

# **Rätt eller fel?**

**Ett försök med ett nytt verbalt prov för högskoleprovet**

**Kerstin Andersson**

## **Abstract**

The SweSAT (Swedish Scholastic Aptitude Test) is used for selection to higher education. One of the five subtests is Vocabulary (WORD), a synonym test measuring the comprehension of words and concepts. This subtest has had the same format ever since the SweSAT was introduced in 1977. The aim of this study was to develop a new type of verbal test focusing on usage and correctness and to find out how such a test functions. The new test had 20 items consisting of one or several sentences. Some sentences were correct and some sentences contained an error. The task was to identify which sentences were correct and which ones contained an error. The results showed that the test was very difficult even if there were practically no gender differences. The biserial correlations were also very high which indicates that the test separated the high-performing students from the low-performing ones. New efforts to develop a test of this kind should be made in cooperation with teachers of Swedish in upper secondary school.

## Inledning

Alltsedan våren 1977 då högskoleprovet<sup>1</sup> infördes har ORD-provet haft samma format. Den enda egentliga förändringen är att antalet uppgifter ökat från 30 till 40. Denna förändring trädde i kraft våren 1996 som en följd av att AO-provet utgick. Rutinerna för både konstruktions- och granskningsarbetet har med tiden skärpts och förfinats men detta har inte synbart påverkat provets utseende.

Varje uppgift i ORD-provet inleds med ett fetstilt ord eller uttryck. Uppgiften är att avgöra vilket av de fem svarsförslagen som bäst motsvarar innebörden av det fetstilta ordet eller uttrycket. Orden tillhör någon av ordklasserna substantiv, verb, adjektiv eller adverb. Varje prov innehåller något eller några ålderdomliga ord, dvs. ord som var vanligare förr.

Under årens lopp har andra typer av verbala prov efterfrågats, t.ex. produktionsprov i form av uppsatsprov eller vokabulärprov i form av luckprov. Argumenten har varit att dessa skulle vara bättre instrument för att mäta språkförmåga och att det är onaturligt att ord hanteras helt uttryckta ur sitt sammanhang. Enligt Stage & Jonsson (1995) har ett antal studier (Murphy, 1982; Bolger & Kellaghan, 1990; Hellekant, 1994 m.fl.) visat att flickor missgynnas av flervalsuppgifter som uppgiftsformat. Dessutom får flickor bättre betyg än pojkar i svenska p.g.a. sina prestationer i skrivning (Garne, 1993). Ett problem med uppsatsprov är dock svårigheten att genomföra en objektiv bedömning och rättning. Försök med s.k. luckprov (Stage & Jonsson, 1995) har genomförts och kan möjligen vara en framkomlig väg. Det är dock inte helt klart om luckprov mäter ordförståelse eller läsförståelse.

Högskoleprovet som helhet har inspirerats mycket av SAT – the Scholastic Assessment Tests – i USA. Nedan ges därför en kort beskrivning av detta prov.

I mars 1994 infördes ett nytt SAT i USA. Förändringarna återspeglar nya trender i amerikansk utbildning, t.ex. fokuseringen på högre kognitiva färdigheter och det allmänna missnöjet med prov som enbart består av flervalsuppgifter. Tonvikten i de verbala delarna av SAT ligger på kritisk läsning och ord i sammanhang (critical reading and vocabulary in context) (Rigol, 1994; Young, 1994; The College Board, 1990).

---

<sup>1</sup> För en mer utförlig beskrivning av högskoleprovet, se Handbok inför Högskoleprovet, Del 1.

## ***SAT-det amerikanska exemplet***

Våren 1994 infördes ett nytt SAT i USA, **SAT I: Reasoning Test** och **SAT II: Subject Tests**. Tillsammans kallas de **SAT-the Scholastic Assessment Tests**<sup>2</sup>.

SAT I mäter verbal och matematisk förmåga och består av:

- två 30-minuterspass och ett 15-minuterspass som mäter ordförråd, verbal resonemangsförmåga och kritisk läsning (Vocabulary, Verbal reasoning, Critical reading skills)
- två 30-minuterspass och ett 15-minuterspass som mäter förmågan att lösa problem inom aritmetik, algebra och geometri (Arithmetic, Algebra, Geometry)
- ett 30-minuterspass med antingen verbala eller matematiska frågor som används i utprovningssyfte och som inte räknas in i slutpoängen. Provtagarna vet inte vilket pass det är.

