

Förord

Den högre utbildningen har under de senaste decennierna expanderat kraftigt i Sverige. Framför allt har de medelstora och små högskolorna fått större resurser. Få studier som gjorts handlar om kvaliteten i högskoleutbildningen. Är det lättare att bli godkänd vid vissa institutioner eller lärosäten? Moderna Tider gjorde nyligen ett försök att rangordna institutionerna efter kvalitet. Rykten - grundade eller ogrundade - cirkulerar kring var utbildningen är bäst eller var det är enklast att ta sig igenom studierna.

Ur ett bredare demokratiskt perspektiv är detta djupt otillfredsställande. För att det skall vara möjligt att på allvar bredda den sociala rekryteringen till högre studier är det viktigt att alla, inte bara den initierade över- eller medelklassen har tillgång till bättre information.

ESO bad *Kjell Härnqvist*, professor emer. i pedagogik och tidigare rektor för Göteborgs universitet, att ta fram och pröva en metod för att granska kvaliteten i utbildningen i ett par universitetsämnen. Den metod han valt för att granska C-uppsatser beskrivs ingående i denna rapport. Vidare redovisas resultatet av kontrollbedömningen av uppsatser som godkänts eller fått betyget Väl godkänd i nationalekonomi och historia. I undersökningen ingår de nationalekonomiska och historia institutionerna vid gamla universitet som Stockholm, Lund och Uppsala, vid nya universitet som Växjö och Örebro samt vid Mitthögskolan och högskolan i Jönköping. Vidare ingick den nationalekonomiska institutionen vid högskolan i Skövde och den historiska institutionen vid högskolan i Halmstad.

Syftet har i första hand varit att pröva en metod som skulle kunna vara möjlig att använda i större omfattning. Ett motsvarande arbete med att ta fram olika *performance indicators* bedrivs f.n. i

vid *Higher Education Funding Council for England*, på den brittiska regeringens uppdrag. Studien omfattar endast ett mindre antal uppsatser, varför resultaten tyvärr endast redovisas per ämne och högskolekategori.

Jag hoppas att denna rapport leder till en bred och öppen diskussion kring behovet av motsvarande mätningar av kvaliteten i den svenska universitetsutbildningen - inte minst kring behovet av en öppen redovisning av vilka institutioner som står för bästa eller sämsta prestationerna.

Som vanligt i ESO-sammanhang svarar författaren för innehållet i rapporten.

Stockholm i november 1999

Anna Hedborg

Ordförande för ESO

Författarens förord

I en PM den 3 juni 1998 skisserade ESO en undersökning för att belysa examenskraven i några centrala ämnen vid universitet och högskolor. Man tänkte sig att som jämförelsegrund använda de uppsatser som i studieplanerna för många akademiska ämnen ingår som den mest väsentliga delen av tredje terminens studier i ett ämne. Dessa s. k. C-uppsatser motsvarar trebetygsuppsatserna enligt en äldre terminologi och utgör ett krön på studierna för filosofie kandidatexamen.

Den skisserade undersökningen kan ses som en parallell till det arbete för kvalitetssäkring inom högskolan som bedrivs såväl vid enskilda universitet och högskolor som inom Högskoleverket. En annan parallell aktivitet utgör den ranking av högskolor som görs av media med Moderna Tidens undersökning (1999) som mest välkända exempel.

Högskoleverkets kvalitetsarbete har omfattat granskningar dels av olika universitets och högskolors egna åtgärder i kvalitetssäkrande syfte, dels som underlag för prövningar av examensrätt på magisternivå, inrättande av vetenskapsområden och professurer samt regionala högskolors uppflyttning till universitetsnivå (se referenser i avsnitt 3). Dessa granskningar har genomförts av expertgrupper på basis av skriftligt material från högskolorna och vanligtvis också studiebesök.

I ett fall har Högskoleverket låtit ämnesexperter granska C-uppsatser i ämnena engelska och statsvetenskap från olika årgångar som underlag för en bedömning om det över tiden skett förändringar i kvalitet.

Den föreliggande undersökningen kommer närmast den sistnämnda. Den bygger på C-uppsatser från olika universitet och högskolor men här granskade under mer kontrollerade betingelser

och av inbördes oberoende bedömare. Jämförelsen görs inte över tid utan mellan högskolor av olika storlek och status.

Den rapport som följer inleds med en sammanfattning. Den riktar sig till en större läsekrets som vill ta del av de viktigaste resultaten och de resonemang och slutsatser som följer av dessa. Den andra delen är en detaljerad dokumentation skriven ungefär som en beteendevetenskaplig forskningsrapport. Kanske den också kan fungera som en tänkbar modell för eventuella likartade undersökningar, eftersom den inte minst pekar på den arbetsinsats och de komplikationer som en seriös behandling av rankingproblematiken medför.

Samtliga institutioner vid universitet och högskolor med undervisning på C-nivå i historia och nationalekonomi har besvarat en enkät om verksamheten på C-nivå. Sexton av dem har gett underlag till undersökningen i form av listor över C-uppsatser och kopior av de uppsatser som valts ut för bedömning. Sex erfarna universitetslärare från vardera av ämnena historia och statsvetenskap har granskat och bedömt uppsatserna. Utan all denna medverkan från ämnena hade undersökningen inte kunnat genomföras. Ett stort tack för dessa insatser.

Under arbetets gång och vid rapportens utarbetande har jag fått synpunkter av Anders Björklund, Göran Dahlbäck, Robert Erikson och Lars Hultkrantz som ingått i en referensgrupp för projektet. Jag har också fått stöd och hjälp från Martin Hörnqvist, Marja Lemne, Lena Svanberg, Björn Sylvén och Anne Tanskanen vid ESO:s sekretariat samt vid den första insamlingen av uppgifter om institutionerna av Lena Conde. Marja Lemne har lotsat rapporten genom det maskineri som tar vid när ett manus skall förvandlas till en tryckt bok. Även om jag valt att under arbetets gång vara synnerligen hemlighetsfull också mot de närmsta kollegorna vid min gamla institution så har jag fått värdefulla råd av Berit Askling och Jan-Eric Gustafsson och datateknisk hjälp av Åsa Berndtsson. Jag tackar dem alla liksom ESO för att ha blivit anförtrodd detta uppdrag.

Kjell Härnqvist

Innehåll

1	Sammanfattning och slutsatser	13
1.1	Högre utbildning	13
1.2	Att granska kvalitet	13
1.3	Tidigare granskningar	14
1.4	Bedömning	15
1.5	Så här gjordes studien	16
	1.5.1 Uppsatser och handledare	16
	1.5.2 Mål och källor	17
	1.5.3 Tillförlitlighet	17
	1.5.4 Jämförelser mellan kategorier av högskolor	19
	1.5.5 Källornas betydelse	20
	1.5.6 Författarantalets betydelse	21
	1.5.7 Ämne, kön och formalia	21
1.6	Slutsatser	22
	1.6.1 Skillnaderna är stora	22
	1.6.2 Undersökningsmetoden	23
	1.6.3 Metodens användbarhet	24
2	Bakgrund och syfte	27
3	Tidigare undersökningar	31
4	Uppläggnig	39
4.1	Urval av ämnen	39
4.2	Institutionsenkät	40
4.3	Underlag för kvalitetsbedömning	40

4.4	Urvalet av uppsatser	42
4.5	Bedömningskriterier	44
4.6	Bedömare	45
4.7	Kontrollvariabler	45
5	Historia	47
5.1	Institutioner och resurser	47
	5.1.1 Tillströmning och examination	48
	5.1.2Handledning vid uppsatsskrivning	49
5.2	C-uppsatsens syften	50
5.3	Bedömningarnas statistiska egenskaper	51
5.4	Några extremfall	57
	5.4.1 Överensstämmelser	57
	5.4.2 Diskrepanser	58
5.5	Jämförelser mellan kategorier av högskolor	60
5.6	Studentrekryteringen inom olika högskolekategorier	63
5.7	Uppsatsernas källmaterial	68
5.8	Uppsatsämnen	71
5.9	Formella drag hos uppsatserna	71
5.10	Kvinnliga och manliga författare	76
6	Nationalekonomi	77
6.1	Institutioner och resurser	77
	6.1.1 Tillströmning och examination	78
	6.1.2Handledning vid uppsatsskrivning	79
6.2	C-uppsatsens syften	80
6.3	Bedömningarnas statistiska egenskaper	81
6.4	Några extremfall	86
	6.4.1 Överensstämmelser	86
	6.4.2 Diskrepanser	87
6.5	Jämförelser mellan kategorier av högskolor	88
6.6	Studentrekryteringen inom olika högskolekategorier	89
6.7	Uppsatsernas källmaterial	93
6.8	Uppsatsämnen	93

6.9	Formella drag hos uppsatserna	94
6.10	Kvinnliga och manliga författare	96
6.11	En eller två författare	97
7	Diskussion och slutsatser	101
	Referenser	109
Bilaga 1	Instruktion för bedömning av C-uppsatser	111
Bilaga 2	Formulär för bedömning av C-uppsats	113
Bilaga 3	Beräkning av antalet ord	115
	An Academic Question - A Summary	123
	Förteckning över ESO:s rapporter	133

Figurer och tabeller

Figur 5.1	Omdömenas fördelning på olika skalvärden	52
Figur 5.2	Högskolekategoriernas genomsnittsomdömen i förhållande till medelbetygen från gymnasiet	66
Figur 5.3	Högskolornas genomsnittsomdömen i förhållande till genomsnittsbetygen från gymnasiet	67
Figur 5.4	Sambandet mellan referenser och genomsnittsomdöme	74
Figur 5.5	Sambandet mellan fotnoter och omdöme om problemformulering	75
Figur 6.1	Omdömenas fördelning på olika skalvärden	82
Figur 6.2	Högskolekategoriernas genomsnittsomdömen i förhållande till medelbetygen från gymnasiet	91
Figur 6.3	Högskolornas genomsnittsomdömen i förhållande till medelbetygen från gymnasiet	92
Figur 6.4	Sambandet mellan antal ord och omdöme om forskningsanknytning	96
Tabell 4.1	Antal listade uppsatser i historia och nationalekonomi	42
Tabell 5.1	Tillströmning och examination på C-nivå i historia (medeltal)	48
Tabell 5.2	Antalet handledare och deras befattningsnivå	49
Tabell 5.3	C-uppsatsernas syften (uppställda i rangordning)	51
Tabell 5.4	Medeltal och spridning för omdömena om uppsatser i historia samt samstämmigheten inom bedömarparen	53
Tabell 5.5	Samstämmigheten inom bedömarparen i fråga om totalomdömet om uppsatser i historia	54
Tabell 5.6	Korrelationer mellan olika bedömningsaspekter	55
Tabell 5.7	Uppsatsbetygen och deras samband med omdömena	56
Tabell 5.8	Jämförelse mellan olika kategorier av högskolor i historia (efter korrektion för betygsskillnader)	61
Tabell 5.9	Jämförelse av <i>eta</i> -värden för två olika uppsättningar av uppsatser	63
Tabell 5.10	Genomsnittliga avgångsbetyg från grundskola och gymnasium för studerande registrerade på B- och C-nivå i historia inom olika kategorier av högskolor	64
Tabell 5.11	Källmaterial för uppsatser i historia	68

Tabell 5.12	Jämförelse mellan uppsatser i historia baserade på olika slags källor	69
Tabell 5.13	Tvåvägsvariensanalyser av högskolekategori x källa förbedömare 1 och 2 var för sig (p-värden)	70
Tabell 5.14	Medeltal och spridning för formella drag hos uppsatserna i historia	72
Tabell 5.15	Korrelationen mellan formella drag och uppsatsernas kvalitet	73
Tabell 6.1	Tillströmning och examination på C-nivå i nationalekonomi (medeltal)	78
Tabell 6.2	Antalet handledare och deras befattningsnivå	79
Tabell 6.3	C-uppsatsens syften (uppställda i rangordning)	81
Tabell 6.4	Medeltal och spridning för omdömena om uppsatser i nationalekonomi samt samstämmigheten inom bedömarparen	83
Tabell 6.5	Samstämmigheten inom bedömarparen i fråga om totalomdömet om uppsatser i nationalekonomi	84
Tabell 6.6	Korrelationer mellan olika bedömningsaspekter	84
Tabell 6.7	Uppsatsbetygen och deras samband med omdömena	85
Tabell 6.8	Jämförelse mellan olika kategorier av högskolor i nationalekonomi (efter korrektion för betygs-skillnader)	89
Tabell 6.9	Genomsnittliga avgångsbetyg från grundskola och gymnasium för studerande registrerade på B- och C-nivå i nationalekonomi inom olika kategorier av högskolor	90
Tabell 6.10	Källmaterial för uppsatser i nationalekonomi	93
Tabell 6.11	Medeltal och spridning för formella drag hos uppsatserna i nationalekonomi	94
Tabell 6.12	Korrelationen mellan formella drag och uppsatsernas kvalitet enligt genomsnittet för två bedömare	95
Tabell 6.13	Jämförelse mellan uppsatser i nationalekonomi med manliga och kvinnliga författare	97
Tabell 6.14	Jämförelse mellan uppsatser i nationalekonomi med en respektive två författare	98
Tabell 6.15	Tvåvägsvariensanalys av högskolekategori och antal författare (p-värden)	99
Tabell 6.16	Högskolekategori i kombination med antal författare. Medeltal i variabler med signifikanta samspelseffekter	99

1 Sammanfattning och slutsatser

1.1 Högre utbildning

Högskoleutbildningen i Sverige har under 1990-talet expanderat snabbt. En allt större andel av de nya resurserna har tillförts regionala högskolor utan universitetsstatus. Tre högskolor blev 1999 universitet, andra fick fasta forskningsresurser i enstaka ämnen och/eller magisterutbildningar med en fjärde studietermi i vissa ämnen.

Högskoleverket har medverkat i denna process genom att granska högskolornas förutsättningar för undervisning och forskning på nivåer över 40 poäng och genom att följa upp beslutade åtgärder, främst när det gäller magisterutbildningarna.

Högskolorna har också granskats i media, främst i Moderna Tider, som genom att väga samman synpunkter på deras anseende, attraktivitet, resurser och examination gjorde en rangordning.

1.2 Att granska kvalitet

Dessa tidigare granskningar har, med undantag för uppgifterna om examination, avsett inputsidan av utbildningen i form av resurser och studerande. Inte i något fall har man gjort systematiska studier av utbildningens output i form av de studerandes prestationer efter avslutad utbildning. I all synnerhet saknas undersökningar av högskolans nettobidrag, mellanskillnaden mellan input och output.

Syftet med denna undersökning är:

- Att med C-uppsatser som underlag jämföra grundutbildningens resultat vid högskolor av olika storlek och status.
- Att basera dess jämförelser av uppsatsernas kvalitet på externa bedömningar som utförts under experimentellt kontrollerade betingelser.
- Att vid utvärderingen av resultaten beakta ett antal faktorer som kan antas påverka kvaliteten, däribland studentunderlagets sammansättning.
- Att bedöma om metoden går att tillämpa i ett program för högskolans kvalitetskontroll.

Arbetet på C-uppsatsen motsvarar halva studietiden under den tredje termin (från 40 till 60 poäng) som ett ämne läses. 60 poäng i ett ämne är numera ett av villkoren för fil.kand.-examen. Därför kan man betrakta uppsatsen som ett krön på studierna för denna examen och dess kvalitet som ett lämpligt mått på de studerandes prestationer.

I första hand belyses bruttoskillnader i output, men även nettoeffekter beskrivs genom att resultaten relateras till de studerandes förutsättningar att tillgodogöra sig högskoleutbildning.

1.3 Tidigare granskningar

Förutom till Högskoleverkets granskningar och Moderna Tidens rangordning (se kapitel 3) hänvisas till två studier av C-uppsatser, en i statsvetenskap och en i statsvetenskap och engelska.

Paralleller dras till hur censorer medverkar i examinationen vid danska och norska universitet och examinatorer från andra lärosäten vid brittiska universitet. En likartad försöksverksamhet i Sverige finansieras av Högskoleverket.

Det har visat sig svårt att i den internationella forskningslitteraturen finna studier av det här slaget. Mellan bedömare av uppsatser med förelagda ämnen, typ skoluppsatser, är samstämmigheten vanligtvis, efter särskild träning av bedömarna, relativt god. Mellan granskare av artiklar för publicering i vetenskapliga tidskrifter är samstämmigheten troligen låg.

1.4 Bedömning

Bedömarna har granskat uppsatser i historia och nationalekonomi, två ämnen som intar en central ställning inom sina fakulteter. Båda ämnena är representerade såväl vid stora universitet som små högskolor. Urvalet, 30 uppsatser från vardera ämnet, 15 från de gamla universiteten, 6 från de nya universiteten och 9 från högskolorna, gjordes med hjälp av slumpstal ur listor över examinerade vårterminen 1996 – vårterminen 1998.

I ett förberedande steg gjordes en enkät vid samtliga institutioner i landet i historia och nationalekonomi om C-undervisningens omfattning. C-uppsatser samlades in från tre kategorier av högskolor:

1. Gamla universitet (Uppsala, Lund, Stockholm)
2. Nya universitet (Växjö, Örebro)
3. Högskolor (Mithögskolan och Jönköping i båda ämnena, Halmstad i historia, Skövde i nationalekonomi).

För vardera ämnet anlätades sex bedömare från högskolor som inte ingick i undersökningen. De delades in i tre par med en bedömare från institution med egen forskarutbildning och en bedömare från institution utan fasta resurser för forskning och forskarutbildning. Uppsatserna fördelades slumpmässigt på bedömarparen. Författar- och institutionsnamn togs bort från titelbladen för att bedömarna inte skulle känna till författarens kön, var uppsatsen

skrivits och om den hade en eller två författare (vilket ofta förekom i nationalekonomi).

