

Godkänd i det nationella provet i svenska?

Eva Östlund-Stjärnegårdh

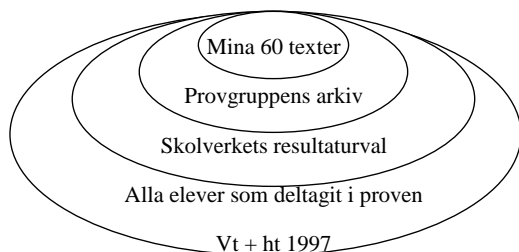
När Svenskans beskrivning hölls i Lund hösten 1996 och hade skolan och svenskämnet som tema hade jag precis tillträtt som konstruktör av de nationella proven i svenska för gymnasieskolan och var med på konferensen för att lyssna och lära. Sedan dess har jag i anslutning till provarbetet påbörjat en undersökning av hur lärare bedömer elevtexter. Fokus i denna undersökning är betygsstegen Godkänd – Icke godkänd, en betygsgräns som vållar svensklärare mycken vånda, eftersom det för elever i gymnasiet och komvuxutbildningen är oerhört viktigt att nå godkända betyg med tanke på studier och arbete efter gymnasieutbildningen.

Undersökningens uppläggning

Min undersökning består av tre delar. Den första är en lärarenkät som är bearbetad och presenterad i provgruppens rapportserie Svenska i utveckling (Östlund-Stjärnegårdh 1999). Enkäten har besvarats av 120 svensklärare från hela landet. Resultaten visar att lärarna i stort är överens om vilka kriterier som är de viktigaste när det gäller att avgöra om en elevtext ska godkännas eller inte. Hur lärare ser på helhetsbedömning kontra bedömning av delaspekter som innehåll, språklig korrekthet och stilistisk anpassning till skrivsituationen belyses av enkätsvaren.

Ett 20-tal lärare, som anmält sig frivilligt på enkäten, har varit sysselsatta med bedömning inom min undersöknings andra del. Denna fas av undersökningen har pågått under läsåret 1998–99. Den tredje delen av undersökningen kommer att utgöras av analys av elevtexterna. Genomgående har jag anledning att jämföra mina resultat med Gymnasistsvenska (Hultman & Westman 1977), även om författarna där beskriver ett system med relativa sifferbetyg och dagens gymnasister betygsätts med kursrelaterade bokstavs-betyg. Det är ändå inte möjligt att syssla med svenska elevtexter utan att ta avstamp i denna klassiker. Jag refererar också på flera ställen till Berges avhandling (1996) om det norska bedömnings-systemet med ”sensorer” och hur de sensorerna tänker och gör när de betygsätter i studentexamen. Även om Sverige och Norge har olika system kan intressanta jämförelser göras mellan Berges och mina resultat.

Figur 1 beskriver de olika nivåerna som jag rör mig mellan. Hur många som faktiskt deltar i provet vet vi inte exakt, men provet trycks i 100 000 exemplar på våren och 60 000 på hösten. Skolverkets statistiker gör ett urval bland de skolor som deltar och för eleverna på de skolorna skickas resultat på de olika provdelarna in. Dessa resultat redovisas i en rapport från Skolverket varje termin (Gymnasieskolans kursprov vårterminen 1997 etc.). För ett mindre urval av de inrapporterade eleverna skickas lösningarna, de skrivna texterna, in till provgruppen. De elevtexter som jag använder mig av kommer alltså från provgruppens arkiv, närmare bestämt från de två proven år 1997.



Figur 1. Materialurval på olika nivåer.

Hur stort Skolverkets resultaturlval är för år 1997 framgår av tabell 1. Som synes är det en betydande skillnad i hur många som deltar i provet på höst- respektive vårtermin. Elevlösningar skulle inom det urvalet skickas in för elever födda ett visst datum, och detta har till provgruppens arkiv gett totalt 250 texter för våren och knappt 100 för hösten. Av dem har jag valt ut elevtexter som av elevens egen lärare bedömts med betygen Godkänd eller Icke godkänd.

