15 uppgifter verkar med andra ord inte troligt.

Orsaken till ökningen av korrelationsvärdena är att de latent faktorerna visar det sanna sambandet som är opåverkat av mätfel. De höga korrelationerna indikerar även att proven kan anses vara endimensionella med avsaknad av specifika inslag. En fördel med att arbeta med strukturella ekvationsmodeller är att man kan testa hur väl den hypotetiska modellen stämmer med det faktiska datamaterialet. Root mean Square error of approximation (RMSEA) är ett statistiskt mått som ger en indikation på hur väl modellen är anpassat till data. Modeller med RMSEA värden mindre än 0,05 betraktas

som väl anpassade modeller. RMSEA är dessutom oberoende av gruppstorleken. Den modell vi valt har en mycket god anpassning till data vid samtliga undersökta provtillfällen, eftersom RMSEA värdena är klart under 0,05. För att vara exakt är RMSEA för både ORD och NOG mindre än 0,00 vid samtliga provtillfällen som ingår i undersökningen.

## Diskussion

Vi kan konstatera för den strukturella ekvationsmodellen att den hypotetiska modellen överensstämmer med samtliga faktiska datamängder som vi använt oss av i undersökningen. Ytterligare modifikation av modellen skulle inte leda till någon modellförbättring. Vi kan därför acceptera modellen som den är.

Vad kan man då säga om den konstanta faktormodellen och den linjära regressionsmodellen?

Vi har testat den strukturella ekvationsmodellen och funnit att RMSEA är 0,00 vilket indikerar att modellen inte behöver modifieras ytterligare. Den funna modellen är enkel och okomplicerad med endast ett lokalt beroende, mellan de latent faktorerna REG och försöksprov NOG respektive ORD. Eftersom såväl den konstanta faktormodellen som den linjära modellens skattningar inte nämnvärt skiljer sig från den strukturella ekvationsmodellen, kan man anta att dessa modeller skattar utprövningsvärdena lika väl eller nästan lika väl som den strukturella ekvationsmodellen.

Om man ser till användarvänligheten så är den konstanta faktormodellen eller den linjära regressionsmodellen att föredra när man ska arbeta praktisk med utprövningsuppgifter för kommande högskoleprov. Den strukturella ekvationsmodellen är kanske mer anpassad för forskning eftersom de genererar ytterligare information om uppgifterna och provet.

I normalfallet ska de justerade medelvärdena inte nämnvärt avvika mellan de olika provtillfällena. Eftersom försöksprovet är detsamma vid samtliga provtillfällen kan en avvikelse i justerad medelpoäng mellan två provtillfällen tolkas som ett mått på skillnaden i förmåga på två grupper av provdeltagare. Det resultat som framkom vid provtillfälle 5 avseende NOG-provet avviker kraftigt från övriga justerade provtillfällesvärden. Fortsatt forskning bör inriktas på att finna förklaringar till varför vissa justerade värden avviker från det normala. Med den nyvunna kunskapen i beaktande är det sedan möjligt att ytterligare förbättra modellerna för justering av utprövningsdata till kommande högskoleprov. Vidare bör även delproven LÄS, DTK och ELF omfattas av modeller för justering av utprövningsdata.



## Referenser

Arbuckle, J. L. (1997). *Amos User's Guide Version 3.6*. Chicago: SmallWaters Corporation.

Gustafsson, J-E., Stahl P-A (1999). *Streams User's Guide Version 2.5 Windows 95/98/NT*. Göteborg: MultivariateWare.

Henriksson, W (1992). *Modell för prognos av provutfall utifrån delprovspecifika ankaruppgifter* (Pm nr 64) . Umeå: Umeå universitet, Enheten för pedagogiska mätningar.

Predhazur, E. J., Predhazur S. L. (1991). *Measurement, Design, and Analysis: An Intergrated Approach*. Lawrence Erlbaum Associates, Publishers.

Reuterberg, S-E., Ohlander, S (1999). *Engelsk läsförståelse i högskoleprovet* (IPD-rapporter nr 1999:09). Göteborg: Göteborgs universitet, Institutionen för pedagogik och didaktik.

**Modell för NOG**

$$\text{NogA} = n_1 + n_4 + n_{10} + \dots + n_{22}$$

$$\text{NogB} = n_2 + n_5 + n_8 + \dots + n_{20}$$

$$\text{NogC} = n_3 + n_6 + n_9 + \dots + n_{21}, \text{där } n \text{ är uppgift i reguljära delprovet NOG}$$

Bilda två latenta variabler nog och försök

Relationen nog  $\Rightarrow$  NogA, NogB, NogC

Relationen försök  $\Rightarrow u_1, \dots, u_{22}$ ; där u är uppgift i försöksprovet NOG

Bilda kovariansen mellan nog och försök, dvs. ett lokalt beroende mellan de latenta variablerna nog och försök. Som resultat erhålls sedan skattningar av de manifesta variablerna  $u_1, \dots, u_{22}$  för dem som har resultat i de manifesta variablerna NogA, NogB och NogC, dvs. för samtliga deltagare.

**Modell för ORD**

$$\text{OrdA} = o_1 + o_4 + o_{10} + \dots + o_{40}$$

$$\text{OrdB} = o_2 + o_5 + o_8 + \dots + o_{38}$$

$$\text{OrdC} = o_3 + o_6 + o_9 + \dots + o_{39}; \text{där } o \text{ är uppgift i reguljära delprovet ORD}$$

Bilda två latenta variabler ord och försök

Relationen ord  $\Rightarrow$  OrdA, OrdB, OrdC

Relationen försök  $\Rightarrow u_1, \dots, u_{40}$ ; där u är uppgift i försöksprovet ORD

Bilda kovariansen mellan ord och försök, dvs. ett lokalt beroende mellan de latenta variablerna ord och försök. Som resultat erhålls sedan skattningar av de manifesta variablerna  $u_1, \dots, u_{40}$  för dem som har resultat i de manifesta variablerna OrdA, OrdB och OrdC, dvs. för samtliga deltagare