Varje uppsats bedömdes i fråga om forskningsanknytning, problemformulering, teoretisk medvetenhet, metod, genomförande, slutsatser, språklig utformning och formalia på en skala från 1 till 5. Även uppsatsens källmaterial klassificerades och uppsatsämnets svårighetsgrad bedömdes (se bilagorna 1 och 2).

1.5 Så här gjordes studien

1.5.1 Uppsatser och handledare

Det förelåg betydande skillnader mellan de båda ämnena i fråga om studerandeantal, handledarresurser och värderingar. Institutionenkäten i historia visade att det under fem terminer hade godkänts mer än 700 C-uppsatser vid de gamla universiteten, cirka 130 vid de nya universiteten och cirka 80 vid högskolorna. Av 35 handledare var 4 professorer och 29 docenter eller universitetslektorer. I genomsnitt blir det 26 uppsatser per handledare över en period på fem terminer, 5 uppsatser per termin. Varje handledare fick tillgodoräkna sig två eller tre undervisningstimmar per uppsats vid de stora universitetsinstitutionerna – och sannolikt mer vid högskolorna. I nationalekonomi godkändes 280 uppsatser vid de gamla universiteten, 50 vid de nya och 30 vid högskolorna. Av 75 handledare var 21 professorer, 47 docenter eller universitetslektorer och 7 var personer utan doktorsexamen. I genomsnitt blir det knappt 5 uppsatser per handledare eller en enda under varje termin vid jämn fördelning mellan handledarna, men troligen har fördelningen varit mycket mera ojämn. Där uppgift har lämnats har varje handledare tillgodoräknats fem undervisningstimmar per uppsats eller mera.

1.5.2 Mål och källor

När institutionerna i en skala från +1 till -1 fick gradera ett antal syften med C-uppsatsen uppträdde intressanta skillnader.

Viktigast i historia:

- Bidra till kunskapsutvecklingen
(differens mellan ämnena +0,56)
- Förstå innebörden av forskning (+0.31)
- Lära sig arbeta självständigt (+0.25)
- Utveckla intresse för forskning (+0.21)
- Utveckla kritiskt tänkande (+0.15)

Viktigast i nationalekonomi:

- Lära sig samarbeta med andra (+0.65)
- Tillämpa kunskaper från tidigare utbildning (+0.38)
- Få möjlighet att fördjupa sig inom något eget valt område (+0.19)
- Lära sig söka, läsa, värdera och sammanställa information (+0.14)

Skillnaderna är i allmänhet inte stora, men det totala intrycket är att ämnena ser olika på C-uppsatsens roll – i historia är den en förberedelse för forskning, i nationalekonomi en fördjupning. Dessa skillnader slår igenom också i valet av uppsatsämnen och i bedömarnas värderingar. Medan en stor andel av uppsatserna i historia baseras på primärkällor, hämtar uppsatserna i nationalekonomi sitt material främst ur sekundärkällor.

1.5.3 Tillförlitlighet

Bedömningarna av uppsatsernas kvalitet avgavs i en femgradig skala.

Uppsatserna i historia fick för kvalitet omdömet 2.84 i genomsnitt med en spridning på 1.15. De lägsta värdena fick de för teore-

tisk medvetenhet och metod, de högsta för forskningsanknytning, språklig utformning och formalia.

Också i nationalekonomi låg genomsnittsomdömet för kvalitet nära tre och spridningen var drygt ett skalsteg. Fördelningen över skalan var emellertid skev med relativt få femmor och ett överskott av fyror. De lägsta medeltalen gällde totalomdömet samt språklig utformning och formalia. Högst låg omdömena om problemformulering och teoretisk medvetenhet – nära nog omvänt resultatet i historia.

Samstämmigheten mellan bedömarna varierade starkt. I historia uppgick korrelationen som mest till 0.56 (i skalan 0–1) för genomsnittsomdömet, tämligen ordinärt för bedömningar av detta slag. Minst eniga var bedömarna om den språkliga utformningen (0.29) och ämnets svårighetsgrad (0.15). Differensen mellan de två bedömarna i varje par varierade kring ett steg på den 5-gradiga skalan, bara extremfallen hade en differens på två eller t. o. m. tre skalsteg. De vid institutionen satta uppsatsbetygen hade korrelationen 0.48 med genomsnittsomdömet från de externa bedömarna – således mindre än mellan bedömarna 1 och 2. Mer än hälften av uppsatserna hade fått betyget Väl godkänd vilket tyder på en överrepresentation av VG-uppsatser, särskilt från universiteten.

I nationalekonomi var samstämmigheten mellan bedömarna störst för teoretisk medvetenhet (0.60) och genomsnittsomdömet (0.58). Helt oeniga var bedömarna om svårighetsgraden. Differensen mellan bedömarna i varje par varierade även här runt 1 skalsteg men kunde i extremfall uppgå till 3. Få uppsatser hade fått betyget Väl godkänd och korrelationen med den externa bedömningen var lägre än i historia, som mest 0.38.

Kommentarer från bedömarna har varit till hjälp vid studiet av de två extremfallen – där bedömarna är helt eniga om totalomdömena 1 respektive 5 och där totalomdömena skiljer sig minst två skalsteg.

I historia blev två uppsatser samstämmigt underkända på grund av ett okritiskt förhållande till tryckta källor och i ett fall också underlåtenhet att utnyttja primärkällor. En uppsats fick samstäm-

mycket Mycket väl godkänd. Nio uppsatser uppvisade skillnader på två eller tre skalsteg.

I nationalekonomi underkändes tre uppsatser samstämmigt och ingen fick Mycket väl godkänd av båda bedömarna. Fyra uppsatser fick totalomdömen som skilde sig två eller tre skalsteg. De samstämmigt underkända uppsatserna föll på bristande struktur och dåligt språk i referaten av tryckta källor.

Även när det gäller uppsatser med stora skillnader i bedömningen spelar i nationalekonomi språket och förhållandet till sekundärkällor en stor roll. I historia hänvisades oftare till brister i vetenskaplighet vilka bedömdes olika strängt av olika bedömare.

1.5.4 Jämförelser mellan kategorier av högskolor

I historia var kvaliteten högre i uppsatserna från de gamla och nya universiteten än i uppsatserna från högskolorna. Mellan de gamla universiteten och de nya var skillnaden obetydlig och de turades om att ligga högst. Som högst låg den bästa ett drygt skalsteg högre än högskolorna. Värdena varierade mellan 1.32 skalsteg för problemformulering till 0.77 skalsteg för språk och formalia med ett medeltal av 1.06 och var nästan överlag statistiskt signifikanta. Som kontrast till kvalitetsomdömena var skattningen av uppsatsämnets svårighetsgrad lika på andra decimalen när i alla tre kategorierna.

I nationalekonomi fick de nya universiteten lägst omdömen även om skillnaderna mellan kategorierna inte var signifikanta. Som mest uppgick skillnaden mellan högsta och lägsta kategori till 1.21 för problemformulering.

För att försöka förklara dessa skillnader undersöktes vilken roll studentrekryteringen till olika högskolor spelade. Att det fanns skillnader kunde betraktas som givet och visar sig bl. a. i de återkommande listorna över söktryck och lägsta poängtal för tillträde vid den centrala antagningen. Denna undersökning har mer precisa mått att jämföra med, hämtade ur en stor oidentifierad databas vid pedagogiska institutionen i Göteborg. Där kan man för alla födda

åren 1972–1979 få uppgift om medelbetyget vid avgången från grundskolan respektive gymnasiet för alla som registrerats för studier på B- och C-nivå i historia och nationalekonomi under vissa läsår. Uppgiften om medelbetyget i gymnasiet för dem som under läsåren 1995/96 och 1996/97 registrerats på B- och C-nivå vid olika högskolor togs som ett kollektivt mått på studenternas förkunskaper vid övergången till högre utbildning och jämfördes med de omdömen som lämnats om uppsatserna från respektive högskolor. Eftersom varje högskola bidragit med högst fem uppsatser får resultaten tolkas med försiktighet.

Om de externa bedömarnas kvalitetsomdömen i historia (jämför figur 5.3) ställs mot gymnasiebetygen ligger de gamla universiteten (med ett undantag) högst på båda axlarna, därefter följer de nya och sist högskolorna. Ett universitet ligger dock långt under denna tänkta kurva trots att institutionen satt betyget Väl godkänd på fyra av de fem bedömda uppsatserna. Detta tyder på en markant avvikelse från den norm som styr de externa bedömarna och från den nivå som studentgruppens sammansättning gett anledning att förvänta.

Även i nationalekonomi har medelbetygen från gymnasiet använts. Resultaten visas i figur 6.3. För sex av de åtta högskolorna följer genomsnittsomdömena medelbetygen från gymnasiet. Högst i båda variablerna kommer två av de gamla universiteten, därefter två av högskolorna och sist de två nya universiteten. Men ett av de gamla universiteten ligger avsevärt under vad gymnasiebetygen gett anledning att förvänta och en av högskolorna ligger avsevärt över.

1.5.5 Källornas betydelse

En viktig förklaring till att omdömena varierar visade sig för historia vara att uppsatser från de gamla och nya universiteten oftare utnyttjade primärkällor, vilket gav mer än ett skalsteg högre omdömen för allt utom för språk och svårighetsgrad. Vid högskolorna dominerade sekundärkällor. Bakom denna olikhet kan ligga skillnader i såväl synen på C-uppsatsens syfte som tillgången till käll-

material. Universiteten torde oftare se uppsatsen som ett underlag för rekryteringen till forskarutbildning och de befinner sig närmre arkiv som kan leverera källmaterial.

Detta styrktes när de båda indelningsgrunderna högskolekategori och typ av källor kombinerades och dessutom gjordes för bedömare 1 och 2 var för sig. Då visade det sig att valet av källor var viktigare för bedömare 1, som kom från universitetsinstitutioner, och kategori av högskola viktigare för bedömare 2, som kom från institutioner utan fasta resurser för forskning och forskarutbildning. Den samlade effekten därav blev dock för båda typerna av bedömare att uppsatserna från högskolorna fick lägre kvalitetsomdömen.

1.5.6 Författarantalets betydelse

Eftersom endast två uppsatser i nationalekonomi baserades på primärkällor kunde uppsatsernas källmaterial där inte användas som förklaringsgrund. Däremot visade sig resultaten i hög grad sammanhånga med antalet författare. Störst var skillnaden till förmån för uppsatser med två författare i språklig utformning, teorimedvetenhet och totalomdöme samt i institutionens betyg på uppsatsen. Endast en av 16 uppsatser med en författare hade fått VG mot sex av 14 med två författare.

När högskolekategori och författarantal analyserades gemensamt visade sig författarantalet vara den mest utslagsgivande. Effekten av dubbelt författarskap gällde dock i första hand vid de nya universiteten och vid högskolorna medan den inte gav utslag vid de gamla universitetsinstitutionerna.

1.5.7 Ämne, kön och formalia

Flera ytterligare faktorer har prövats, bl. a. olika kategorier av uppsatsämnen, författarnas kön och formella drag hos uppsatserna.

Vad gäller formella drag framkom några skillnader som kan vara av intresse. Uppsatserna i historia var längre och hade fler fotnoter och referenser, medan uppsatserna i nationalekonomi innehöll fler figurer och tabeller. I historia var referensernas antal positivt korrelerat med de flesta kvalitetsomdömen och med uppsatsernas längd. I nationalekonomi var antalet tabeller och figurer mest positivt korrelerat med omdömena. Beträffande dessa resultat varnas för en alltför enkel tolkning: att uppsatsförfattarna spekulat i högre betyg genom att skriva långt, använda fler referenser och tabeller, eller att bedömnarna styrts av sådana formalia.

1.6 Slutsatser

1.6.1 Skillnaderna är stora

Undersökningen ger stöd för uppfattningen att det finns en skillnad i kvaliteten på C-uppsatser mellan högskolor av olika storlek och status. I historia har uppsatserna från högskolorna fått de lägsta omdömena. I nationalekonomi är det de nya universiteten som ligger lägst men skillnaderna är ur statistisk synpunkt mindre säkra.

Det finns avvikelser såväl uppåt som nedåt.Handledarnas kompetens och engagemang, ämnets ställning inom högskolan, C-nivåns status och tillgång till underlag är säkerligen av stor betydelse. Efter att studenternas ingångskvalifikationer beaktats är det på de variationerna som utsagor om högskolornas kvalitet bör grundas.

Det finns anledning att tro att de resultat som uppnåtts här också skulle kunna avse andra ämnen och andra högskolor. Betydelsen av studentgruppens sammansättning torde gälla inte bara inom de flesta ämnen utan också mellan linjer och ämnen som skiljer sig i attraktivitet och är utsatta för olika söktryck. Alltför ofta bortser man i diskussionen från den viktiga faktor som skillnader i förkunskaper och studiekapacitet utgör.

Resultaten tillåter också vissa pedagogiska slutsatser. Det är uppenbart att C-uppsatserna spelar olika roller inom historia och nationalekonomi – som forskningsförberedelse eller fördjupning. Detta understryks ytterligare av att historieinstitutioner med ansvar för forskarutbildning värderar uppsatser som bygger på primärkällor högre. Det finns skäl att institutionerna inom ämnet uppmärksammar den betydelse som valet av källor har för bedömningen av uppsatsernas kvalitet. Förutsatt att forskningsförberedelse accepteras som mål bör institutionerna om möjligt underlätta för de studerande att få tillgång till primärkällor.

I nationalekonomi har en tredjedel av uppsatserna mer än en författare, men variationerna mellan institutionerna är stora. Både externa och interna bedömare värderar uppsatser med flera författare högre än ensamuppsatser. Examinatorerna visar heller ingen tendens att korrigera betygen nedåt för dubbelt författarskap. Att uppsatserna blir bättre med två författare beror sannolikt på att studenterna nalkas uppgiften med olika perspektiv och arbetar på den under fortlöpande diskussion. Detta är viktiga komplement till lärarens handledning. Det kan å andra sidan försämra det individuella prognosvärdet av uppsatsen, men om fördjupning snarare än forskningsförberedelse är målet så kan svårigheten att urskilja individuella insatser vara väl värd att ta. För övrigt ligger sådant samarbete i linje med den färdigutbildade forskarens arbetssituation. Emellertid finns det skäl att mellan institutionerna i ämnet diskutera den praxis som utvecklats.

1.6.2 Undersökningsmetoden

C-uppsatser kan tyckas vara en smal bas som underlag för värderingen av grundutbildningens resultat, men bristen på alternativa prestationer som är åtkomliga för enhetlig bedömning är inte det enda skälet till detta val. Uppsatsskrivandet integrerar en rad ämnesanknutna kunskaper och metodologiska färdigheter med sådant som tillhör högskolestudiernas allmänna mål beträffande självständigt arbete, kritisk förmåga och effektiv kommunikation. Inget av

detta kan i någon strängare mening värderas objektivt utan värderingen måste ske med hjälp av ämnesförtrogna med erfarenhet av uppsatsbedömning. Valet av bedömare är därför avgörande och kan inte lösas enligt något enkelt stickprovsförfarande utan får göras så att olika uppsatskulturer blir företrädda. Resultaten får utvisa i vilken grad samstämmighet har uppnåtts och vad som har skapat mest oenighet.

Hörnstenen i den här metoden är att bedömningarna görs oberoende av varandra. Bedömarna skall vara minst två och ovetande om varandra. De skall grunda sina bedömningar på själva uppsatsen utan att påverkas av några förhandsinställningar till de institutioner uppsatserna kommer från, författarnas kön eller antalet författare till varje uppsats. Uppsatserna skall fördelas så att alla bedömare får ett likartat urval av uppsatser från samtliga undersökta högskolor. Dessa villkor är i detta fall väl uppfyllda.

1.6.3 Metodens användbarhet

Den metod som använts i denna undersökning bör kunna tillämpas på andra ämnen än historia och nationalekonomi och även för D-nivån och för licentiatexamen och doktorsexamen.

Man får inom varje grupp av ämnen eller institutioner räkna med olika uppsatskulturer när det gäller organisationen av ämnesval, handledning, examination och uppsatsernas betydelse för rekryteringen till forskarutbildning. Förhållandena inom ett ämne kan också förändras över tiden och när studerandeantalet ökar kraftigt.

Men metoden kan också användas i ett större rangordningsförfarande, som kan och bör omfatta en rad ytterligare kriterier som prövats i andra sammanhang.

Metoden kräver konsekvens och en fast struktur både när en undersökning planeras och när den genomförs. Den kräver medverkan från såväl de institutioner, som levererar underlag för bedömning, som de som ställer upp med ämnessakkunniga som utför bedömningarna. Metoden är relativt billig eftersom huvuddelen av kostnaderna består av arvoden till bedömarna, däremot blir det inga

kostnader för resor eftersom granskarna inte behöver resa omkring. Man får räkna med att insamling av uppsatser, rekrytering av bedömare och själva bedömningen tar tid. För att få fullt utbyte av det insamlade materialet krävs en ingående resultatanalys, som dock i föreliggande fall blivit mer tidskrävande än som är nödvändigt vid en upprepning.

Metoden lämpar sig för punktvisa insatser som kontroller när det kommer signaler om att bedömningen är ojämn. Det är just kontroller, både när det gäller omdömenas frihet från systematiska fel och studentgruppens sammansättning, som gör metoden användbar. Därigenom kan den fungera som en måttstock för andra försök till rangordning.