Tabell 1. Betygsfördelning för B-uppgiften i Skolverkets resultaturlval

Prov	Gy/Kx	Betygsfördelning i procent				Antal elever	Andel Gy/Kx per termin
		IG	G	VG	MVG		
Vt-97	Gy	9	53	33	6	6741	89
	Kx	4	45	44	6	814	11
Ht-97	Gy	15	48	28	9	1667	71
	Kx	9	48	36	6	685	29

B-uppgiften är den långa skrivuppgiften och det är lösningar av den som jag plockat fram. Andelen IG har legat runt 10 % i samtliga prov hittills. Provet innehåller också en muntlig uppgift och en kortare, mer bunden skrivuppgift.

Sammanlagt 60 texter från de två proven ingår i mitt material. För hösten -97 är det samtliga texter insända med IG eller G, 26 stycken. Från våren -97 har jag helt enkelt fyllt på upp till 60 texter med de 34 först registrerade i provgruppens databas. Andelen komvuxlösningar är sammantaget 18 %, en rimlig snittandel om vi ser till hur många som deltar i proven från respektive skolform.

Bedömning av elevtexter

Bedömningen har i nästa steg gått till så att varje text har granskats av tre bedömare oberoende av varandra, och några texter av fler som en kontroll. Denna fas av arbetet är avslutad och har resulterat i ett antal siffror och tabeller som jag ska presentera valda delar av i denna artikel. Den enklaste redovisningen ser ut som i tabell 2. Det kan diskuteras hur många bedömare per text man ska arbeta med. Berge skriver om det statistiskt "sanna" betyget, som skulle uppnås med fem bedömare (1996:214). För mig har det varit en anpassning till den praktiska tillgången på frivilliga bedömare (se nedan) som fått bestämma antalet.

Tabell 2. Insända betyg kontra snittbetyg i bedömningen

Insänt betyg	Bedömarnas snittbetyg			
	IG	G	VG	S:a
IG	9	1	–	10
G	26	23	1	50
S:a	35	24	1	60

Av de tio texter som skickats in som IG har 9 fått det betyget också av mina bedömare. Av de femtio texter som skickats in som G har inte mindre än 26 underkänts. Det kan finnas flera förklaringar till det förhållandet. Den första förklaringen som inställer sig är att det är lättare att ge betyget IG till en elev man inte känner. I sina egna elevers texter försöker man alltid se till de positiva inslagen. Man har också kunskap om allt annat eleven presterat, t.ex. kanske goda insatser i den muntliga provdelen. Slutsatsen av detta resonemang är att den elevlösning som av elevens egen lärare får betyget IG på det nationella provet verkligen förtjänar det. Viktigare för eleven är ju sedan vad han eller hon får i betyg på kursen, men elevernas kursbetyg är inte ämnet för denna studie. En annan förklaring till den stora andelen IG i mitt bedömningsmaterial kan vara den instruktion bedömarna fick: "Dessa texter har skickats in med IG eller G." Bedömarna har tagit sitt uppdrag på stort allvar och möjligen överbetonat IG, trots att jag också uppmanade dem att använda högre betyg om de ansåg det motiverat.

Bedömarna är i huvudsak lärare som vid de två tillfällen jag genomfört enkäten svarat att de är intresserade av att delta som frivilliga bedömare. Intresse är en sak och deltagande en annan: trots ca 30 anmälda har jag under bedömningsfasen blivit tvungen att fylla på med tre personer som inte medverkat i enkätundersökningen. De har valts bland lärare som utför utprovningar av skrivuppgifter åt provgruppen men inte ingår i den trängre cirkeln som deltar i möten och diskussioner flera gånger under året. Sammanlagt har 21 bedömare medverkat och de har i möjligaste mån fördelats så att de fått texter från ett prov de angett att de tidigare arbetat med, naturligtvis för att bespara dem extra arbete med inläsning av okänt material. Skälet till att jag velat ha den här kopplingen mellan enkätbesvarande och bedömande är att jag tänker mig att undersöka elevtexterna utifrån de kriterier lärarna i sina enkätsvar angett att de använder sig av i arbetet med att skilja Godkänd från Icke godkänd. En motsvarande amerikansk undersökning av Breland & Jones (redovisad i Huot 1990) visar att de fyra viktigaste kriterierna i enkäten med bedömarna stämde med den verkliga bedömningen.