Denna förändring av SAT innebär bl.a. att en större andel av den totala provtiden används till de verbala delarna av provet jämfört med tidigare även om det totalt innehåller färre frågor av verbal karaktär.

SAT II mäter kunskaper och färdigheter i speciella ämnen och förmågan att tillämpa denna kunskap. SAT II kan delas in i fem ämnesområden:

- *Engelska*: skrivning, litteratur (Writing, Literature)
- *Historia och Samhällsvetenskap*: amerikansk historia och samhällsvetenskap, världshistoria (American History and Social Studies, World History)
- *Matematik*: matematik nivå I, matematik nivå IIC (miniräknare) (Mathematics Level I, Mathematics Level IIC calculator)
- *Naturvetenskap*: biologi, kemi, fysik (Biology, Chemistry, Physics)

---

<sup>2</sup> För en mer utförlig beskrivning av SAT, se t. ex. Taking the SAT I, Taking the SAT II och Graff, 1993.

- *Främmande språk*: Chinese with Listening, French, French with Listening, German, German with Listening, Modern Hebrew, Italian, Japanese with Listening, Latin, Spanish, Spanish with Listening

Provtiden för vart och ett av dessa ämnesspecifika tester är 60 minuter och de består av flervalsfrågor, utom det som avser skrivning som består av en 20-minuters uppsats och 40 minuter med flervalsfrågor. Detta prov (The Writing Subject Test) mäter förmågan att uttrycka sig skriftligt, att identifiera fel i språkbruk och språkstruktur och att ha känsla för ords innebörd. Provet innehåller tre uppgiftstyper: att identifiera fel i meningar, att förbättra meningar och att förbättra stycken. The Writing Subject Test skiljer sig från de övriga ämnesspecifika proven (Subject Tests) genom att det avser att mäta förmågor man har utvecklat under många år. Dessa förmågor sägs vara förvärvade genom omfattande erfarenhet med språket, speciellt med det skrivna språket (Taking the SAT II: Subject Tests). Flervalsfrågorna i The Writing Subject Test kräver inte att man ska kunna definiera eller använda grammatiska termer, de avser heller inte stavning eller bruket av stor bokstav. I vissa frågor är skiljetecken, såsom semikolon, viktiga för att man ska komma fram till det rätta svaret men i dessa frågor gäller det framför allt strukturen där detta skiljetecken förekommer. Däremot kan frågorna avse tempus, pronomen, ordföljd och idiom. För en mer uttömmande lista över dessa typer av skrivproblem se Taking the SAT II: Subject Tests.

## Syfte

Syftet med denna studie är, att med utgångspunkt i förändringen av de verbala delarna av SAT, framförallt i The Writing Subject Test i SAT II, undersöka möjligheterna att konstruera en ny typ av verbalt prov för högskoleprovet och att studera hur ett sådant prov fungerar. Hur blir svårighetsgraden? Hur blir relationen mellan betyg och ett sådant prov? Gynnar ett sådant prov kvinnor eller män? Vad bör man tänka på vid konstruktionen av uppgifter till ett sådant prov? Detta är några av de frågor vi försöker besvara i denna studie.

## **Metod**

### ***Uppgifterna***

Försöket bestod av två delar: 15 s.k. ankaruppgifter (se bilaga 1) och 20 uppgifter som utgjorde det nya verbala provet (se bilaga 2), hädanefter kallat försöksprovet. Facit till både ankaruppgifterna och försöksprovet presenteras i bilaga 3.

Ankaruppgifterna bestod av 15 ord som ingått i något av de reguljära proven våren 1979 eller 1987. Dessa 15 uppgifter användes också under lång tid som ankaruppgifter i de utprövningar av nya ord som genomfördes i gymnasieskolan. Uppgifterna användes som ett komplement till medelbetygen för att kontrollera om den aktuella utprövningsgruppen avvek från den normala gruppen. Dessa uppgifter, för vilka det alltså finns data från en följd av år, valdes som en indikator på nivån för gruppen i detta försök. Provtiden var 8 minuter.