2 Bakgrund och syfte

Uppsatsskrivandet och seminariebehandlingen av den färdiga uppsatsen är universitetsutbildningens kärna. I uppsatsarbetet och i seminariesituationen skall de enskilda studenterna visa att de förmått tillägna sig det metodologiska kunnande och det analytiska tillvägagångssätt som kännetecknar vetenskaplig verksamhet. Genom att ge de studerande möjlighet att i självständiga undersökningar bearbeta vetenskapliga problemställningar blir uppsatsarbetet den tydligaste skiljelinjen mellan universitetsutbildningen och gymnasieskolan.

Med denna programförklaring inleder en grupp av statsvetare i Göteborg och Uppsala sin rapportering av erfarenheter från ett kvalitetssäkringsprogram som fokuserat just "den goda C-uppsatsen" (Bjereld m.fl., 1996).

Samma bedömning ligger bakom Högskoleverkets jämförelse av nivån på C-uppsatser i engelska och statsvetenskap skrivna med tio års mellanrum (Högskoleverket, 1997:8).

När ESO initierade en granskning av kvaliteten på grundutbildningen vid högskolor av olika storlek och status var det naturligt att använda C-uppsatserna som underlag för en sådan bedömning. Detta förfarande kompliceras av en rad samverkande faktorer som påverkar uppsatsernas kvalitet. Å andra sidan finns det – i avsaknad av centrala prov eller externa examinationer – ingen annan jämförbar skriftlig dokumentation av de studerandes prestationer. Och vilka prov som än hade funnits hade resultaten påverkats av ungefär samma uppsättning av faktorer.

Med en grov indelning kan man bakom (uppsats)kvaliteten vid en viss högskola räkna med inflytanden från bl.a.

- den begåvnings- och kunskapsmässiga standarden hos de studenter som söker sig till ämnet och fortsätter till C-nivå
- den utbildning de erhållit på tidigare nivåer i ämnet och inom grundutbildningen över huvud taget
- den kunskapssyn och de normer som präglar ämnet och bestämmer vad som är viktigt för god kvalitet
- tillgången till källmaterial samt biblioteksmässiga och tekniska resurser hos institutionen
- den undervisning och handledning de får under (uppsats)arbetet vilken kan variera både i kompetens, tidsinsats och engagemang hos lärare/handledare.
- formerna för undervisning, seminarieverksamhet och examination.

Dessa och andra faktorer bidrar till att bestämma uppsatsernas kvalitet. Men utfallet av en *undersökning av* kvaliteten påverkas också av andra komplikationer, t.ex. valet av

- ämne
- institutioner
- bedömningskriterier
- bedömare
- uppsatser som är representativa för de enheter som skall jämföras.

Medan den första gruppen (elevstandard etc.) utgörs av sådana faktorer som erbjuder alternativa tolkningar av jämförelsernas resultat, avser den senare gruppen sådant som åtminstone hjälpligt kan kontrolleras genom experimentella arrangemang. Det är framför allt den senare typen av faktorer som skall diskuteras i kapitel 4 om undersökningens upplägning.

Syftet med undersökningen kan sammanfattas som

- att jämföra grundutbildningens resultat vid högskolor av olika storlek och status med C-uppsatser som underlag,
- att basera dessa jämförelser på externa bedömningar av uppsatsernas kvalitet som utförts under experimentellt kontrollerade betingelser,
- att vid utvärderingen av resultaten beakta ett antal alternativa faktorer som kan antas påverka uppsatsernas kvalitet, däribland studerandeunderlagets sammansättning samt
- att göra en bedömning av den använda metodikens tillämpbarhet som del av ett program för högskolans kvalitetskontroll.

Allmänt gäller naturligtvis det förbehållet att resultaten endast avser de universitetsämnen som C-uppsatserna kommer från. Generalisering därutöver kan resonemangsvis göras till närliggande ämnen men inte till hela högskolor av en viss storlek och status.

3 Tidigare undersökningar

Redan inledningsvis har två svenska undersökningar av C-uppsatser nämnts. Den första av dem, *Den goda C-uppsatsen* (Bjereld m.fl., 1996), redovisar ett kvalitetssäkringsprogram i samverkan mellan institutionerna för statsvetenskap i Göteborg och Uppsala. Förutom uppgifter om studerandeantal och genomströmning innehåller rapporten ett bedömningsexperiment där tre uppsatser graderats på en 9-gradig skala av tre bedömare från Göteborg och fem bedömare från Uppsala. I genomsnitt placeras uppsatserna i samma rangordning både av Göteborgs- och Uppsala-bedömare, men skillnaderna i poängtal är betydande. En mera realistisk experimentsituation med samma totala arbetsinsats skulle ha varit att ta en bedömare från vardera institutionen och låta dem oberoende av varandra bedöma 12 uppsatser.

I den andra undersökningen bedömdes C-uppsatser i engelska från Göteborgs och Uppsala universitet samt C-uppsatser i statsvetenskap från Stockholms och Umeå universitet (Högskoleverket, 1997:8R). Cirka 30 uppsatser hämtades från var och en av de fyra institutionerna. Ungefär en tredjedel av uppsatserna kom från mitten av 1970-talet, en tredjedel från mitten av 1980-talet och resten från mitten av 1990-talet. Avsikten var att belysa om innehåll och kvalitet förändrats mellan åren.

För de tidigare åren är det i stort sett samtliga uppsatser under en period som bedömts, medan det för 1990-talet är ett av vederbörande institution gjort urval. Samtliga uppsatser från en institution bedömdes av en och samma lärare från en annan institution inom ämnet. Dubbla bedömare förekom således inte. Varje bedömare tillämpade sina egna principer för bedömningen. Redovisningen var av resonerande karaktär i fråga om engelska. I statsve-

tenskap var den delvis baserad på i poängtal uttryckta omdömen. Även om metodiken för jämförelser inte är invändningsfri ger rapporten en intressant bild av de olika ämnena.

I författarens förord hänvisades till Högskoleverkets och de enskilda högskolornas arbete för kvalitetssäkring av den högre utbildningen. Den del av detta arbete som ligger närmast föreliggande undersökning är prövningen av rätt att ge magisterexamen för vilken godkända prov på D-nivån i ett ämne krävs. Sådana prövningar har ägt rum sedan 1993 och redovisats i ett stort antal rapporter från Högskoleverket. Vid prövningen har externa sakkunniga anlåtats.

De kriterier som de sakkunniga använt vid prövningen är lärarkompetens och kompetensutveckling, forskningsaktivitet, ämnesdjup och ämnesbredd – miljö, utbildningsaktivitet över 40-poängsnivån, bibliotek och litteraturförsörjning, lokaler och utrustning, internationalisering, behörighet och övergångsmöjligheter till forskarutbildning. (Högskoleverket, 1997:38R, s. 6)

I besluten om tillstånd att utfärda magisterexamen i ett ämne sades att Högskoleverket hade för avsikt att efter ca tre år följa upp sina beslut. Sådana uppföljningar har genomförts och publicerats i Högskoleverkets rapportserie (1997:40R, 1997:45R och 1999:3R).

Den första av dessa rapporter avser bl.a. magisterutbildningen i historia och nationalekonomi vid de dåvarande högskolorna i Växjö och Örebro. Behandlingen är ganska summarisk jämfört med senare rapporter och fokuserar mera på resurser än resultat.

Den senaste av rapporterna granskar bl.a. magisterutbildningen i historia i Göteborg, Växjö, Linköping och vid Mitthögskolan (Sundsvall, Härnösand och Östersund). Denna rapport ger en inträngande analys av förutsättningar och resultat, inte minst när det gäller uppsatsarbetet på D-nivå. De sakkunniga finner i samtliga fall att de uppsatser de läst är godtagbara, men någon detaljerad redovisning ges inte. Uppföljningsrapporter av denna art och kvalitet förtjänar stor uppmärksamhet även inom ämnen och institutioner som inte själva varit föremål för granskning.

Ett steg längre bort från uppsatsbedömningen ligger Högskoleverkets granskningar av rätten att inrätta professorer (1997:37R), rätten att bedriva forskarutbildning inom visst vetenskapsområde (1998:27R) och förutsättningarna för att bli universitet (1998:11R).

Mest uppmärksammat och omdiskuterat bland svenska försök att beskriva nivån hos olika högskolor är tidskriften *Moderna Tiders "ranking"* (Berg, 1999). Den rapporterade en poängbedömning inom som mest 20 ämnen, däribland historia, vid 23 universitet och högskolor. Totalomdömet byggde på fyra huvudkategorier:

- *status* i fråga om grundutbildning och forskning enligt bedömningar gjorda av lärare i ifrågavarande ämne
- *attraktion* mätt genom söktryck och antagningskrav
- *resurser* (pedagogiska resurser och akademisk nivå)
- *prestation* (poängproduktion inom grundutbildning och disputationer i förhållande till antalet doktorander).

Någon jämförande kvalitetsbedömning av produkter i form av uppsatser och avhandlingar ingår inte.

Universitetskanslern har som motdrag initierat en s.k. *franking* avseende vissa övergripande drag såsom jämställdhet, studentinflytande samt social och etnisk mångfald (Universitetskanslerns skrivelse 1994-04-20). De olika högskolorna har fått i uppdrag att beskriva sina åtgärder i dessa avseenden, och de kommer också att besökas av en expertgrupp för ytterligare diskussion.

En avhandling från pedagogiska institutionen i Lund (Gerrevall, 1992) handlar om de högskolestuderandes erfarenheter av självständigt arbete i projektform inom läkarutbildning, lärarutbildning och ekonomutbildning. Avhandlingen belyser på ett intressant sätt de syften som projektarbetet kan ha för de studerande men innehåller inte heller den någon bedömning av färdiga produkter.

En ytterligare svensk studie som berör den akademiska uppsatskrivningen är gjord av ekonomerna Curt Löfgren och Henry Ohlsson (1999). I deras artikel utvecklas en närmast ekonomisk mo-

dell för att med hjälp av studenternas bakgrund, studieprogram och typ av uppsats beräkna det sannolika antalet studerande som fullföljer uppsatsskrivandet inom viss tid, vanligen flera terminer mer än vad studieplanen anger.

I Sverige ligger examinationen för högskoleexamina sedan gammalt hos den enskilda läraren och några krav på insyn från andra universitet eller ämnen föreligger inte utom för doktorsexamen med dess föreskrifter om betygsnämnd. Detta skiljer sig radikalt från förhållandena i Danmark och Norge, som anlitar särskilt utsedda censorer vid examinationen, och från Storbritannien, som har *external examiners* (Nilsson, 1997). Till skillnad från Sverige finns i dessa länder en fingraderad betygsättning som spelar en stor roll för examinas brukbarhet på arbetsmarknaden. De externa bedömare har den dubbla uppgiften att dels garantera studenternas rättssäkerhet i examina, dels svara för upprätthållandet av en enhetlig och kvalitativt tillfredsställande standard.

En avsevärd litteratur om det brittiska systemet föreligger men innefattar knappast några systematiska kontroller av samstämmigheten mellan lokala och externa bedömare. Den svenska högskoleutredningen i början av 1990-talet föreslog att extern medverkan i examinationen skulle införas även i Sverige (SOU 1992:1). Så har ej blivit fallet men Högskoleverket stöder försöksverksamhet med samordning av examinationen mellan olika högskolor inom ett antal olika utbildningsområden. Vanligen är det examensarbeten som granskas med hjälp av externa bedömare (Nilsson, 1997).

En besläktad verksamhet är benchmarking en kvalitetskontroll med ursprung i industriella och kommersiella sammanhang, som har prövats även inom högre utbildning. Yorke m.fl. (1999) ger exempel från en brittisk undersökning där man studerade jämförbarheten mellan syftena inom till namnet likartade utbildningsprogram och kurser, mellan programmens examenskrav och examinationens utfall. Slutsatserna förefaller väl värda att citera.

Academic standards are underpinned by complexity in terms of academic purposes, curricular content, assessment practices, and regulatory frameworks for awards. In any diverse higher education system what is needed is not a

simplistic set of statements about academic standards, since the evidence presented above is probably sufficient to raise doubts about the viability of such an approach. Rather, benchmarking should be seen as a method through which academic standards can be clarified, and as a result of which institutions will better be able to articulate the standards to which they are prepared to attach their imprimatours. (a.a., s. 93)

En kommitté tillsatt av Higher Education Funding Council for England har nyligen publicerat en rapport *om performance indicators* (1999). Den inventerar olika tänkbara indikatorer för såväl utbildning som forskning. På utbildningssidan bygger flertalet indikatorer på kvantitativa uppgifter om deltagande i högre utbildning, genomströmning och de examinerade inträde på arbetsmarknaden. Ett program för utveckling av "benchmarks" för tröskelvärden för godkännande och högre nivåer i olika ämnen skisseras, men systemet beräknas inte vara klart för avrapportering förrän tidigast år 2006.

I forskningslitteraturen är det svårt att finna systematiska undersökningar av uppsatsförfattande och den känsliga frågan om samstämmighet mellan olika uppsatsbedömare. En relativt omfattande datorsökning i såväl ERIC (databas för utbildningsforskning) som SSCI (databas för samhällsvetenskap) har gett få referenser, i varje fall när det gäller uppsats- och avhandlingsbedömning inom högre utbildning. Sökord såsom "rating", "marking", "scoring", "grading", "peer" eller "referee review", i kombination med "essay", "thesis", "dissertation", "research paper" och med "reliability" och "consistency" har använts.

Det slags uppsatser man får information om är antingen vad vi skulle kalla skoluppsatser, där betoningen ligger på den skriftliga framställningen, eller också akademiska "term papers" som redovisar inläringen av ett visst kursinnehåll. Däremot tycks få ha studerat samstämmigheten i bedömning av uppsatser som redovisar forskningsuppgifter av mindre omfång – det som är aktuellt i detta sammanhang.

Från England kommer flera jämförelser mellan vad studenter tror att lärare grundar sin bedömning på och lärares egen bild av

detta (Hounsell, 1984, Norton, 1990, Penny & Grover, 1996, Dracup, 1997). Dessa uppfattningar stämmer sällan väl överens.

Från Förenta Staterna kommer bl.a. försök att genom databehandling av uppsatsernas språk använda antalet ord, antalet olika ord, frekvensen av olika ordklasser och liknande kvantitativa mått, vilka jämförs med motsvarande data från modelluppsatser som enligt samstämmiga bedömare är av viss angiven kvalitet. Den mest avancerade metodiken kallas Latent Semantic Analysis (LSA) som gör det möjligt att för varje uppsats beräkna en "profil" med en mängd dimensioner för denna jämförelse (Landauer & Dumais, 1994).

En av pionjärerna för datoriserad bedömning av uppsatser (Page, 1995), som redan 1966 skrev en artikel med titeln "The imminence of grading essays by computer", hävdar att ett poängtal beräknat ur formella drag, tillsammans med skattningar från en tränad bedömare, uppnår minst lika hög grad av tillförlitlighet som två personers samlade bedömning. I ett välkontrollerat experiment uppgick den genomsnittliga korrelationen mellan sex bedömare till 0.65 och den genomsnittliga korrelationen mellan "datorpoängen" och de sex bedömarna tagna var för sig till 0.74, således högre. Detta resultat gäller dock uppsatser över förelagda ämnen som del i t.ex. ett högskoleprov med tusentals provtagare. Även om datormetodiken inte är tillämplig på akademiska C-uppsatser, kan erfarenheterna av den ändå göra det motiverat att studera formella drag hos de uppsatser som denna undersökning bygger på.

Det närmaste till kvantitativt uttryckta bedömningar av vetenskapliga uppsatser som återfunnits vid litteraturgenomgången gäller *peer* eller *referee review* av manuskript som insänts för publicering i vetenskapliga tidskrifter, således på en högre vetenskaplig ambitionsnivå än uppsatser som avslutning på den akademiska grundutbildningen.

Marsh och Ball (1981, 1989) redovisar psykometriska studier av bedömningen av artiklar insända till *Journal of Educational Psychology*. De beräknar samstämmigheten mellan två bedömare av samma manuskript och jämför resultaten med dem från ett antal liknande publicerade försök. Totalomdömet för en uppsats hade i

den ena studien korrelationen 0.34 mellan bedömarna, i den andra 0.31. För enskilda delasppekter av bedömningen var samstämmigheten lägre.

Generaliseringen av dessa resultat kompliceras av att sammansättningen av bedömarparen har varierat, paren har granskat olika många manuskript och manuskripten har inte fördelats mellan bedömarna så att likvärdighet garanterats. Försök att i analysen hålla i sär manuskriptens kvalitet och bedömarnas kravnivå ledde dock inte till några stora förändringar. I jämförelse med resultaten av andra liknande studier var de av Marsh och Ball beräknade värdena snarast bättre. De kan dock inte konkurrera med de resultat för skoluppsatser som redovisats av Page (jfr ovan).

Som sammanfattning av denna kortfattade litteraturgenomgång kan konstateras att kvalitetskontrollen inom svensk högre utbildning i första hand bygger på expertgranskningar av befintlig lärarkompetens, resurser, läroplaner och organisation av undervisning. I vissa andra länder medverkar utifrån kommande ämnesexperter i examinationen. Kontrollerade mätningar av produkternas kvalitet är sällsynta och därför är denna undersökning ett viktigt komplement till övriga ansatser för kvalitetssäkring av högskolans arbete.

4 Uppläggnig

4.1 Urval av ämnen

I förberedelserna för undersökningen ansågs två ämnen vara ett rimligt omfång för en första studie. Båda ämnena borde vara stora och väletablerade både vid universitet och mindre högskolor. De borde vidare komma från olika fakultetsområden, förslagsvis samhällsvetenskap resp naturvetenskap och teknik. Jämförelsen borde i första hand avse uppsatser på C-nivå eller examensarbeten av motsvarande karaktär.

Vid valet av ämnen borde hänsyn tas till det förhållandet att C-uppsatser i engelska och statsvetenskap redan varit föremål för en granskning initierad av Höskoleverket (Kvalitet och förändring, 1997).