Texterna har skrivits ut från dator för läsbarhetens och anonymitetens skull. De har bara försetts med ett nummer; skribentens kön, gymnasieprogram eller ålder har inte angetts och självklart inte heller insänt betyg. På ett bedömningsprotokoll har bedömarna fått fylla i ett betyg per text, gärna modifierat med plus- eller minustecken. Det har varit frivilligt att kommentera bedömningen, men majoriteten av bedömarna har lämnat kommentarer till de flesta betygen. Där finns ett stort material som kommer att vara till god nytta när jag kommer in på textanalysen.

För att kunna räkna på betygsättningen i denna bedömningsundersökning har jag valt att tillämpa en modell för sammanvägning av olika betyg som används i det nationella provet för skolår 9 (Vägledning för bedömning, vt -99). På en 10-gradig skala har bokstavsbetygen (G, VG, MVG) värden enligt nedanstående sammanställning. För grundskolan finns inte betyget Icke godkänd.

- Målen inte uppnådda 1 – 2
- Godkänd 3 – 5
- Väl godkänd 6 – 8
- Mycket väl godkänd 9 – 10

”Normalläget” för Godkänd är 4 och för Väl godkänd 7 enligt modellen. Talen 2, 5 och 8 motsvarar prestationer som ligger nära gränsen för högre betyg. I betygssystemet för gymnasieskolan finns fyra steg och t.ex. G+ används inte i Skolverkets publikationer eller på elevers betygsblanketter. Men många lärare laborerar ändå med plus- och minustecken vid betygsättningen av enstaka prestationer, t.ex. skrivuppgifter. Så är också fallet i de protokoll som mina bedömare fört vid bedömningen, något som jag i och för sig uppmuntrat dem till, varför jag väljer att använda följande omräkningsskala.

IG	1
IG+	2
G-	3
G	4
G+	5
VG-	6
VG	7
VG+	8
MVG-	9
MVG	10

En elevtext betygsatt av fem lärare får utgöra exempel på hur omräkningskalan används. Texten har fått betygen IG, IG+, G+, G och G. Enligt den här skalan blir sifferbetygen 1, 2, 5, 4 och 4, summa 16. Medeltalet blir 3,2 och det sammanvägda betyget G-, alltså ett betyg i underkanten av Godkänd. Vanliga avrundningsregler medför att spannet för IG kommer att vara 1–2,4, för G 2,5–5,4, för VG 5,5–8,4 och för MVG 8,5–10. I den här studien kommer dock inte hela modellen till användning.

För symmetriens skull borde kanske IG- och MVG+ finnas. Flera lärare har dock försäkrat mig att de aldrig använder IG-. En text som skulle få ett sådant betyg är inte färdig, egentligen inte bedömningsbar, utan återlämnas till eleven för vidare arbete. MVG+ är möjligen tänkbart som betyg, men då den kombinationen inte dykt upp i en utprovning av beräkningsmodellen räknar jag tills vidare inte med den. MVG är ju heller inte aktuellt i det material jag nu håller på med.

En annan möjlig beräkningsgrund vore att använda det omräkningssystem som finns för ansökan till högskolan. Där ger kursbetyget IG 0 poäng, G 10, VG 15 och MVG 20. I bedömningen av enskilda elevtexter anser jag dock att det systemet implicerar en större skillnad mellan icke godkända och godkända texter än vad det faktiskt är frågan om, och jag har därför valt att använda 1–10 för omräkningen.