Det nya verbala provet, som bestod av 20 nykonstruerade uppgifter, avsåg att mäta kunskaper och färdigheter som är nödvändiga för att man ska kunna uttrycka sig korrekt i skrift. Provet avsåg att mäta kunskaper om grammatiska strukturer, ordval, idiomatiska uttryck och stavning. Varje uppgift bestod av en eller flera meningar. Vissa meningar var språkligt korrekta medan vissa innehöll ett fel. Uppgiften var att avgöra om meningen var korrekt eller om den innehöll ett fel. Varje uppgift innehöll tre understrykningar, som innehöll det eventuella felet, markerat A-C. Om man ansåg att meningen var helt korrekt skulle man svara D. Provtiden var 20 minuter.

Försöksprovet konstruerades så att först definierades vilken aspekt av språket varje enskild uppgift avsåg att mäta, t.ex. att det heter "valet mellan fast och rörlig ränta" i uppgift 1. Sedan konstruerades en kontext bestående av en eller flera meningar med tänkbara felaktiga svarsförslag, s.k. distraktorer. Svårigheten vid uppgiftskonstruktionen var just att skapa denna trovärdiga kontext med två distraktorer.

### ***Försökspersonerna***

Försöket genomfördes i november 1997 på naturvetenskapsprogrammet, NV, och samhällsvetenskapsprogrammet, SP, i årskurs 3 i gymnasieskolan.

Totalt deltog 131 elever i försöket, 48 från NV och 83 från SP. Av dessa var 69 flickor och 62 pojkar.

## Resultat

### *Betyg*

Eleverna ombads uppge sitt senaste betyg i svenska. Eleverna fick ange sina betyg enligt följande skala: Icke godkänd=1, Godkänd=2, Väl godkänd=3 och Mycket väl godkänd=4. Resultatet presenteras i nedanstående tabell.

Tabell 1. Senaste medelbetyg i svenska uppdelat på program, kön och totalt.

NV	SP	Män	Kv	Tot
3.15	2.65	2.68	2.97	2.83

Eftersom betygen är angivna enligt det nya systemet kan de inte jämföras med det tidigare femgradiga betygssystemet. Man kan dock konstatera att medelvärdet 2.83 är över medel och att flickorna har något högre betyg i svenska än pojkarna.

### *Ankaruppgifterna*

Medelvärdet för ankaruppgifterna blev 7.58 för totalgruppen. Motsvarande medelvärde för fyra utprövningar i början av 90-talet var 7.90. Denna minskning skulle tidigare lett till en uppskrivning av p-värdet för en uppgift utprövad i denna grupp med .02. Medelvärdet för män/kvinnor blev vid detta försök 7.58/7.58. Motsvarande värde för 90-talet var 8.09/7.76. Medelvärdet för ankaruppgifterna vid detta försök blev 8.71 för NV och 6.93 för SP. Resultatet för ankaruppgifterna är genomgående lägre vid detta försök jämfört med tidigare data men då måste man komma ihåg att det vid detta försök handlar om en relativt liten grupp.

### *Försöksprovet*

Nedanstående tabell visar hur de 20 uppgifterna i försöksprovet fungerade statistiskt. Medelvärdet för NV blev 7.02 och för SP 5.95. Korrelationen, sambandet, mellan ankaruppgifterna och försöksprovet var .49. Korrelatio-

nen mellan betyg och ankaruppgifter var .41 och korrelationen mellan betyg och försöksprov var också .41. Korrelationen mellan de två proven var alltså något högre än korrelationen mellan vart och ett av proven och betygen, medan korrelationen mellan vart och ett av proven och betygen var likvärdig.

Tabell 2. p-värden totalt, p-värden för män och kvinnor samt rbis för de 20 uppgifterna i försöksprovet.