Inom samhällsvetenskap befanns nationalekonomi vara ett lämpligt ämne. Det ingår dels obligatoriskt i ekonomutbildningar, dels som huvudämne i andra samhällsvetenskapliga utbildningsprogram. Ämnet har en stark teoretisk kärna och en etablerad bedömningspraxis.

Inom naturvetenskap och teknik föreföll fysik, kemi och biologi vara ämnen som uppfyllde motsvarande krav. Vid närmare granskning befanns emellertid att dessa ämnen saknar motsvarigheter till C-uppsatserna inom humanistisk och samhällsvetenskaplig fakultet. Ett alternativ som diskuterades var examensarbeten från de korta ingenjörsutbildningar som förekommer dels vid tekniska högskolor, dels vid vissa regionala högskolor, men dessa examensarbeten torde vara förlagda till sena stadier i utbildningen

och därvid till skilda delämnena inom programmen. Utbildningarna är dessutom så nya att någon bedömningspraxis knappast ännu har hunnit utvecklas.

I stället riktades intresset mot humaniora. Historia befanns därvid vara ett ämne som var väl etablerat inte bara vid universiteten utan också vid ett flertal mindre högskolor. Man kunde också förvänta sig en väl utvecklad bedömningspraxis. Valet blev därför historia inom humanistisk fakultet och nationalekonomi inom samhällsvetenskaplig fakultet. Båda ämnena intar en central ställning inom sina fakultetsområden och finns vid många högskolor.

4.2 Institutionsenkät

När valet av ämnen skett skickades en enkät till institutioner med undervisning på C-nivå i historia och nationalekonomi. Enkäten omfattade frågor om tillströmning och examination på C-nivå, valet av uppsatssämnena, handledningens omfattning, C-uppsatsens syften samt uppgifter om utvärdering och kvalitetskontroll. Enkäten besvarades av alla institutioner, nämligen av 18 institutioner med C-undervisning i historia och 13 institutioner med C-undervisning i nationalekonomi. Utanför universiteten var dessa studier oftast förlagda till institutioner som omfattade flera examensämnen inom humaniora eller ekonomi och samhällsvetenskap. Resultaten av enkäten redovisas som inledning till avsnitten om historia respektive nationalekonomi.

4.3 Underlag för kvalitetsbedömning

Uppsatser på C-nivå förutsattes vara lämpliga som underlag för kvalitetsbedömningen. Om man i stället använt D-nivån skulle underlaget minskat kraftigt. En flyttning nedåt till B-nivå skulle ha kunnat göras i historia, men då hade den önskvärda ambitionsnivån gått förlorad och möjligheten att jämföra med nationalekonomi försvunnit.

Insamlingen av uppsatser som underlag för bedömning fick ske i flera steg:

- Först fastställdes genom förfrågningar vid de olika högskolorna om C-undervisning bedrivits i ämnena historia och/eller nationalekonomi under de senaste läsåren.
- Därefter begärdes listor över godkända uppsatser på C-nivå in till ESO. Eftersom denna insamling påbörjades under ferietid blev den en långvarig process.

Enligt den ursprungliga planen skulle universitet jämföras med mindre högskolor. Det visade sig emellertid att de båda ämnena var ojämnt företrädade vid de minsta högskolorna. Därför gjordes i stället en tredelning av högskolorna enligt följande:

Kategori 1. Universiteten i Uppsala, Lund, Stockholm, Göteborg, Umeå och Linköping

Kategori 2. Högskolorna i Karlstad, Växjö och Örebro som numera är universitet samt Luleå tekniska universitet.

Kategori 3. Övriga högskolor med C-undervisning i något av ämnena historia och nationalekonomi.

Den ekonomiska ramen medgav att cirka 30 uppsatser per ämne bedömdes av vardera två bedömare. Dessa uppsatser fördelades på följande sätt:

	Kategori 1	Kategori 2	Kategori 3	Totalt
Historia	3x5 uppsatser	2x3 uppsatser	3x3 uppsatser	30
National-ekonomi	3x5 uppsatser	2x3 uppsatser	3x3 uppsatser	30

I kategori 1 valdes universiteten i Uppsala, Lund och Stockholm. I kategori 2 valdes de nya universiteten i Växjö och Örebro, vilka under den tid bedömningen skulle avse var regionala högskolor. I kategori 3 valdes Mitthögskolan och högskolan i Jönkö-

ping (båda ämnena), Halmstad (historia) och Skövde (nationalekonomi). Ett önskemål att jämföra samma högskolor i båda ämnena kunde sålunda inte ut tillgodoses.

I tabell 4.1 redovisas antalet listade C-uppsatser från perioden vt 1996 – vt 1998 i historia och nationalekonomi.

Tabell 4.1 **Antal listade uppsatser i historia och nationalekonomi**

Ort	Historia			National-ekonomi		
	1 förf	>1 förf	S:a	1 förf	>1 förf	S:a
Uppsala	184	0	184	25	45	70
Lund	275	2	277	59	22	81
Stockholm	256	0	256	112	21	133
Kategori 1	715	2	717	196	88	284
Växjö	58	1	59	19	7	26
Örebro	71	2	73	10	11	21
Kategori 2	129	3	132	29	18	47
Mitthögskolan	38	2	40	8	7	15
Jönköping	8	0	8	2	6	8
Halmstad	35	1	36			
Skövde				2	4	6
Kategori 3	81	3	84	12	17	29
Totalt	925	8	933	237	123	360

I historia fanns det sammanlagt 934 uppsatser varav 8 med mer än en författare. I nationalekonomi var antalet 360, varav en tredjedel hade mer än en författare. Av dessa fanns det utrymme för att välja 30 från vartdera ämnet.

4.4 Urvalet av uppsatser

Målet vid urvalet var att få med uppsatser som tillsammans var representativa för respektive högskola, samtidigt som uppsatserna

inte kunde identifieras med avseende på vare sig författare eller högskola.

Antalet uppsatser varierade starkt mellan högskolor och ämnen. Som framgick av tabell 4.1 innehöll listorna för historieuppsatser hela 277 uppsatser vid Lunds universitet varav 2 med mer än en författare mot som lägst 8 uppsatser vid högskolan i Jönköping. Vid urvalet medtogs i historia endast uppsatser med en författare. I nationalekonomi hade en stor del av uppsatserna två författare. Där valdes uppsatser med en resp två författare var för sig. För Stockholms universitet upptog listorna 112+21 uppsatser mot 2+4 uppsatser från högskolan i Skövde. Urvalskvoterna kom på detta sätt att skilja sig starkt.

Själva urvalet gjordes på följande sätt (jfr bilagetabell 4.1). Antalet tillgängliga uppsatser delades med antalet som skulle tas ut, t.ex. för historia i Lund $275/5=55$. Med hjälp av en slumpstalstabell (Fisher & Yates, 1943, s. 90) valdes som startnummer det första tal mellan 1 och 55 som uppträdde i kolumn 2, nämligen 47. Därefter drogs nr $47+55=102$, $102+55=157$ etc. För varje högskola användes en ny kolumn i slumpstalstabellen.

Efter urvalet av historieuppsatser upptäcktes att ett antal uppsatser med lokalhistoriska ämnen hade titlar som nästan oundvikligen avslöjade varifrån de kom. Eftersom lokalhistoriska uppsatser var vanliga vid de mindre högskolorna skulle representativiteten ha skadats om dessa uppsatser hade bytts ut mot sådana med mer neutrala beteckningar. Som en kontrollåtgärd utökades i stället urvalet med fem likaså slumpmässigt valda ”neutrala” uppsatser som kan alternera med de ursprungligen valda i den kommande resultatanalysen.

En uppsats visade sig vid bedömningen sakna ett flertal sidor som inte heller återfanns i institutionens arkivexemplar. Denna uppsats utbyttes mot den som stod närmast före i listan.

4.5 Bedömningskriterier

Före överlämnandet till bedömarna aidentifierades uppsatserna med avseende på högskola samt författarens, handledarens och examinatorns namn. Det nya titelbladet kom fortsättningsvis att bära endast uppsatsens namn.

Bedömningen av uppsatser skedde enligt anvisade mallar (bilagorna 1 och 2).

Dels skattades ett antal specificerade variabler. Dessa valdes bland sådana som i andra sammanhang, t.ex. vid Höskoleverkets granskning av uppsatser i engelska och statsvetenskap (1997:8R), visat sig användbara, närmare bestämt

- *forskningsanknytning*
- *problemformulering*
- *teoretisk medvetenhet*
- *metod*
- *genomförande och slutsatser samt*
- *språklig utformning och formalia.*

Skalorna gick från 1 till 5. (Deras avsedda innehåll beskrivs i bilaga 1.) Dels gjordes en helhetsbedömning - totalomdöme - av uppsatsen i skalan *underkänd, med tvekan godkänd, godkänd, väl godkänd, mycket väl godkänd.*

Utöver dessa kvalitetsgraderingar ombads bedömarna att ange om uppsatsen huvudsakligen byggde på primärkällor eller sekundärkällor. Likaså att i 5-gradig skala skatta uppsatsämnets svårighetsgrad. Det angavs som önskvärt att bedömarna också skrev en kortfattad kommentar eller karakteristik i anslutning till varje uppsats. Denna uppmaning har samtliga bedömare följt.

4.6 Bedömare

Vid planeringen konstaterades att huvuddelen av examinationen på C-nivå numera utförs av universitetslektorer med eller utan docentkompetens och undantagsvis av professorer eller biträdande professorer. Detta bekräftades av svaren på den enkät samtliga institutioner med C-undervisning i historia och nationalekonomi (jfr avsnitten 5.1 och 6.1). Det verkade därför rimligt att de bedömare som anlätades låg på motsvarande befattningsnivå. De borde också vara jämnt fördelade mellan universitet och mindre högskolor och komma från institutioner inom ämnet som inte ingår i underlaget för bedömningen. Dessa principer har också kunnat följas.

Sålunda har varje uppsats inom ett ämne bedömts av en lärare från ett universitet (kategori 1) och en lärare från en högskola utan fasta forskningsresurser i vederbörande ämne (kategori 2 eller 3). I resultatredovisningen betecknas de Bedömare 1 resp Bedömare 2. Bedömarna har arbetat i par – tre par i varje ämne – och varje par har bedömt 10 uppsatser i nationalekonomi och 11 eller 12 uppsatser i historia, i båda ämnena spridda över samtliga institutioner. De båda bedömarna i ett par har kommit från olika högskolor, gjort sina bedömningar oberoende av varandra och veterligen inte känt till vem deras partner har varit.

4.7 Kontrollvariabler

I kapitel 2 nämndes ett antal faktorer som kan påverka uppsatsernas kvalitet. Mycket litet av sådan information har emellertid kunnat insamlas och direkt kopplas till de enskilda uppsatserna.

På individnivå finns sålunda endast uppgifter om författarnas kön, det betyg uppsatsen erhållit (G eller VG) och, när det gäller nationalekonomi, om uppsatsen skrivits av en eller två personer. Likaså finns för varje uppsats uppgift om vilken typ av källmaterial – primärt eller sekundärt – som enligt bedömarens uppfattning

huvudsakligen använts. För att erhålla mer ingående uppgifter på individnivå, t.ex. om handledarens insats, hade det varit nödvändigt att ta kontakt med författaren och/eller handledaren. Detta bedömdes som mindre lämpligt och dessutom förenat med stora felkällor.

På kollektiv nivå kan däremot vissa uppgifter hämtas från institutionsenkäten, t.ex. om hur handledningen varit anordnad, handledarnas kompetensnivå och institutionens syn på C-uppsatsens syfte. Det har också visat sig möjligt att få fram uppgifter om den allmänna nivån på de studerande vid respektive institution.

Kontrollen av alternativa faktorer som kan förklara de skillnader som finns inom materialet får emellertid till stor del ske resonemangsvis och med stöd i analyser av samspelet mellan olika faktorer i den information som föreligger. Men dessförinnan skall resultaten redovisas och diskuteras ämnesvis.

Den ämnesvisa beskrivningen av resultaten (i kapitlen 5 och 6, historia respektive 6 nationalekonomi) är disponerad på samma sätt i syfte att underlätta jämförelser. En del information måste då tyvärr upprepas från det ena ämnet till det andra med tanke på de läsare som huvudsakligen är intresserade av resultaten i ettdera ämnet.

5 Historia

5.1 Institutioner och resurser

Vid den enkät som hösten 1998 skickades ut till institutioner med C-undervisning i historia erhöles svar från 18 institutioner vid följande högskolor enligt den uppdelning som genomgående tillämpats i undersökningen:

1. Universiteten i Uppsala, Lund, Stockholm, Göteborg, Umeå och Linköping
2. De nya universiteten i Luleå, Karlstad, Växjö och Örebro
3. Övriga högskolor med C-undervisning i historia, nämligen högskolorna i Dalarna (Falun), Gävle/Sandviken, Halmstad, Jönköping, Karlskrona/Ronneby, Mitthögskolan (Härnösand), Mälardalens (Eskilstuna) och Södertörn.

Vid redovisningen av enkätresultaten har valet stått mellan att redovisa svar från samtliga högskolor eller bara svaren från de högskolor som ingår i uppsatsbedömningen. En jämförelse visade att resultaten skilde sig endast obetydligt åt och därför har med ett undantag uppgifterna från samtliga högskolor redovisats. När det gäller tillgången på handledare och deras kompetens har dock endast de högskolor tagits med som lämnat uppsatser för granskning (tabell 5.2).

5.1.1 Tillströmning och examination

En första grupp av enkätfrågorna gällde tillströmning och examination på C-nivå, närmare bestämt:

- Antalet studerande som börjat studier på C-nivå under läsåren 1996/97 och 1997/98.
- Antalet C-uppsatser som godkänts dessa två läsår, därav det antal som godkänts samma termin som C-studierna påbörjats.

Medeltalen för svaren på dessa frågor dels totalt, dels uppdelade efter kategori av högskolor redovisas i tabell 5.1. Eftersom samtliga högskolor som bedrivit C-undervisning i respektive ämne under ifrågavarande läsår har svarat, rör det sig inte om jämförelser mellan stickprov.

Tabell 5.1 Tillströmning och examination på C-nivå i historia (medeltal)

Läsår		Kategori 1	Kategori 2	Kategori 3	Totalt
96/97	Start	110	43	15	53
	Godkänd	88	28	8	39
	Inom termin	56	11	2	22
97/98	Start	105	39	29	56
	Godkänd	84	20	16	40
	Inom termin	49	10	8	22

Det finns en tydlig rangordning mellan de tre kategorierna av högskolor. Först kommer de stora institutionerna vid de gamla universiteten som i medeltal har mer än 100 nybörjare på C-nivån per läsår. Sist kommer de mindre högskolorna (kategori 3) där ämnet på vissa högskolor nådde C-nivå först 1997/98.

Bortsett från kategori 3 där antalen ökat, är frekvenserna tämligen lika de båda läsåren. Över de båda åren är andelen godkända i förhållande till nybörjare samma år 72 % och andelen godkända

samma termin 40 %. I vissa fall har studierna dragit ut på tiden på grund av att utbildningen getts på halvfart eller bara varannan termin.

Utöver detta kan följande noteras. I allmänhet värderas uppsatsen som 10 poäng av 20 erforderliga för godkänd på C-nivå, vid fyra institutioner dock som 12 och vid en som 11 poäng. Vid samtliga institutioner framläggs uppsatserna vid seminarier med opponent från den egna institutionen.

I genomsnitt får drygt en tredjedel av de godkända uppsatserna betyget Väl godkänd. Ingen tydlig variation finns mellan ämnen och högskolekategorier. Uppsatsen tillmäts något större vikt än litteraturkursen för slutbetyget. Vid urvalet till forskarutbildning, där sådan förekommer, får uppsatsen stor eller mycket stor betydelse.

Uppgifterna om den tid de studerande tar på sig för C-studier och uppsatsarbete är av mycket varierande precision. Vid de flesta institutionerna ligger värdena mellan 1 och 2 terminer där drygt halva tiden går åt för uppsatsen.

5.1.2Handledning vid uppsatsskrivning

Som framgår av tabell 5.2 handleds C-uppsatserna oftast av universitetslektorer och docenter. Professorer medverkar vid universitetsinstitutionerna.

Tabell 5.2 Antalet handledare och deras befattningsnivå

	Kategori 1	Kategori 2	Kategori 3	Summa
Professor	4			4
Docent	9	2	2	13
Lektor	9	4	3	16
Adjunkt	2			2
Summa	24	6	5	35
Antal högskolor	3	2	3	8
Medeltal	8.0	3.0	1.7	4.4

Om det totala antalet handledare enligt tabell 5.2. sätts i relation till antalet handledda uppsatser enligt tabell 4.1, så blir antalet uppsatser per handledare över hela den studerade perioden 26, närmare bestämt 30 i kategori 1, 22 i kategori 2 och 16 i kategori 3. Enligt denna mycket grova beräkning är den relativa tillgången på handledare störst vid de minsta högskolorna och minst vid universiteten. Å andra sidan är handledarnas befattningsnivå högre vid universiteten.

Enligt enkäten ges handledningen vanligen enskilt till de studerande och av en enskild handledare. Inga skillnader mellan kategorier av högskolor finns därvidlag att notera. Antalet handledningstimmor skiljer sig inte mellan olika slags högskolor.

En ytterligare aspekt på handledarskapet är tillvägagångssättet vid val av uppsatsämne. Sex alternativ presenterades och man kunde markera ett valfritt antal. Skillnaderna mellan kategorier av högskolor är obetydliga. I alla dominerar svaret Väljs av den studerande i samråd med handledaren. Därefter kommer svaren Väljs från förelagd lista över ämnen och Väljs helt fritt av den studerande. Ämnet anvisas sällan vare sig av handledaren eller inom ett institutionsprojekt. Höskoleverkets uppföljning av tillstånden att ge magisterexamen i historia (1999:3R) ger intressanta exempel på försök att bättre strukturera och styra uppsatsarbetet på D-nivå, något som också borde kunna tillämpas på C-nivå.