Betyg och variation

Min idé vid planeringen var att bedömningsomgången skulle resultera i tre ”högar” med texter: säkra G, säkra IG och en ”mittimellan-hög” med texter för vilka bedömarna var tveksamma om betyget. Nu har resultatet kanske inte blivit så enkelt. Om man bara ser på snittbetygen enligt den omräkningsmodell jag presenterat, kan man tala om tre högar. Tabell 3 förklarar först sorteringsprincipen:

Tabell 3. Antal texter per betygssteg

Snitt-betyg	IG	IG+	G-	G	G+	VG-	
	1,0–1,4	1,5–2,4	2,5–3,4	3,5–4,4	4,5–5,4	5,5–6,4	
Antal texter	18	17	10	9	5	1	S:a 60

Det är alltså så många som 35 texter som hamnar på IG-nivån. Om man tänker sig IG+ och G- som gränssfall utgör de texterna mellanhögen. Klart godkända är 15 texter som får G eller mer. Med en smärre förändring i indelningen blir sorteringen den som visas i tabell 4.

I spannet IG+ till G- skulle man kunna avgränsa snittbetygen 2,3 till 2,7. Där finns tolv texter som ju verkligen ligger på gränsen mellan Godkänd och Icke godkänd. I en insändare i Skolvärlden (Lärarnas riksförbunds tidning) efterlystes betyget MTG för den kategorin; betygssteget Med tvekan godkänd har dock ännu inte införts av skolmyndigheterna. I den här högen måste i alla fall gränsen mellan IG och G gå. Frågan är hur jag ska definiera den.

Tabell 4. Antal texter per betygssteg inkl. MTG = Med tvekan godkänd

Snitt-betyg	IG	IG+	MTG	G-	G	G+	VG-	
	1,0–1,4	1,5–2,2	2,3–2,7	2,8–3,4	3,5–4,4	4,5–5,4	5,5–6,4	
Antal texter	18	8	12	7	9	5	1	S:a 60

Om jag tar med det betyg som elevens egen lärare satt och alltså räknar snittbetyg på fyra lärares bedömning händer en hel del, vilket man kan förstå av tabell 2 över insända betyg kontra snittbetyg. Med ett snitt från fyra bedömare får 14 texter högre betyg än de hade med bara tre bedömare, dvs. passerar gränsen och går från IG+ till G-. En enda text får lägre betyg och det är den enda som med tre bedömare tagit sig upp till VG- men som alltså stannar på G+ med fyra bedömare. I fortsättningen tänker jag dock arbeta med snittbetygen baserade på tre oberoende bedömare. I stället får jag ständigt ha i åtanke vid analysarbetet att det är så här stor skillnad på insända betyg och mina snittbetyg.

Om man ser till variationen i bedömning kan man sortera högarna lite annorlunda, vilket framgår av tabell 5. De flesta texterna i IG-högen är självklara; av dem får 15 IG av samtliga bedömare och övriga något IG+. Två texter får också G- av samtliga bedömare. De elevtexter som alla bedömare är

Tabell 5. Antal texter på olika variationsnivåer

	Antal texter	Stdavv
Stor variation	2	2,31
	10	1,73
	3	1,53
Viss variation	12	1,15
	1	1,13
	1	1,01
	3	1,00
	11	0,58
Ingen variation	17	0,00

överens om är enligt Löfqvist (1989:89) särpräglade. En 5:a (högsta betyg) som alla är ense om är "inte alls typisk utan snarare extrem". Mindre särpräglade och extrema är en stor grupp, 28 texter, alltså ca hälften av de här redovisade texterna som uppvisar en variation uttryckt i standardavvikelse på 0,58 till 1,15. I de fallen är två av tre bedömare överens till och med om eventuella plus- eller minustecken. Inom den gruppen förekommer samtliga utnyttjade betygssteg från IG till VG-. I den tredje gruppen, den med störst variation i betygsättningen, ingår 15 texter med snittbetyg från IG+ till G+. Flest texter finns med IG+ (5 stycken) och G- (6 stycken), vilket väl egentligen bara visar det jag utgått ifrån, nämligen att det är svårt att dra gränsen mellan IG och G. Allra störst variation, med en standardavvikelse på 2,31, uppvisar två texter som båda fått betygserien IG, IG och G+. Snittbetyget blir 2,3, det vill säga IG+, men för varje text finns det alltså en bedömande lärare som vill ge texten G+. En av dessa texter är en skönlitterär text, första kapitlet till en roman utgående från en konstbild, och det har visat sig i lärarenkäter och på andra sätt att det är en svårbedömd skrivuppgift, som ett antal lärare ogillar av den anledningen.