Uppg. nr.	p tot	p män/ p kv	rbis
1	.24	.31 /.19	.37
2	.53	.55 /.52	.27
3	.25	.24 /.26	.58
4	.43	.39 /.46	.22
5	.21	.26 /.17	.40
6	.58	.50 /.65	.39
7	.41	.40 /.42	.56
8	.24	.24 /.23	.36
9	.11	.10 /.13	.34
10	.30	.27 /.32	.24
11	.67	.68 /.67	.53
12	.54	.50 /.58	.55
13	.18	.24 /.12	.69
14	.18	.18 /.19	.35
15	.18	.16 /.20	.24
16	.27	.31 /.23	.50
17	.21	.21 /.22	.24
18	.45	.40 /.49	.18
19	.18	.21 /.16	.74
20	.16	.21 /.12	.62
	6.34	6.35 /6.33	

Detta prov bestod alltså av 20 uppgifter och eftersom varje rätt svar gav 1 poäng kunde totalpoängen maximalt bli 20. Medelvärdet för totalgruppen blev 6.34 och standardavvikelsen  $s = 2.63$ . Medelvärdet för män/kvinnor blev 6.35/6.33. Det kan således konstateras att provet blev svårt men att könsskillnaden blev mycket liten. NV hade det bästa resultatet på försöksprovet. En jämförelse med betyg och resultat på ankarord visar att NV också har det högsta betyget i svenska och det bästa resultatet på ankarorden.



Analysen av de enskilda uppgifterna i försöksprovet har gjorts på basis av tre av de viktigaste värdena vid sammansättningen av prov till högskoleprovet:

1. lösningsfrekvenserna, p-värden
2. den biseriala korrelationen, rbis
3. könsskillnader

Lösningsfrekvensen anger hur många procent av provdeltagarna som besvarat frågan rätt.

Den biseriala korrelationen anger hur väl uppgiften i fråga samvarierar med totalpoängen på provet. Detta värde används också som kontroll på provets homogenitet. Gränsen har satts vid .30, dvs. samtliga uppgifter i ett prov ska ha ett rbisvärde som är minst .30 för att ingå i ett reguljärt prov.

Könsskillnaderna anger hur många procent av pojkarna respektive flickorna som besvarat frågan rätt.

I ett "verkligt läge" används inget av de här två värdena isolerat utan en värdering görs alltid av relationen mellan dem. I just denna studie har dessa variabler studerats var för sig för att se om utfallet blir olika beroende på vilken av dessa variabler som ligger till grund för analysen.

### ***Analys av lösningsfrekvenser***

Lösningsfrekvensen, p-värdet, för 6 av uppgifterna i försöksprovet blev mindre än .20. Det var uppg. nr. 9, 13, 14, 15, 19 och 20 som därför måste sägas ha blivit för svåra. Dessa uppgifter kommer inte att diskuteras under denna rubrik.

Tio uppgifter, nr. 1, 3, 4, 5, 7, 8, 10, 16, 17 och 20 fick p-värden mellan .20 och .50.

Fyra uppgifter, nr. 2, 6, 11 och 12 fick p-värden över .50.

Fjorton uppgifter fick alltså acceptabla p-värden. Vid urval av uppgifter till reguljära prov är det också viktigt att hela uppgiften har fungerat, dvs. rätt svar ska ha det högsta p-värdet och resten ska fördela sig någorlunda jämnt över distraktorerna. Enligt detta krav skulle endast en uppgift, nr. 11, kunna godkännas. Denna uppgift hade också ett bra rbis, .53.

I detta prov fanns endast två noll-distraktorer, dvs. svarsförslag som ingen hade valt. Omvänt kan sägas att ett alltför stort antal distraktorer hade lockat alltför mycket. Detta tyder på att provet är för svårt och att gissningsfrekvensen är hög.

### ***Analys av biseriala korrelationer***

Den biseriala korrelationen,  $r_{bis}$ , är som redan nämnts också mycket viktig vid provsammansättningen. Man kan säga att  $r_{bis}$  ”sällar agnarna från vetet”. Kombinationen acceptabelt p-värde och  $r_{bis}$  över .30 är förstås idealet men en uppgift med lågt p-värde men bra  $r_{bis}$  säger att just den uppgiften har lösts av de duktiga eleverna.

Totalt fjorton uppgifter i försöksprovet fick  $r_{bis}$ -värden över .30, en del av dem t.o.m. mycket höga sådana.