5.2 C-uppsatsens syften

Institutionerna ombads gradera ett antal olika syften med uppsatser på C-nivå. Syftena har formulerats med hjälp av den taxonomi som utvecklats av Gerrevall (1992, särskilt s. 88–89). Graden av instämmande kunde markeras med plus (kod 1), minus (kod -1) eller också hoppas över (kod 0). I tabell 4.3 visas medeltalen för olika syften för hela gruppen av institutioner.

Tabell 5.3 C-uppsatsens syften (uppställda i rangordning)

- Utveckla kritiskt tänkande (1.00)
- Förstå innebörden av forskning (1.00)
- Lära sig arbeta självständigt (0.94)
- Utveckla intresse för forskning (0.67)
- Lära sig söka, läsa, värdera och sammanställa information (0.61)
- Få möjlighet att fördjupa sig inom något eget valt område (0.50)
- Tillämpa kunskaper man fått genom tidigare utbildning (0.39)
- Bidra till kunskapsutvecklingen inom området (0.33)
- Lära sig samarbeta med andra (-0.50)

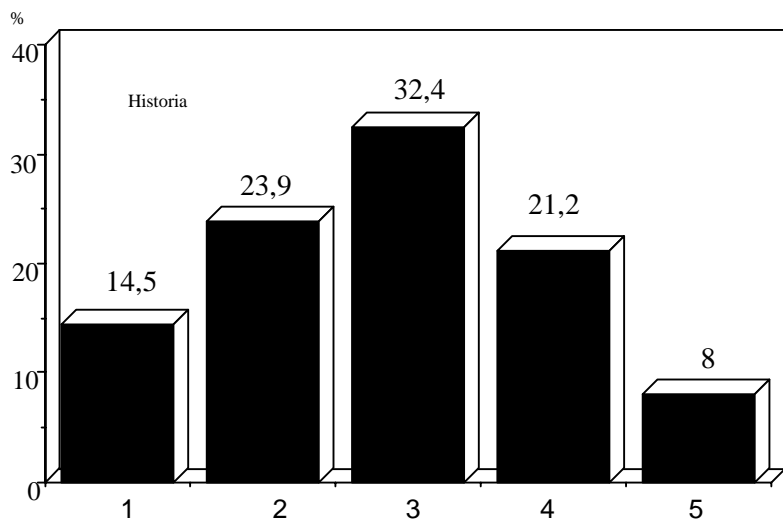
Ett syfte som inte nämndes bland svarsmöjligheterna men uppenbarligen är betydelsefullt, är att uppsatsen skall bidra med underlag för urvalet till forskarutbildning.

De tre första syftena är samstämmigt högt prioriterade. Därefter finns det utrymme för variationer. I historia där praktiskt taget alla uppsatser har en ensam författare är det naturligt att samarbete med andra hamnar på minussidan.

De största skillnaderna gäller hanteringen av information (medeltal 0.61) och tillämpningen av tidigare kunskaper (medeltal 0.39). Båda syftena tillmäts mer vikt vid de gamla universiteten än av övriga. I det förstnämnda syftet kommer de mindre högskolorna sist; i det andra är det de nya universiteten som ligger lägst med medeltalet på minussidan.

5.3 Bedömningarnas statistiska egenskaper

De sex bedömarnas huvuduppgift var att gradera uppsatsernas kvalitet dels i sex olika avseenden, dels totalt. Skalan gick från 1 till 5. Ingen särskild instruktion gavs hur man skulle fördela sina omdömen utefter denna skala. Vad totalresultatet blev för 35 uppsatser, vardera bedömd av två personer för sju olika aspekter, visas i figur 5.1.

Figur 5.1 Omdömenas fördelning på olika skalvärden

Denna fördelning påminner om den femgradiga skala som använts för de numeriska avskaffade relativa betygen i grundskola och gymnasium. Genomsnittsomdömet är här 2,84 med en spridning av 1,15.

Av tabell 5.4 framgår hur de olika aspekterna bedömdes.

Tabell 5.4 Medeltal och spridning för omdömena om uppsatser i historia samt samstämmigheten inom bedömarparen

	Medeltal	Spridning	Korrelation	Medeldifferens inom paren
Forskningsanknytning	3.01	0.94	.54	0.80
Problemformulering	2.91	1.04	.45	1.20
Teoretisk medvetenhet	2.54	1.05	.51	0.97
Metod	2.57	1.02	.48	0.91
Genomförande och slutsatser	2.94	0.94	.38	0.97
Språklig utformning och formalia	3.01	0.69	.29	0.77
Totalomdöme	2.90	1.10	.48	1.06
Genomsnittsomdöme	2.84	0.88	.56	0.76
Svårighetsgrad	3.26	0.62	.15	0.86

Skattningarna av teoretisk medvetenhet och metod har fått de lägsta värdena (2.54 resp 2.57) och forskningsanknytning och språklig utformning de högsta (3.01). Totalomdömet ligger däremellan på 2.90. Spridningen håller sig i allmänhet kring värdet 1, som motsvarar ett steg på den femgradiga skalan.

Korrelationen mellan dem som bedömt samma uppsats varierar tydligt mellan olika aspekter. Endast i få variabler når den över nivån 0.50. Skattningen av språk och formalia och uppsatsämnets svårighetsgrad, där höga värden betecknar högre svårighetsgrad, har uppenbarligen berett bedömarna större bekymmer än övriga aspekter. Det framgår av den låga spridningen och den låga korrelationen inom paren.

Genomsnittsomdömet för samtliga variabler utom svårighetsgrad visar som väntat högre samstämmighet mellan bedömarna ($r=0.56$) än de olika delaspekterna. Detta värde ligger mellan de värden som citerats i avsnitt 3, nämligen 0.65 för skoluppsatser enligt Page (1995) och 0.31–0.34 för refereebedömning av vetenskapliga artiklar enligt Marsh och Ball (1981, 1989).

Med hjälp av Spearman-Browns formel kan man skatta den mätningstekniska tillförlitligheten (reliabiliteten) i medeltalet för de

två bedömarna i denna studie till 0.72. Detta värde är inte så lågt för att gälla subjektiva bedömningar, men för att komma upp i nivå med den reliabilitet man förväntar sig vid objektiva mätningar, t.ex. 0.90, skulle man behöva ha sju i stället för två bedömare.

De två bedömarna fyller emellertid inte bara funktionen att gemensamt höja reliabiliteten från 0.56 till 0.72 utan medger också en s.k. korsvalidering av de medeltalsskillnader som uppträder. Detta åstadkoms genom att samma analys av medeltalen görs för bedömare 1 och 2 var för sig – något som kommer att ske i de följande resultatredovisningarna.

Ett alternativt sätt att mäta samstämmigheten mellan bedömarna är att beräkna den genomsnittliga differensen (bortsett från tecken) inom paren. I flertalet variabler ligger den i närheten av ett skalsteg.

En mer konkret bild av sambandens storlek ges i tabell 5.5 som visar olika kombinationer av omdömen från de två bedömarna av samma uppsats. Endast totalomdömet redovisas här men tendensen är likartad i övriga variabler.

Tabell 5.5 Samstämmigheten inom bedömarparen i fråga om totalomdömet om uppsatser i historia

	Bedö- mare 2	1	2	3	4	5	S:a
Bedömare 1	1	1	3	3	1	-	9
	2	-	1	3	-	-	4
	3	3	2	1	4	1	11
	4	-	1	1	4	3	9
	5	-	-	1	-	1	2
	S:a	5	7	9	9	5	35

Båda bedömarna har utnyttjat hela skalan från 1 till 5. Bedömare 2 är något mer positiv i sina omdömen än bedömare 1. Det är framför allt de många "underkännande" omdömena som drar ned medeltalet för bedömare 1. I bara två fall hamnar båda på skalvärdet 1 och i ett fall på 5. Dessa överensstämmelser liksom de större diskrepanserna mellan bedömarna kommer att exemplifieras i avsnitt 5.4.

Omdömena om olika aspekter är högt korrelerade inbördes och med totalomdömet, vilket visas i tabell 5.6 som avser medeltalen av de båda bedömnarnas skattningar.

Tabell 5.6 Korrelationer mellan olika bedömningsaspekter

	Foankn	Problem	Teori	Metod	Slutsats	Språk
Problem	.83					
Teori	.78	.88				
Metod	.77	.84	.86			
Slutsats	.66	.82	.81	.88		
Språk	.59	.55	.59	.75	.74	
Total	.78	.85	.88	.92	.93	.74

Överlag är korrelationerna höga, vilket betyder att det varit svårt att hålla i sär de olika aspekterna. Helhetsintrycket har tenderat att ta överhanden. Detta är ett mycket vanligt fenomen vid subjektiva bedömningar och går under namnet "haloeffekt". Här finns det dock en antydning till att den språkliga utformningen och formalia (genomsnittskorrelation med övriga 0.64) bedömts mer fristående från helhetsintrycket än övriga aspekter (0.81).

Om man på ett optimalt sätt väger samman de olika delomdömena så blir den multipla korrelationen mot totalomdömet 0.97. Till denna höga korrelation är det omdömet om genomförande och slutsatser som bidrar mest (beta-koefficient 0.52) och språklig utformning och formalia minst (beta=0.04). Det finns uppenbarligen inte mycket utrymme för variationer mellan olika aspekter på den enskilda uppsatsen. Detta utesluter dock inte möjligheten att resultat för t.ex. olika kategorier av högskolor kan uppvisa en viss profilering i omdömena.

Utöver omdömena kan analysen utnyttja de betyg som resp institution satt på uppsatserna, dvs. Godkänd (här kodad 1) eller Väl godkänd (2). I tabell 5.7 redovisas data om betygen.

Tabell 5.7 Uppsatsbetygen och deras samband med omdömena

Betygsmedeltal	1.57
Spridning	0.50
	Korrelation med omdömen
Forskningsanknytning	.32
Problemformulering	.41
Teoretisk medvetenhet	.43
Metod	.47
Genomförande och slutsatser	.48
Språklig utformning och formalia	.44
Totalomdöme	.48
Genomsnittsomdöme	.48
Svårighetsgrad	.18

Medeltalet 1.57 grundas på att 15 uppsatser fått betyget Godkänd och 20 uppsatser betyget Väl godkänd, vilket jämfört med resultaten från enkäten tyder på att VG-uppsatser blivit överrepresenterade vid det slumpmässiga urvalet, särskilt från de gamla och nya universiteten.

Betygen korrelerar som mest 0.48, närmare bestämt med totalomdömet, genomsnittsomdömet samt genomförande och slutsatser. Detta är ett något svagare samband än mellan de två bedömarna inbördes som maximalt nådde 0.56. Men situationen för lärare och externa bedömare är ju också olika. Som underlag för lärarnas betygssättning ligger inte bara uppsatsen utan även intryck av de studerande under handledning och seminariebehandling – information som de externa bedömarna inte haft tillgång till.

Sammanfattningsvis har materialöversikten gett vid handen

- att bedömarna väl har utnyttjat de olika skalstegen för sin bedömning;
- att de olika aspekter som bedömts i genomsnitt väl motsvarar det totalomdöme de gett på uppsatserna;

- att överensstämmelsen mellan bedömarna i varje par är måttlig men sannolikt inte lägre än den man brukar få mellan två bedömare av samma uppsats;
- att användningen av två bedömare medger separata medeltalsanalyser och därmed en extra kontroll av eventuella skillnaders tillförlitlighet, något som är särskilt viktigt här med ett litet antal uppsatser att jämföra;
- att sambanden mellan externa omdömen och intern betygsättning ligger på en rimlig nivå;
- att rätt många uppsatser fått totalomdömet Underkänd av endera bedömaren och inte fullt så många Mycket väl godkänd. Däremot är det ovanligt att de båda bedömarna är överens om sådana extremvärden. I nästa avsnitt skall bedömarnas kommentarer användas för att belysa såväl samstämmiga som mycket divergerande omdömen.

5.4 Några extremfall

5.4.1 Överensstämmelser

Enligt tabell 5.5 var det två uppsatser som samstämmigt bedömdes som underkända (1) och en uppsats som fick det högsta totalomdömet Mycket väl godkänd (5) av bägge bedömarna.

Båda de underkända uppsatserna byggde på sekundärkällor – den ena biografiska verk om en kontroversiell person i modern historia, den andra huvudsakligen handbokslitteratur om förvaltningen på olika nivåer i ett historiskt perspektiv och blev inte alls den genomgång av primärkällor som uppsatsens titel gav anledning att förvänta.

Huvudinvändningen mot den förstnämnda uppsatsen är från båda bedömarna att författaren förhåller sig okritiskt till sitt underlag som ibland hanteras som källor för vad som faktiskt ägt rum och andra gånger som tolkningar med ett stort inslag av eget tyckande.

Den ena bedömaren betraktar ämnet som mycket svårt, den andra som ordinarie. Den ena sätter sitt lägsta omdöme om teoretisk medvetenhet, den andra om metod. De är överens om att uppsatsen kunde ha blivit bättre om den gjorts till en historiografisk jämförelse mellan olika författares behandling av ämnet. Uppsatsen hade tidigare fått betyget Väl godkänd (VG).

Den andra enstämigt underkända uppsatsen får nästan genomgående ettor av den ena bedömaren men t.o.m. några treor av den andra. Gemensamt för deras invändningar är att uppsatsen huvudsakligen refererar handboksuppgifter som inte är relevanta i sammanhanget och i stället försummar den undersökning av tillgängligt källmaterial som skulle kunnat vara av intresse. Uppsatsen hade tidigare fått betyget Godkänd (G).

Den tredje samstämigt bedömda uppsatsen är *outstanding* inte bara för att båda bedömarna gett den en femma i totalomdöme utan också därför att de är överens på samtliga enstaka punkter – femmor på allt utom språklig utformning (4) och svårighetsgrad (4). Uppsatsen sägs uppfylla alla krav på vetenskaplighet, visa stor förtrogenhet med ämnet och ha sin styrka i metodisk och teoretisk medvetenhet. Denna uppsats hade tidigare fått betyget VG.

5.4.2 Diskrepanser

För nio uppsatser skiljer sig totalomdömena åt med två skalsteg och en med hela tre skalsteg. Diskrepanserna är ojämnt fördelade på bedömarparen – ett par med sex och ett par med tre differenser om två skalsteg. Det tredje paret skiljer sig bara en gång men då med hela tre steg. Diskrepanserna uppvägs emellertid av nästan lika många överensstämmelser så det är svårt att peka ut något av paren som genomgående mindre samstämigt än övriga.

I sju av de tio här utplockade uppsatserna hade den ena bedömaren stannat för Underkänd (1) och den andra för Godkänd (3) eller t o m Väl godkänd (4). I två av de övriga låg skillnaden mellan omdömena 3 och 5, och slutligen i en mellan 2 och 4. Det är i och för sig inte anmärkningsvärt att stora differenser uppträder när

extremvärden är inblandade. Det kan dock vara skäl att försöka spåra några drag i de skilda värderingarna.

Först uppsatser bedömda med 1 resp 3 eller 4, av vilka fyra fick betyget G och tre betyget VG:

I två av dem har bedömarna angivit olika typ av källor. Den bedömaren som uppfattat att uppsatserna huvudsakligen bygger på primärkällor har gett det högre betyget.

De delaspekter som ger störst utslag mellan den högre och den lägre bedömningen av samma sju uppsatser är i ungefär lika mån problemformulering, teoretisk medvetenhet, metod samt genomförande och slutsatser. I dessa aspekter skiljer sig omdömena i genomsnitt 1.6 eller 1.7 skalsteg mellan godkända och underkända uppsatser. I forskningsanknytning är skillnaden 1.0, i språklig utformning 0.3 och i svårighetsgrad 0. Omdömena skiljer sig mest i de variabler som med en gemensam beteckning kan kallas vetenskaplighet, och därvidlag är skillnaderna betydande.

Både i fråga om de uppsatser som fått 3 eller 4 och dem som fått 1 som totalomdöme uppehåller sig bedömarna huvudsakligen vid uppsatsernas brister och är tämligen eniga om vilka aspekter de avser. Men de hamnar ändå på olika totalomdömen vilket får ses som uttryck för olika grad av stränghet i bedömningen och kanske i sin tur olika referensramar. I några fall tillfogar de viss reservation, t ex "hyggligt" men ej mera eller "gränsfall" (nedåt) som kommentar till en trea, eller "möjligen 2" som kommentar till en etta.

I två uppsatser som fått totalomdömet 1 betonas att de har en sådan refererande karaktär att de snarast är att betrakta som B-uppsatser, och man finner det anmärkningsvärt att en författare bakat in flera sidor ur sin tidigare B-uppsats. En uppsats som fått 4 av den ena bedömaren får sin pluspoäng för "flyhänt redogörelse" medan kollegan i sin kommentar till 1 betecknar uppsatsen närmast som en essä med oprecisa frågeställningar.

Så till uppsatser bedömda med 3 resp 5 och 2 resp 4, en av dem med betyget G och två med betyget VG från institutionen:

För tre av dessa uppsatser modifieras totalomdömet i kommentaren: 3 blir 3+ att jämföra med 5, och 4 blir 4- att jämföra med 2. Omdömena ligger något närmre varandra än två hela skalsteg.