Tre av de 60 texterna har bedömts av nio bedömare vardera för att utgöra kontrolltexter. Resultaten för dessa tre visas i tabell 6. Kontrollen består i att se om någon bedömare avviker väldigt mycket från övriga genom att vara särskilt "sträng" eller särskilt "snäll". Vid all bedömning finns risken att någon bedömare arbetar med en idiosynkratisk betygsskala och inte ingår i bedömningsgemenskapen. Enligt Berge (1996:258) finns ett visst "normfelleskap" bland de norska "sensorerna" i hans undersökning men denna gemenskap har oklara gränser. I Gymnasistsvenska (Hultman & Westman 1977:25) redovisas "en gemensam skolnorm", dvs. de ingående lärarna är sinsemellan någorlunda över-

ens om betygen. Lärarna är i alla fall mer överens med varandra än med de i bedömningen medverkande journalisterna. Min undersökning av variationen mellan bedömarna är inte avslutad men intrycket är att ingen av de ingående bedömarena behöver diskvalificeras.

Tabell 6. Kontrolltexternas betyg och snittbetyg

Text										Snittbetyg	Stdavv
10	4	4	3	1	3	4	4	4	4	3,4 = G-	1,01
34	1	1	3	1	4	1	1	1	1	1,6 = IG+	1,13
50	1	1	5	1	5	3	3	1	4	2,7 = G-	1,73

Text 10 har skickats in som Godkänd och är också helt klart godkänd av de oberoende bedömarna; endast en av nio har satt IG. Variationen blir inte heller särskilt stor. Rutan markerar de tre bedömare som är inblandade i två av kontrolltexterna. I övrigt är det olika bedömare för de olika texterna. Text 34 är inskickad som Icke godkänd och det blir den ju också här. De två bedömare som har satt G- respektive G (3 och 4 poäng) visar sig ligga generellt högt och är också de två som gett Text 50 G+ (5 poäng). I rutan syns att de tre lärare som skiljer Text 34 och Text 50 åt är helt överens om rangordningen: text 50 är den bättre av de två. Berge anser att de få "sensorer" som rangordnar systematiskt annorlunda än majoriteten avviker så mycket att de inte kan sägas ingå i bedörmargemenskapen. Det gör däremot de som är "snille sensorer" (1996:298). Just Text 50 är den intressantaste för mig. Den är inskickad som Godkänd och får här lägsta tänkbara G med stor variation. Texten är en lösning av en skrivuppgift som innebär att skriva manus till ett tal, en uppgiftstyp som diskuterats nästan lika mycket som den skönlitterära som nämnts ovan.

Att olika uppgiftstyper är olika lätta att bedöma är knappast något nytt. Inte heller att lärare har olika åsikter om olika provuppgifter. Berge (1996:514, 516) redovisar djupintervjuer med norska "sensorer" där en stor skepsis gentemot skönlitterärt skrivande som examensuppgift redovisas. Också av Berges lärarenkäter framgår att novellskrivande är den uppgift som uppskattas minst. I Norge har i stället så kallade pensumrelaterade uppgifter mottagits väl (Berge 1993:162 ff.).