### ***Analys av könsdifferenser***

Totalt sett blev det ingen nämnvärd könsskillnad på försöksprovet, 6.35 för pojkarna och 6.33 för flickorna. Det varierar däremot en del mellan uppgifterna. Elva uppgifter har en könsskillnad på högst .05 till antingen flickornas eller pojkarnas fördel. Den största enskilda skillnaden är .15. Sammantaget kan sägas att könsskillnaderna inte är alltför stora och att uppgifter som gynnar pojkar kan utjämnas med uppgifter som gynnar flickor.

## **Diskussion**

Syftet med denna studie var att undersöka möjligheterna att konstruera en ny typ av verbalt prov för högskoleprovet och att undersöka hur ett sådant prov fungerar.

Mätinstrumentet bestod av två prov, ett ankarprov bestående av 15 uppgifter, dvs. uppgifter för vilka man sedan tidigare har statistiska data för olika grupper samt ett nykonstruerat försöksprov. Tonvikten i detta försöksprov låg på korrekt språkbruk och avsåg att mäta kunskaper om grammatiska strukturer, ordval, idiomatiska uttryck och stavning.

Resultatet för ankarprovet visade att denna utprövningsgrupp var något svagare än den grupp för vilken ankarprovsdata föreligger sedan tidigare. Resultatet för försöksprovet visade att det blev ett mycket svårt prov. Provet kunde ge maximalt 20 poäng och resultatet för detta försök blev 6.34. Där emot fick de flesta uppgifter bra och t.o.m. höga  $r_{bis}$ -värden vilket talar för

att provet skiljer ut de högpresterande från de lågpresterande. En orsak till att det blev så svårt är provkonstruktörernas bristande erfarenhet av just denna typ av uppgiftskonstruktion. Svårigheten låg i att skapa en trovärdig kontext till just det man ville mäta med respektive uppgift.

Vad gäller könsskillnader verkar det som om man med denna typ av prov kan balansera ut dem ganska bra.

Sambandet mellan ankaruppgifterna och försöksprovet är något högre än sambandet mellan betygen och vart och ett av proven.

Sammanfattningsvis kan konstateras att det säkert är fullt möjligt att konstruera ett nytt verbalt prov av denna typ även om just detta försöksprov visade sig bli alltför svårt. Det vore intressant att göra ett nytt försök med ett lättare prov. För att finna rätt svårighetsnivå bör man samarbeta med gymnasielärare i svenska som har kunskap om och erfarenhet av vad eleverna kan. Svårigheterna att konstruera uppgifter av denna typ kan avhjälpas genom träning och genom användning av autentiskt material, t.ex. tidningsartiklar.

## Referenser

Bolger, N. & Kellaghan, T. (1990). Method of measurement and gender differences in scholastic achievement. *Journal of Educational Measurement*, 27, 165-174.

College Board. (1990). *Beyond prediction. Commission on new possibilities for the admissions testing program*. New York: College Entrance Examination Board.

College Board. (1994). *Taking the SAT I: Reasoning test*. New York: College Entrance Examination Board.

College Board. (1994). *Taking the SAT II: Subject tests*. New York: College Entrance Examination Board.

Enheten för pedagogiska mätningar. (1998). *Handbok inför högskoleprovet, Del 1*. Umeå universitet.

Garne, B. (1993). *I tal & skrift*. Malmö: Gleerups Förlag.

Graff, A. S. (1993). The new SAT: the future of transition assessment. *Education Libraries*, 17, 3, 7-13.

Hellekant, J. (1994). Are multiple-choice tests unfair to girls? *System*, 22, 3, 349-352.

Murphy, R.J.L. (1980). Sex differences in GCE examination entry statistics and success rates. *Educational Studies*, 6, 169-178.

Rigol, G.W. (1994). The new SAT and education reform in the 1990s. *The College Board Review*, 170, 26-31.

Stage, C., Jonsson, I. (1995). Luckprov med flervalsuppgifter. Ett alternativt ORD-prov? *Pedagogiska mätningar*, 96. Umeå universitet.

Young, J.W. (1994). The new SAT. *College and University*, 2, 146-149.