Liksom för föregående grupp av uppsatser är de särskiljande delaspekterna problemformulering, teoretisk medvetenhet, metod samt genomförande och slutsatser. Därtill kommer nu forskningsanknytning. Även här tycks det vara övergripande bedömningar av graden av vetenskaplighet som ger utslag. Detta intryck förstärks av kommentarer som "ovanligt intelligent studie" och "hög vetenskaplig klass" (för C-nivå) på plussidan och "sorglös inställning till metod" och "påklistrad allmän teori" på minussidan.

Såväl överensstämmelser som diskrepanser tyder på att bedömnarna i stort sett låtit sig styras av den grad av vetenskaplighet som de funnit hos uppsatserna, likaså att denna vetenskaplighet i stort sett sammanfaller med de aspekter som specificerats i bedömningsformuläret. Det finns en tydlig åtskillnad mellan forskningsanknytning etc. å den ena sidan och språklig utformning och formalia å den andra. Anmärkningsvärt är att fyra av nio lågt bedömda uppsatser fått betyget VG.

5.5 Jämförelser mellan kategorier av högskolor

Undersökningsmaterialet omfattar som tidigare nämnts tre kategorier av högskolor, nämligen

- tre universitet med tillsammans 15 uppsatser från historia (kategori 1)
- två nya universitet som före högskolereformen 1977 utgjorde universitetsfilialer och därefter intill 1999 var stora regionala högskolor; de representeras av 8 uppsatser i historia (kategori 2)
- tre mindre högskolor med tillsammans 12 uppsatser i historia (kategori 3).

I det stickprov som ligger till grund för beräkningarna är betyget Väl godkänd vanligare i kategori 1 och därefter i kategori 2 än

som rapporterats i enkäten, medan andelarna stämmer överens för kategori 3. Detta tyder på att uppsatser med Väl godkänd är över-representerade i kategorierna 1 och 2. Det samband mellan omdömena och uppsatsernas betyg som redovisades i tabell 5.7 kan användas för att korrigera för överrepresentationen. Metoden kallas kovariansanalys.

Metoden justerar de observerade medelvärdena till den genomsnittliga betygsnivån för hela materialet. I detta fall minskas värdena för kategorierna 1 och 2 och ökas för kategori 3 i proportion till sambandet mellan omdömen och erhållna betyg. Därmed minskas också differenserna mellan kategorierna.

I de medeltal byggda på de båda bedömarnas skattningar som redovisas i tabell 5.8 har en sådan korrektion införts. Eftersom en del av betygsskillnaderna torde vara "sanna" skillnader mellan kategorierna och inte en effekt av slumpen innebär förfarandet en risk för viss överkorrektion som dock inte allvarligt påverkat de skillnader i kvalitetsomdömena som observerats.

Tabell 5.8 Jämförelse mellan olika kategorier av högskolor i historia (efter korrektion för betygsskillnader)

	Kategori 1 n=15	Kategori 2 n=8	Kategori 3 n=12	p	eta
Foankn	3.33	3.50	2.30	.014	.49
Problem	3.40	3.24	2.08	.009	.51
Teori	2.86	2.86	1.94	.095	.38
Metod	2.65	3.09	2.13	.103	.37
Sluts	3.36	3.24	2.23	.010	.51
Språk	3.01	3.48	2.71	.037	.44
Totalomdöme	3.24	3.35	2.18	.029	.45
Genomsnittsomdöme	3.12	3.25	2.22	.015	.49
Svårighet	3.27	3.24	3.25	.993	.00

Två mått redovisas för skillnadernas storlek,

- dels p-värden där värden på högst 5 % (0.05) betraktas som signifikanta,
- dels *eta* som är ett mått på intraklasskorrelationen beräknad genom variansanalys – värdena kan gå mellan 0.00 och 1.00

där högre värden betyder större skillnader mellan kategorierna.

Kategorierna 1 och 2 turas om att få de högsta skattningarna som utom för teoretisk medvetenhet och metod ligger en bit över 3.00. Kategori 3 ligger mellan 2.50 och 2.00 utom i svårighetsgrad där den inte skiljer sig från de övriga kategorierna. Alla utom den sistnämnda variabeln samt teori och metod visar starkt signifikanta samband mellan tillhörighet till högskolekategori och omdömen. Skiljelinjen går mellan de gamla och nya universiteten på den ena sidan och de regionala högskolorna på den andra. Att det rör sig om kvalitetsskillnader bestyrks av det förhållandet att alla kategorierna har samma medeltal i bedömningen av uppsatsämnenas svårighetsgrad.

I bilagetabell 5.1 redovisas på motsvarande sätt de olika kategoriernas medelvärden uppdelade på de två bedömarna i varje par. Den första bedömaren kommer från en universitetsinstitution, den andra från en institution utan fasta resurser för forskning och forskarutbildning – i samtliga fall dock från andra högskolor än de vilkas uppsatser blivit föremål för bedömning.

Uppdelningen av bedömarparen på detta sätt ger upphov till intressanta och oväntade resultat. De kan sammanfattas med hjälp av *eta*-värdena. Medan *eta* i genomsnitt uppgick till 0.46 i tabell 5.8 (svårighetsgrad oräknad) blev det 0.28 för bedömare 1 i paren och 0.48 för bedömare 2. För bedömare 2 var alla skillnader mellan kategorierna (med undantag för metod) signifikanta, medan de inte var det för bedömare 1. De bedömare som kom från de gamla universiteten differentierade mindre mellan uppsatser som kom från olikstora högskolor. Fortfarande blir dock rangordningen densamma för båda bedömargrupperna, men alla skillnader är inte tillräckligt stora för att ge signifikans vid det låga antal uppsatser som materialet omfattar.

De hittills redovisade beräkningarna är baserade på 35 uppsatser. Av dessa är fem ett tillskott utöver planerat antal för att kontrollera för oönskade effekter av lätt identifierade uppsatser med lokalhistoriskt innehåll (jfr avsnitt 4.3). Kontrollen har gjorts genom beräkningar motsvarande dem som redovisats i tabell 5.8 och

bilagetabell 5.1. De neutralt benämnda tilläggsuppsatserna har lämnats utanför och beräkningen gjorts på 30 uppsatser. Även om de siffermässiga resultaten undergår små förändringar så står sig resultaten från hela materialet av 35 uppsatser.

Tabell 5.9 Jämförelse av *eta*-värden för två olika uppsättningar av uppsatser

	N=35	N=30
Bedömare 1	.28	.26
Bedömare 2	.48	.48
Bedömarna tillsammans	.46	.44

I de två uppsättningarna är resultaten nästan identiska. Även här uppvisar bedömare 1 lägre samband med högskolekategori än bedömare 2, vars värden ligger nära medelvärdet för de två bedömarna. Det föreligger inga skäl att inte utnyttja hela materialet om 35 uppsatser även i de fortsatta analyserna.

5.6 Studentrekryteringen inom olika högskolekategorier

En naturlig första ansats till förklaring av skillnaderna i uppsatsbedömningen är att försöka fastställa kunskapsnivån på de studenter som söker sig till ämnet vid olika högskolor. Moderna Tiders ranking använde sig av söktryck och antagningskrav som mått på olika ämnens attraktion. De användes dock inte som variabler mot vilka man kan ställa studieresultaten, vilket i detta sammanhang är det primära syftet.

Här har undersökningen fått möjlighet att använda en vid vid Göteborgsinstitutionen i pedagogik upprättad databas (UGUR) för att belysa rekryteringen till ämnena historia och nationalekonomi. Databasen omfattar bl.a. de genomsnittliga slutbetygen i grundskola och gymnasium för hela kohorterna födda 1972–1979, sammanlagt 842 800 individer. I databasen finns bl.a. alla registre-

ringar inom högskolan för dessa åldersgrupper (Gustafsson, Andersson & Hansen, 1999).

Ur databasen har alla individer tagits ut som registrerats för kurser inom historia under läsåren 1995/96 och 1996/97. Uppgifterna om registrering har sammanförts med de studerandes genomsnittliga grundskole- och gymnasiebetyg. Därefter har medelbetygen beräknats för dem som registrerats på B- och C-nivå vid de institutioner som ingår i undersökningen. B-nivån inkluderades därför att rekryteringen till C-nivå sker från denna. Däremot lämnades den kvantitativt mycket större A-nivån utanför analysen.

Genom dessa beräkningar har ett antal mått erhållits på den genomsnittliga nivån för *samtliga* B- och C-studerande som varit registrerade vid respektive institutioner under de båda läsåren. Tabell 5.10 visar medeltalen kategorivis för högskolorna.

Tabell 5.10 Genomsnittliga avgångsbetyg från grundskola och gymnasium för studerande registrerade på B- och C-nivå i historia inom olika kategorier av högskolor

		Kategori	Kategori	Kategori	p-värde	eta
		1	2	3		
B95	Grundskola	4.05	3.79	3.59	.000	.34
B96		3.97	3.91	3.62	.000	.29
B95	Gymnasium	3.82	3.47	3.22	.000	.34
B96		3.81	3.64	3.35	.000	.31
C95	Grundskola	3.98	3.78	3.59	.007	.23
C96		3.99	3.84	3.76	.027	.16
C95	Gymnasium	3.75	3.61	3.41	.101	.16
C96		3.77	3.53	3.43	.008	.19
Medeltal	B	3.91	3.70	3.44	.000	.32
	C	3.87	3.69	3.55	.034	.18
Medeltal	Grundskola	4.00	3.83	3.64	.008	.26
	Gymnasium	3.79	3.58	3.35	.027	.25

Dessa värden avser givetvis inte just uppsatsförfattarna. För det första finns ingen identifikation som gör det möjligt att ur databasen ta fram betygen från grundskola och gymnasium för dessa studerande. För det andra har många av dem som registrerats under de angivna läsåren ännu inte nått fram till färdiga uppsatser, medan

andra i gengäld är äldre än de som tillhör kohorter för vilka uppgift föreligger. Men som ett kollektivt mått på ämnets attraktivitet vid berörda högskolor har värdena ett betydande intresse.

De medeltal som presenterats i tabell 5.10 bygger, i genomsnitt för de två läsåren, på följande antal inom olika högskolekategorier:

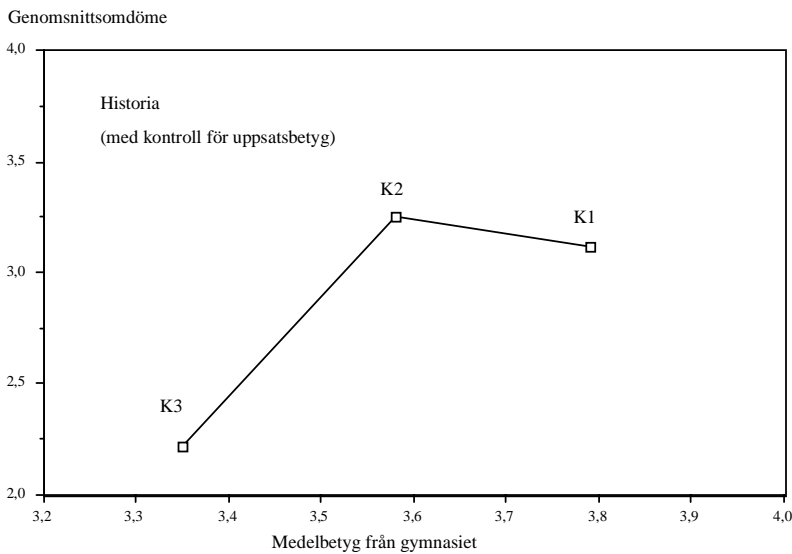
	Kategori 1	Kategori 2	Kategori 3
B-nivå	165	44	30
C-nivå	176	34	13

Dessa antal ger på det hela taget en god stabilitet hos medeltalen. Avtrappningen från kategori 1 till kategori 3 är väntad. Skillnadernas p-värden och eta visar synnerligen betryggande samband på B-nivå och inte fullt så starka på C-nivå, men rangordningen mellan kategorierna är genomgående densamma. De gamla universiteten ligger högst och de mindre högskolorna lägst.

Skillnaderna kan ändå tyckas små, om man t.ex. jämför med omdömen i tabell 5.8, men spridningarna i medelbetyg är samtidigt små jämfört med den för omdömena. Grundskolebetygen har i genomsnitt en spridning på 0.45 och gymnasiebetygen 0.57 mot spridningsvärden runt 1.0 i omdömena (tabell 5.4).

I figur 5.2 har genomsnittsomdömena i tabell 5.8 för olika högskolekategorier, korrigerade för uppsatsbetyg, satts i relation till medelbetygen från gymnasiet med B- och C-nivåerna tagna tillsammans.

Figur 5.2 Högskolekategoriernas genomsnittsomdömen i förhållande till medelbetygen från gymnasiet



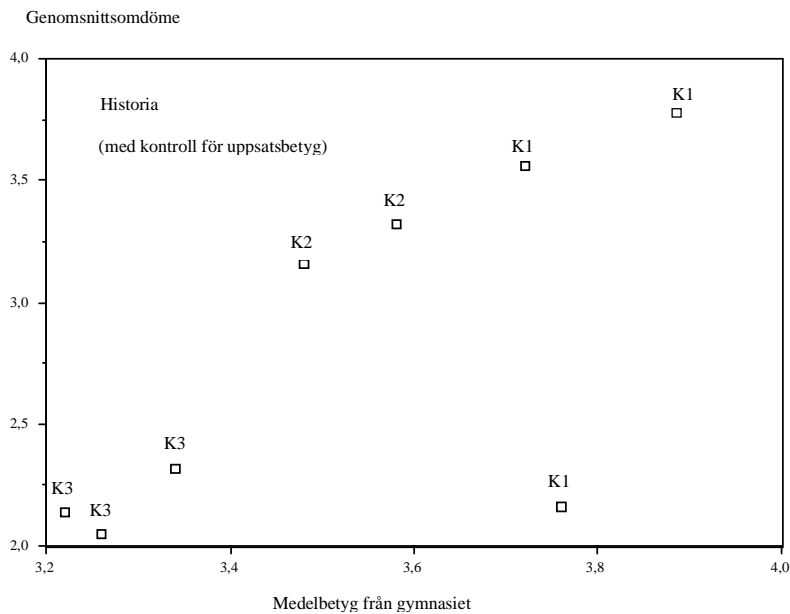
Högskolekategorierna 1 och 2 skiljer sig markant i betygsnivå men har ändå ungefär samma medeltal för genomsnittsomdömet om kategoriens uppsatser. Kategori 3 ligger avsevärt lägre på båda axlarna. Det gör ingen större skillnad om man väljer B- eller C-nivå för jämförelsen. Enligt detta skulle kategori 2 (de nya universitetet) klara sig bäst i uppsatsskrivandet i förhållande till de studerande som kategorin rekryterar.

Såsom visats i avsnitt 5.3 sammanfattar det här använda genomsnittsomdömet om uppsatserna rätt väl de specificerade kvalitetsomdömena. Det skulle tynga rapporten om separata diagram skulle visas för alla delområden. Däremot finns det en annan väg att gå djupare in i materialet, nämligen att se på de olika högskolornas genomsnittsomdömen och sätta dem i relation till de attraktivitetsmått som här använts. Detta görs visserligen med stor tve-

kan eftersom varje enskild högskola är företrädd med bara 3 till 5 uppsatser. Högskolorna bör därför i en sådan analys få förbli anonyma.

I figur 5.3 visas sambandet mellan genomsnittsomdömet om uppsatserna, korrigerat för uppsatsens betyg (jfr avsnitt 5.7), och det attraktivitetsmått som baseras på gymnasiebetyg – motsvarigheten till figur 5.2.

Figur 5.3 Högskolornas genomsnittsomdömen i förhållande till genomsnittsbetygen från gymnasiet



Sambandet är starkt och regelbundet för sju av de åtta högskolorna som i stort sett ligger på en gemensam kurva med kategorierna 1, 2 och 3 i nämnd ordning. En av kategori 1-högskolorna faller dock utanför en sådan trend. Den har ett, av dessa mått att döma, mycket gott studerandeunderlag men svaga resultat i uppsats-

bedömningen trots att fyra av de fem uppsatserna erhållit betyget VG. Kategori 1 blir därigenom heterogen och dess förhållandevis låga placering i figur 5.2 får då sin förklaring.

5.7 Uppsatsernas källmaterial

I materialet finns det några variabler som kan bidra till att förklara skillnaderna mellan högskolorna. I första hand avses uppgiften om källmaterial till uppsatsen. I tabell 5.11 visas vilken typ av källor som enligt bedömarna kommit till huvudsaklig användning.

Tabell 5.11 Källmaterial för uppsatser i historia

	Primärkällor	Sekundärkällor	Bådadera	Samtliga
Kategori 1	10	4	1	15
Kategori 2	5	1	2	8
Kategori 3	1	7	4	12
Samtliga	16	12	7	35

Materialet faller i tre undergrupper:

- primärkällor angivna av båda bedömarna
- sekundärkällor angivna av båda bedömarna
- bådadera som betecknar när primärkällor anges av den ena och sekundärkällor av den andra bedömaren *eller* när bedömaren undantagsvis har markerat båda.

Av tabell 5.11 framgår att primärkällor anges oftare för uppsatser från högskolekategorierna 1 och 2, medan sekundärkällor och bådadera förekommer oftare i kategori 3. Skillnaden är signifikant.

Tabell 5.12 Jämförelse mellan uppsatser i historia baserade på olika slags källor

	Primär n=16	Sekundär n=12	Bådadera n=7	p	eta
Foankn	3.53	2.33	3.00	.002	.57
Problem	3.59	2.08	2.79	.000	.66
Teori	3.12	1.71	2.50	.001	.59
Metod	3.12	1.71	2.79	.000	.64
Sluts	3.47	2.25	2.93	.001	.58
Språk	3.25	2.67	3.07	.081	.38
Totalomdöme	3.53	2.00	3.00	.000	.63
Genomsnittsomdöme	3.38	2.17	2.78	.000	.61
Svårighet	3.31	3.08	3.43	.464	.22
Uppsatsbetyg	1.75	1.42	1.43	.156	.33

Enligt tabell 5.12 har uppsatser baserade på primärkällor värderats högre än uppsatser baserade på sekundärkällor. Undergruppen bådadera ligger däremellan. Skillnaderna är starkt signifikanta utom för språk och svårighetsgrad.