Jag får säkert anledning att titta närmare på t.ex. könsskillnader i elevtexterna. Tabell 7 visar ett genomgående betygsmönster i alla kursprov och de olika provdelarna. Flickor/kvinnor har högre betyg helt enkelt. Det är därför helt enligt mönstret att männen i min undersökning har så mycket större andel IG än kvinnorna. För provgruppens arkiv av insänt material och för det stora resultaturlvalet som Skolverket gjort har jag bara med siffrorna från vårterminen men mönstret är detsamma. Det intressanta är att denna könsskillnad består även när lärare be-

dömer anonyma, neutrala texter. Varken namn eller handstil eller program eller något annat avslöjar skribentens kön och ändå kvarstår betygsskillnaden. Skillnaden har med andra ord inte med kön att göra utan med kvalitet – som i sin tur har med kön att göra.

Tabell 7. Könsskillnader i betyg på B-uppgiften i procent; mina texter, totalt insända och resultaturlvalet

Textgrupp	Kön	IG	G	VG	MVG
Mina 60 texter	Kvinnor	54	46	–	–
	Män	61	36	3	–
Insänt material vt -97 226 elevtexter	Kvinnor	8	45	42	6
	Män	11	58	26	5
Resultaturlval vt -97 6684 elever	Kvinnor	5	46	40	9
	Män	10	56	29	5

Jag har tittat lite på variation kontra kön. Variationen uttryckt i standardavvikelse tycks inte vara högre för någotdera könets betyg. När standardavvikelsen är 1 eller mer, fördelar sig texterna på varje nivå helt jämnt mellan manliga och kvinnliga elever. I den textgrupp där standardavvikelsen är 0 är det många fler män, men det hänger samman med att det i huvudsak är betyget IG bedömarna är helt överens om, och det är de manliga eleverna som i stor utsträckning skrivit IG-texterna.

Avslutning

Tolkningen av alla dessa siffror gör att jag ideligen kommer in på vilka elevtexter det handlar om och vad som karakteriserar dem. Är den ena uppgiftstypen svårare att bedöma än den andra? Jag har nämnt uppgifterna skönlitterära texter och manus till tal som uppgiftstyper som kan vålla problem. Ställs olika krav för Godkänd i olika uppgiftstyper? Kan jag i texterna spåra det som lärarna i enkäten säger sig gå efter: röd tråd, textens funktion i den givna situationen, relevant innehåll osv. Det inställer sig spännande frågor som jag inte hunnit arbeta med än, så svaren på dem ska jag be att få återkomma med i annat sammanhang, kanske nästa Svenskans beskrivning. Skolforskningen kan nog behöva en egen sektion då också.

Litteratur

- Berge, Kjell Lars, 1993: På søking etter tolkningsfellesskap. I: Skrivepedagogisk fornying. Utg. av Inge Moslet & Lars Sigfred Evensen. Oslo. S. 151–190.
- Berge, Kjell Lars, 1996: Norsksensorenes tekstnormer og doxa. En kultursemiotisk og sosio-tekstologisk analyse. Norges teknisk-naturvitenskapelige universitet, Trondheim.
- Gymnasieskolans kursprov höstterminen 1997. Resultatredovisning till skolorna. Skolverket, Stockholm.
- Gymnasieskolans kursprov vårterminen 1997. Resultatredovisning till skolorna. Skolverket, Stockholm.
- Hultman, Tor G. & Westman, Margareta, 1977: Gymnasistsvenska. (Skrifter utgivna av Svenskläraryöreningen 167.) Lund.
- Huot, Brian, 1990: The literature of direct writing assessment: major concerns and prevailing trends. I: Review of educational research 60:2. S. 237–263.
- Löfqvist, Gert, 1989: Lärare bedömer uppsatser. (Pedagogisk-psykologiska problem nr 510.) Malmö.
- Vägledning för bedömning och exempel på elevers lösningar med kommentarer. Nationellt ämnesprov i svenska för skolår 9, vårterminen 1999. Skolverket, Stockholm.
- Östlund-Stjärnegårdh, Eva, 1999: Principen och praktiken. En enkätundersökning av lärares syn på bedömning av gymnasieelevers texter. (Svenska i utveckling nr 12. FUMS Rapport nr 195.) Uppsala.