Det finns också en antydning till att uppsatser baserade på primärkällor fått högre betyg än övriga.

Även här har resultaten uppdelats på bedömare 1 och 2 (bilagetabell 5.2). Liksom när det gällde variabeln högskolekategori skiljer sig resultaten på ett slående sätt mellan de båda bedömarna. De bedömarna som kommer från de gamla universiteten har kraftigt signifikanta skillnader mellan uppsatser baserade på primär- resp sekundärkällor i skattningarna av problemformulering, teoretisk medvetenhet, metod, slutsatser och totalomdöme. Bedömarna från övriga högskolor skiljer sig signifikant endast i fråga om problemformulering. De genomsnittliga *eta*-värdena, bortsett från skattningen av svårighetsgrad, uppgår till 0.55 för bedömare 1 och 0.37 för bedömare 2. För variabeln högskolekategori var motsvarande *eta*-värden 0.28 resp 0.48 – praktiskt taget omkastade värden.

En kompletterande variansanalys där både högskolekategori och källor ingår som förklarande variabler understryker dessa resultat (tabell 5.13). Bedömarna från de gamla universiteten har starkt signifikanta samband mellan primärkälla och omdömen om metod, genomförande och slutsatser samt totalomdöme men inga sådana

samband med högskolekategori. Övriga bedömare har starkt signifikanta samband mellan variabeln högskolekategori och samtliga kvalitetsomdömen men inga samband med källa.

Tabell 5.13 Tvåvägs variansanalyser av högskolekategori x källa för bedömare 1 och 2 var för sig (p-värden)

Bedömare 1

	Högskole- kategori	Källa	Samspel
Forskningsanknytning	.225	.494	.717
Problemformulering	.374	.113	.798
Teoretisk medvetenhet	.480	.103	.717
Metod	.489	.002	.919
Genomförande och slutsatser	.476	.001	.437
Språklig utformning och formalia	.456	.199	.896
Totalomdöme	.368	.005	.811
Genomsnittsomdöme	.322	.017	.764
Svärighetsgrad	.473	.169	.253

Bedömare 2

	Högskole- kategori	Källa	Samspel
Forskningsanknytning	.016	.438	.936
Problemformulering	.011	.487	.393
Teoretisk medvetenhet	.024	.790	.698
Metod	.028	.878	.366
Genomförande och slutsatser	.001	.532	.504
Språklig utformning och formalia	.025	.782	.434
Totalomdöme	.005	.422	.480
Genomsnittsomdöme	.002	.652	.538
Svärighetsgrad	.868	.384	.969

Bedömarna från de gamla universiteten lägger sålunda stor vikt vid primäranalyser, och sådana förekommer oftare i uppsatser från kategorierna 1 och 2. De olika typerna av källor har mer likartad status hos övriga bedömare. Slutresultaten blir ändå på det hela taget desamma, nämligen att uppsatser från kategori 3 bedöms ligga på en lägre kvalitetsnivå än övriga.

5.8 Uppsatsämnena

Med uppsatsernas titlar som underlag har ämnena delats in i olika grupper. Resultatet blev två system, ett efter kronologiska principer och ett efter sakinnehållet i uppsatserna. Fördelningen blev följande:

Äldre och tidigmoderna	Genus/religion	3
	Näring	4
	Politik, inrikes	1
	Politik utrikes	0
	Övrigt	3
	Summa	11
Moderna	Genus/religion	0
	Näring	4
	Politik, inrikes	8
	Politik utrikes	8
	Övrigt	4
	Summa	24

De moderna ämnena är dubbelt så vanliga som de äldre och tidigmoderna. Som moderna ämnen räknas de som avser tiden efter 1850. Gränsen mellan äldre och tidigmoderna går vid 1600, men det var bara två ämnen som kunde betecknas som äldre.

Det finns inga tydliga skillnader mellan de tre kategorierna av högskolor. Det enda sambandet med omdömena gällde fördelningen mellan de två kronologiska grupperna i förhållande till uppsatsernas bedömda svårighetsgrad. De äldre och tidigmoderna uppsatserna fick i genomsnitt värdet 3.68, medan de moderna fick 3.06.

5.9 Formella drag hos uppsatserna

Den forskning som studerat möjligheterna att komplettera lärares eller experters uppsatsbedömning med datoriserade kvantitativa mått har utnyttjat uppsatserns längd, ordförrådets sammansättning o. dyl. språkliga och formella drag (jfr kapitel 3). Veterligen har sådana mått inte prövats på uppsatser av den karaktär som denna

studie avser. Här görs dock ett försök med mått som relativt lätt kan beräknas utan datorhjälp, nämligen

- uppsatsens längd mätt genom skattning av antalet ord i huvudtexten
- svensk eller engelsk text
- antalet underrubriker
- antalet tabeller och figurer
- antalet referenser
- antalet fotnoter
- antalet bilagor.

I tabell 5.14 redovisas medeltal och spridning för dessa formella drag hos uppsatserna.

Tabell 5.14 Medeltal och spridning för formella drag hos uppsatserna i historia

Antal	Medeltal	Spridning	Variationskoeff
Ord	14 657	5 025	0.34
Underrubriker	25.5	13.4	0.53
Tabeller och figurer	2.5	4.0	1.60
Referenser	30.6	12.0	0.39
Fotnoter	121.7	63.7	0.52
Bilagor	2.2	2.9	1.32

Endast två av uppsatserna i historia var skrivna på engelska varför denna variabel faller bort.

Den genomsnittliga C-uppsatsen i historia var nästan 15 000 ord lång och spridningen är hela 5 000. Genomgående är variationerna stora. För tabeller och figurer respektive bilagor är spridningarna t.o.m. större än motsvarande medeltal, vilket innebär att fördelningarna är mycket skeva med en stor andel nollfrekvenser och en uppåt långt utdragen kurva.

Sambandet mellan dessa mått och bedömningarna av uppsatsens kvalitet i olika avseenden redovisas i tabell 5.15.

Tabell 5.15 Korrelationen mellan formella drag och uppsatsernas kvalitet

	Ord	Rubrik	Tab/fig	Ref	Noter	Bilagor
Foankn	.32	.08	.18	.56	.27	-.27
Problem	.21	-.00	.18	.57	.10	-.32
Teori	.37	.11	.22	.57	.20	-.25
Metod	.32	.16	.25	.59	.22	-.25
Slutsats	.36	.23	.23	.60	.30	-.20
Språk	.48	.37	.14	.54	.45	-.16
Totalomdöme	.36	.18	.19	.64	.36	-.21
Genomsnittsomdöme	.37	.16	.22	.64	.29	-.26
Svårighet	-.05	-.13	.01	-.13	-.29	-.12
Uppsatsbetyg	.47	.12	.13	.24	.31	.07

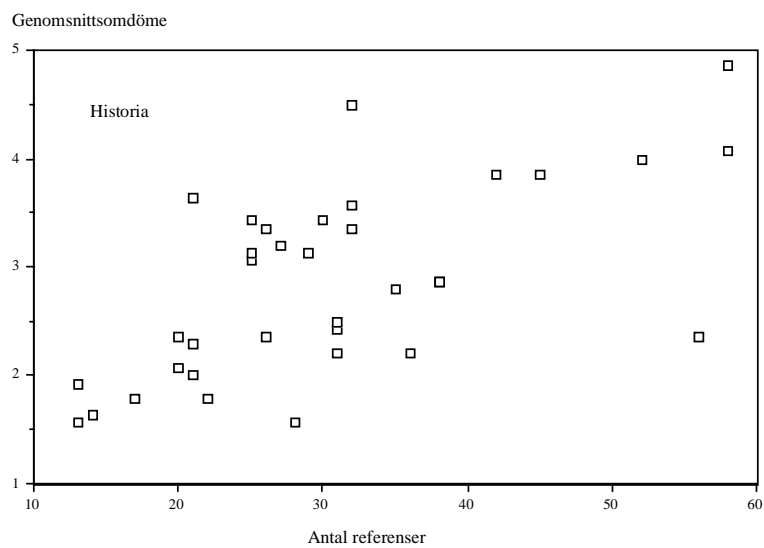
Korrelationer på 0.33 och högre är signifikanta på 5-procentsnivån. Antalet referenser skiljer sig på ett markant sätt från övriga formella variabler. Korrelationerna ligger mellan 0.54 och 0.64 vilket är väl så högt som korrelationen mellan de två bedömarna inbördes (jfr tabell 5.4). Därefter kommer antalet ord med ett högsta värde på 0.48 för bedömningen av språklig utformning och formalia. Också antalet noter har en korrelation av denna storleksordning med språkbedömningen.

De av institutionerna satta uppsatsbetygen har en högsta korrelation på 0.47 och denna gäller antalet ord. Ju längre uppsats desto högre betyg. Därefter kommer antalet noter.

Korrelationerna i tabell 5.15 förutsätter en lineär relation mellan formalia och kvalitet. Det finns skäl att kontrollera att så förhåller sig. Därför har även andragskurvor anpassats till de olika relationerna och multipla korrelationer beräknats. Som regel blir dessa något högre på grund av ett svagt avtagande utbyte efter en viss nivå, men endast i ett fåtal fall är den kvadratiske termen signifikant.

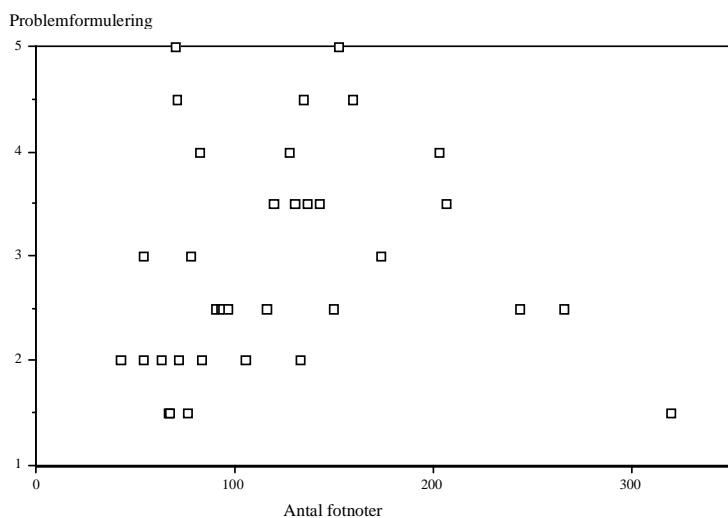
Först den formella variabel som har den starkaste korrelationen med bedömningarna, nämligen referenser (tabell 5.15). Sambandet återges i figur 5.4.

Figur 5.4 Sambandet mellan referenser och genomsnittsomdöme



Så ett exempel på ett samband där korrelationen enligt tabell 5.15 är låg men den kvadratiske termen signifikant och ger en avsevärt högre multipel korrelation (0.49 mot 0.10, figur 5.5). Här stiger omdömena med ökat antal fotnoter till en viss nivå för att sedan sjunka.

Figur 5.5 Sambandet mellan fotnoter och omdöme om problemformulering



Antalet ord är mycket större i uppsatser där bedömarna varit eniga om placeringen i primär- resp sekundärkällor än när placeringen är oklar (jfr 5.6). Kategorin bådadera ligger klart lägst också i fråga om fotnoter och den placeringen kan möjligen bero på att uppsatsen är för kort för att ge en tillförlitlig bild av källanvändningen. Uppsatser baserade på sekundärkällor har flest underrubriker och fotnoter (signifikant) samt flest ord (ej signifikant).

Eftersom uppsatser med sekundärkällor fick lägre omdömen än övriga och dessutom förekom oftare vid de minsta högskolorna (jfr avsnitt 5.7) finns det också skäl att jämföra formalia mellan kategorier av högskolor. Signifikanta skillnader uppträdde i följande variabler.

	Ord	Rubriker	Referenser	Noter
Kategori 1	16 690	23.5	31.9	141.6
Kategori 2	17 940	45.7	41.5	195.0
Kategori 3	11 420	18.2	23.9	63.1

Kategori 2 (de nya universiteten) ligger högst i alla fyra måtten på formalia, dock ej signifikant skild från kategori 1 när det gäller ordmängd. Kategori 3 ligger lägst precis som den gjorde i fråga om kvalitetsomdömena.

5.10 Kvinnliga och manliga författare

Uppsatserna i historia var författade av 14 kvinnor och 21 män. De har jämförts inbördes med samma metodik som tillämpats i tabellerna 5.8 och 5.11 med tillhörande bilagetabeller. Inte en enda differens har varit i närheten av signifikans. Medeltalen för *eta* uppgick till 0.09 för bedömaren 1, 0.11 för bedömaren 2 och 0.06 för bedömarna tillsammans. Dessa ligger långt under *eta*-värdena i övriga jämförelser mellan undergrupper.

6 Nationalekonomi

6.1 Institutioner och resurser

Vid den enkät som hösten 1998 utsändes till institutioner med C-undervisning i nationalekonomi erhöles svar från 13 institutioner vid följande högskolor enligt den uppdelning som genomgående tillämpats i undersökningen:

1. Universiteten i Uppsala, Lund, Stockholm, Göteborg, Umeå och Linköping
2. De nya universiteten i Luleå, Karlstad, Växjö och Örebro
3. Övriga högskolor med C-undervisning i nationalekonomi, nämligen högskolorna i Jönköping och Skövde samt Mitthögskolan (Östersund).

Vid redovisningen av enkätresultaten har valet stått mellan att redovisa svar från samtliga högskolor eller bara svaren från de högskolor som ingår i uppsatsbedömningen. En jämförelse visade att resultaten skilde sig endast obetydligt åt och därför har med ett undantag uppgifterna från samtliga högskolor redovisats. När det gäller tillgången på handledare och deras kompetens har dock endast de högskolor tagits med som lämnat uppsatser för granskning (tabell 6.2).

6.1.1 Tillströmning och examination

De första enkätfrågorna gällde tillströmning och examination på C-nivå, närmare bestämt

- Antalet studerande som börjat studier på C-nivå under läsåren 1996/97 och 1997/98.
- Antalet C-uppsatser som godkänts dessa två läsår, därav det antal som godkänts samma termin som C-studierna påbörjats samt antalet uppsatser med mer än en författare.

Medeltalen för svaren på dessa frågor dels totalt, dels uppdelade efter kategori av högskolor redovisas i tabell 6.1.

Tabell 6.1 Tillströmning och examination på C-nivå i nationalekonomi (medeltal)

Läsår		Kategori			Totalt
		1	2	3	
96/97	Start	84	22	8	50
	Godkänd	38	12	5	23
	Inom termin	22	7	2	11
	> 1 förf	12	4	1	7
97/98	Start	74	28	28	51
	Godkänd	38	14	9	24
	Inom termin	22	11	9	15
	> 1 förf	15	5	5	9

Det finns en tydlig rangordning mellan de tre kategorierna av högskolor. Först kommer de stora institutionerna vid de gamla universiteterna med runt 80 nybörjare på C-nivån per läsår. Sist kommer de mindre högskolorna (kategori 3).

Bortsett från kategori 3 där antalen ökat är frekvenserna tämligen lika de båda läsåren. Över de båda åren är andelen godkända i förhållande till nybörjare samma år 47 % och andelen godkända samma termin 26 %. I vissa fall har studierna dragit ut på tiden på grund av att utbildningen getts på halvfart eller bara varannan termin. Inte så få uppsatser har till skillnad från historia mer än en författare.

Vidare värderas uppsatsen genomgående som 10 poäng av 20 erforderliga för godkänd. Vid samtliga institutioner framläggs uppsatserna vid seminarier med opponenter från den egna institutionen.

I genomsnitt får knappt en tredjedel av de godkända uppsatserna betyget Väl godkänd. Ingen tydlig variation finns mellan de tre högskolekategorierna. Uppsatsen tillmäts något större vikt än litteraturkursen för slutbetyget. Vid urvalet till forskarutbildning, där sådan förekommer, får uppsatsen stor eller mycket stor betydelse.

Uppgifterna om den tid de studerande tar på sig för C-studier och uppsatsarbete varierar mycket. Vid de allra flesta institutionerna ligger värdena mellan 1 och 2 terminer där drygt halva tiden går åt för uppsatsen.

6.1.2Handledning vid uppsatsskrivning

Som framgår av tabell 6.2 handleds C-uppsatserna oftast av universitetslektorer. Professorer medverkar inte bara vid universitetsinstitutionerna som i historia utan inom alla kategorier av högskolor. Även adjunkter eller motsvarande (t.ex. doktorander) medverkar i större utsträckning än i historia.

Tabell 6.2 Antalet handledare och deras befattningsnivå

	Kategori 1	Kategori 2	Kategori 3	Summa
Professor	15	2	4	21
Docent	11	1	2	14
Lektor	27	2	4	33
Adjunkt	5	1	1	7
Summa	58	6	11	75
Antal högskolor	3	2	3	8
Medeltal	19.3	3.0	3.7	9.4

Antalet uppsatser (enligt tabell 4.1) per handledare är fem, närmare bestämt i kategori 1 fem, i kategori 2 åtta och i kategori 3 två och en halv. I alla tre kategorierna understiger det kraftigt antalet uppsatser per handledare i historia enligt denna grova beräkning.

Kategori 3 har i båda ämnena det minsta antalet uppsatser per handledare, och skillnaden mellan ämnena är störst vid de gamla universiteten.

Handledningen ges vanligen enskilt till de studerande och av en enskild handledare. Inga skillnader mellan kategorier av högskolor förekommer. Antalet handledningstimmar skiljer sig heller inte.

En ytterligare aspekt på handledarskapet är tillvägagångssättet vid val av uppsatsämne. Sex alternativ presenterades och man kunde markera ett valfritt antal. Skillnaderna mellan kategorier av högskolor är obetydliga. I alla kategorier dominerar svaret Väljs av den studerande i samråd med handledaren. Därefter kommer svaren Väljs från förelagd lista över ämnen och Väljs helt fritt av den studerande. Ämnet anvisas sällan vare sig av handledaren eller inom ett institutionsprojekt. Enligt uppgift är dock en förändring på gång som kan komma att innebära ökad styrning av uppsatsarbetet.

6.2 C-uppsatsens syften

Institutionerna ombads gradera ett antal olika syften med uppsatser på C-nivå (jfr Gerrevall, 1992). Markeringen kunde göras med plus (kod 1), minus (kod -1) eller också hoppas över (kod 0). I ta-bell 6.3 visas medeltalen för olika syften för hela gruppen av institutioner.

Tabell 6.3 C-uppsatsens syften (uppställda i rangordning)

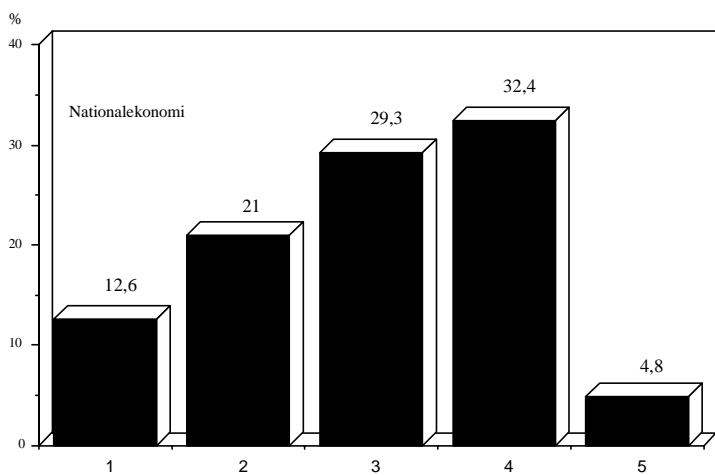
- Utveckla kritiskt tänkande (0.85)
- Lära sig söka, läsa, värdera och sammanställa information (0.85)
- Tillämpa kunskaper man fått genom tidigare utbildning (0.77)
- Förstå innebörden av forskning (0.69)
- Lära sig arbeta självständigt (0.69)
- Få möjlighet att fördjupa sig inom något eget valt område (0.69)
- Utveckla intresse för forskning (0.46)
- Lära sig samarbeta med andra (0.15)
- Bidra till kunskapsutvecklingen inom området (-0.23)

Ett syfte som inte nämndes bland svarsmöjligheterna men uppenbarligen är betydelsefullt är att uppsatsen skall bidra med underlag för urvalet till forskarutbildning.

De största förskjutningarna i rangordningen jämfört med historia gäller samarbete med andra, som ju förekommer rätt ofta i C-uppsatserna i nationalekonomi, därefter att förstå innebörden av forskning och bidra till kunskapsutvecklingen. I de sistnämnda avseendena är förväntningarna i nationalekonomi lägre än i historia – mest så vid universiteten.

6.3 Bedömningarnas statistiska egenskaper

Bedömarnas huvuduppgift var att gradera uppsatsernas kvalitet dels i sex specificerade avseenden, dels totalt. Skalan gick från 1 till 5. Ingen särskild instruktion gavs hur man skulle fördela sina omdömen utefter denna skala. Vad totalresultatet blev för 30 uppsatser, vardera bedömd av två personer för sju olika aspekter, visas i figur 6.1.

Figur 6.1 Omdömenas fördelning på olika skalvärden

Fördelningen är här påtagligt sned med en större anhopning på värdet 4 och en obetydlig användning av värdet 5 på skalan. Genomsnittsomdömet är 2.96 med en spridning av 1.11.

Tabell 6.4 visar hur de olika aspekterna bedömdes.

Tabell 6.4 Medeltal och spridning för omdömena om uppsatser i nationalekonomi samt samstämmigheten inom bedömarparen

	Medeltal	Spridning	Korrelation	Medeldifferens inom paren
Forskningsanknytning	3.03	0.89	.44	0.80
Problemformulering	3.07	0.94	.39	1.07
Teoretisk medvetenhet	3.07	0.98	.60	0.73
Metod	2.97	0.90	.45	0.93
Genomförande och slutsatser	2.95	1.00	.56	0.70
Språklig utformning och formalia	2.83	0.95	.37	1.00
Totalomdöme	2.78	0.98	.44	0.83
Genomsnittsomdöme	2.96	0.82	.58	0.68
Svårighetsgrad	3.30	0.52	-.19	0.90

Skattningen av språklig utformning och formalia har jämte totalomdömet fått de lägsta värdena (2.83 respektive 2.78) och problemformulering och teoretisk medvetenhet de högsta (3.07). Spridningen håller sig i allmänhet kring värdet 1, som motsvarar ett steg på den femgradiga skalan.

Korrelationen mellan dem som bedömt samma uppsats varierar tydligt mellan olika aspekter. I tre variabler når den över nivån 0.55. Skattningen av uppsatsämnets svårighetsgrad, där höga värden betecknar högre svårighetsgrad har bedömarna större bekymmer än övriga aspekter. Det framgår av den låga spridningen och den negativa korrelationen inom paren.

Genomsnittsomdömet har ungefär samma grad av samstämmighet som i historia ($r=0.58$) som dock överträffas av omdömet om teoretisk medvetenhet. Reliabiliteten för de två bedömarna tillsammans blir 0.73. Likaså gäller de resonemang som i historia fördes beträffande nivån på detta värde och behovet av så mycket som, i detta fall, sex bedömare om man strävade efter att uppnå en samlad reliabilitet på 0.90. Den dubbla betydelsen av att ha åtminstone två bedömare gäller också här – en förhöjning av reliabiliteten i omdömena och en möjlighet till intern kontroll av medeltalsdifferenser genom dubblering av resultatanalyserna.

Ett alternativt sätt att mäta samstämmigheten mellan bedömarerna är att beräkna den genomsnittliga differensen (bortsett från tecken) inom paren. I flertalet variabler ligger den i närheten av ett skalsteg.

En mer konkret bild av sambandens storlek ges i tabell 6.5 som visar olika kombinationer av omdömen från de två bedömarerna av samma uppsats. Endast totalomdömet redovisas här men tendensen är likartad i övriga variabler.

Tabell 6.5 Samstämmigheten inom bedömarparen i fråga om totalomdömet om uppsatser i nationalekonomi

	Bedö- mare 2	1	2	3	4	5	S:a
Bedömare 1	1	3	-	-	1	-	4
	2	3	2	-	-	1	6
	3	-	3	4	3		10
	4	1	1	4	3	1	10
	5	-	-	-	-	-	-
	S:a	7	6	8	7	2	30

Bedömare 1 har inte ansett någon vara förtjänt av omdömet 5. Tre uppsatser har blivit "underkända" av båda bedömarerna. Överensstämmelser och stora diskrepanser behandlas närmare i avsnitt 6.4.

Bedömarernas samlade omdömen om olika aspekter är högt korrelerade inbördes och med totalomdömet, vilket visas i tabell 6.6.

Tabell 6.6 Korrelationer mellan olika bedömningsaspekter

	Foankn	Problem	Teori	Metod	Slutsats	Språk
Problem	.65					
Teori	.74	.68				
Metod	.64	.84	.76			
Slutsats	.64	.80	.76	.79		
Språk	.50	.56	.52	.54	.69	
Total	.75	.84	.80	.83	.87	.74

Överlag är korrelationerna höga, vilket betyder att det varit svårt att hålla i sär de olika aspekterna. Helhetsintrycket har tenderat att ta överhanden. Språklig utformning och formalia har dock bedömts mer fristående från helhetsintrycket (genomsnittskorrelation med övriga 0.59) än övriga aspekter (0.71).

Om man på ett optimalt sätt väger samman de olika delomdömena så blir den multipla korrelationen mot totalomdömet 0.94. Det högsta beta-värdet avser språklig utformning och formalia (0.25). Det finns inte heller här så mycket utrymme för variationer mellan olika aspekter på den enskilda uppsatsen. Detta uteluter inte möjligheten att resultat för t.ex. olika kategorier av högskolor kan uppvisa en viss profilering i omdömena.

Utöver omdömena finns också lärarnas betyg på uppsatserna tillgängliga. Dessa redovisas i tabell 6.7.

Tabell 6.7 Uppsatsbetygen och deras samband med omdömena

Betygsmedeltal	1.23
Spridning	0.43
	Korrelation med omdömen
Forskningsanknytning	.16
Problemformulering	.38
Teoretisk medvetenhet	.21
Metod	.38
Genomförande och slutsatser	.23
Språklig utformning och formalia	.31
Totalomdöme	.25
Genomsnittsomdöme	.31
Svårighetsgrad	-.09

Endast 7 av 30 uppsatser hade åsatts betyget Väl godkänd. Detta resulterar i ett medeltal på 1.23 och på det hela taget låga korrelationer med omdömena, som mest 0.38 med problemformulering och metod. Det är svagare än sambanden mellan de två bedömarna som maximalt nådde 0.60. Medan omdömena endast har påverkats av uppsatsernas innehåll och form har lärarnas betygsättning också

kunnat påverkas av intryck från handledning och seminariebehandling.

Även i nationalekonomi har materialöversikten gett vid handen

- att bedömarna väl har utnyttjat de olika skalstegen för sin bedömning;
- att de olika aspekter som bedömts i genomsnitt väl motsvarar det totalomdöme de gett på uppsatserna;
- att överensstämmelsen mellan bedömarna i varje par är måttlig men sannolikt inte lägre än den man brukar få mellan två bedömare av samma uppsats;
- att användningen av två bedömare medger separata medeltalsanalyser och därmed en extra kontroll av eventuella skillnaders tillförlitlighet, något som är särskilt viktigt här med ett litet antal uppsatser att jämföra;

Speciellt för nationalekonomi gäller

- att sambanden mellan externa omdömen och intern betygsättning ligger på en låg nivå;
- att rätt många uppsatser fått totalomdömet Underkänd av endera bedömaren men få Mycket väl godkänd. I nästa avsnitt skall bedömarnas kommentarer användas för att belysa såväl samstämmiga som mycket divergerande omdömen.

6.4 Några extremfall

6.4.1 Överensstämmelser

Enligt tabell 6.5 var det tre uppsatser som samstämmigt bedömdes som underkända (1). Ingen uppsats fick det högsta totalomdömet Mycket väl godkänd (5) av båda bedömarna.

En av de tre underkända uppsatserna skiljer ut sig från de övriga genom att bedömarna gett den ettor inte bara totalt utan praktiskt taget genomgående. Den har karaktären av referat av tre teman rörande ett lands ekonomi utan att dessa relateras till varandra och utan hänvisningar till specifika källor. Uppsatsen hade tidigare fått betyget G.

Nästa uppsats får av den ena bedömaren omdömet 2 för flera delaspekter, men bedömarna är överens om totalomdömet 1 med hänvisning till att syftet är diffust, texten ostrukturerad och språket dåligt. Uppsatsen hade tidigare fått betyget G.

Den tredje underkända uppsatsen får vissa tvåor och treor för delaspekter men faller på extremt dåligt språk och oskicklig beskrivning av den relativt avancerade statistiska behandlingen. Uppsatsen hade tidigare fått betyget VG.

6.4.2 Diskrepanser

För tre uppsatser skiljer sig totalomdömet hela tre skalsteg (1 till 4 och 2 till 5) och för en uppsats två steg (2 till 4).

Eftersom bara två femmor förekom i nationalekonomi får skillnaden mellan 5 och 2 betraktas som den mest extrema diskrepansen. Den ena bedömaren har satt 5 på alla delaspekter utom språk och betecknar uppsatsen som extremt mogen, dock med en liten undran "hur nära någon annan källa följts". Den andra bedömaren har spritt sina delpoäng över hela skalan med dåligt språk (1) som lägsta omdöme. Betyg G.

Dåligt språk (1) tycks också vara avgörande för den ena bedömaren i en uppsats han gett totalomdömet 1 mot 4 från kollegan inom bedömarparet. I övrigt skiljer de sig inte mycket. Den som gett 4 har fem fyror i delaspekterna mot fyra fyror hos den som underkänt hela uppsatsen. Betyg VG.

I det andra extremfallet med 4 resp 1 som totalomdöme ställs en genomgående rad av fyror mot en mera varierad bedömning hos den bedömaren som stannat på 1. Totalomdömet dras ned främst genom brister i avvägningen inom uppsatsen. Den metod som tillämpas i vissa beräkningar presenteras mycket ofullständigt. I gengäld

uppehåller författaren sig utförligt vid mindre centrala frågor. Betyg G.

När det gäller den fjärde uppsatsen bedömd med 4 respektive 2 är båda bedömarna överens i flera delasppekter, likaså om att det är svårt att urskilja författarens egna bidrag i de refererande delarna av uppsatsen, men de stannar på olika sidor om godkändmärket. Betyg G.

Den språkliga utformningen och formalia figurerar över huvud taget oftare som motiv för totalomdömet i nationalekonomi än vad det gjorde i historia. Det kan i sin tur ha att göra med att flertalet av uppsatserna i nationalekonomi bygger på sekundärkällor. Större vikt får därvid uppsatsens organisation och presentation när det gäller att sammanställa och tolka teorier eller resultat som hämtats från litteraturen. I historia var i gengäld egen forskning baserad på primärmaterial den prioriterade aktiviteten. Kanske historikerna även faktiskt var bättre som skribenter.

6.5 Jämförelser mellan kategorier av högskolor

Undersökningsmaterialet omfattar som nämnts tre kategorier av högskolor, nämligen

- tre universitet med tillsammans 15 uppsatser från nationalekonomi (kategori 1)
- två nya universitet som före högskolereformen 1977 utgjorde universitetsfilialer och därefter intill 1999 var stora regionala högskolor; de representeras av 6 uppsatser i nationalekonomi (kategori 2)
- tre mindre högskolor med tillsammans 9 uppsatser i nationalekonomi (kategori 3).

Medeltal byggda på de båda bedömarnas skattningar redovisas i tabell 6.8.

Tabell 6.8 Jämförelse mellan olika kategorier av högskolor i nationalekonomi (efter korrektion för betygsskillnader)

	Kategori 1 n=15	Kategori 2 n=6	Kategori 3 n=9	p	eta
Foankn	3.23	2.27	3.22	.059	.44
Problem	3.36	2.65	2.85	.182	.35
Teori	3.35	2.28	3.12	.071	.43
Metod	3.19	2.64	2.81	.357	.28
Sluts	3.33	2.29	2.75	.073	.43
Språk	2.91	2.88	2.67	.846	.11
Totalomdöme	3.03	2.21	2.76	.219	.33
Genomsnittsomdöme	3.20	2.46	2.88	.153	.37
Svårighet	3.40	3.25	3.18	.633	.19

Medan det i historia var kategorierna 1 och 2 som turades om att få de högsta skattningarna är det i nationalekonomi kategori 2 (de nya universiteten) som skiljer sig från de övriga genom att ligga lägre i alla bedömningar utom i språk och svårighetsgrad. Skillnaderna är emellertid mindre här och räcker inte till signifikans för enskilda variabler. Inte heller medeltalen av samtliga bedömningar skiljer sig signifikant åt.

En uppdelning av analysen på bedömarna 1 och 2 har gjorts (bilagetabell 6.1). Båda bedömarna visar likartade resultat som bekräftar det samlade omdömet. Om någon kategori skiljer sig från de övriga så är det kategori 2 med lägre skattningar i praktiskt taget alla variabler utom språklig utformning och formalia samt svårighetsgrad.

6.6 Studentrekryteringen inom olika högskolekategorier

Med hjälp av registerdata ur databasen UGUR är det möjligt att skapa ett ungefärligt mått på kunskapsnivån hos de studenter som söker sig till undersökningens ämnen och institutioner.

Först redovisas de genomsnittliga grundskole- och gymnasiebetygen för studerande registrerade på B- och C-nivå läsåren 1995/96 och 1996/97 inom de tre kategorierna av högskolor (tabell 6.9).

Tabell 6.9 Genomsnittliga avgångsbetyg från grundskola och gymnasium för studerande registrerade på B- och C-nivå i nationalekonomi inom olika kategorier av högskolor

		Kategori 1	Kategori 2	Kategori 3	p-värde	eta
B95	Grundskola	4.08	3.84	3.89	.000	.22
B96		4.11	3.79	3.80	.000	.34
B95	Gymnasium	3.92	3.63	3.48	.000	.32
B96		3.93	3.44	3.43	.000	.43
C95	Grundskola	4.09	3.85	4.25	.156	.18
C96		4.07	3.80	3.81	.021	.20
C95	Gymnasium	3.90	3.62	4.34	.060	.22
C96		3.90	3.51	3.49	.003	.24
Medeltal	B	4.01	3.68	3.65	.000	.33
	C	4.00	3.70	3.97	.060	.21
Medeltal	Grundskola	4.08	3.82	3.94	.044	.24
	Gymnasium	3.92	3.55	3.68	.016	.30

Dessa värden avser inte uppsatsförfattarna eftersom de inte kan identifieras i det större registret. De kan inte heller väntas vara samlade inom just de här berörda läsåren. Det är fråga om ett allmänt mått på ämnets och högskolornas attraktivitet och som sådant antagligen rätt stabilt.

Medeltalen i tabellen bygger på följande antal:

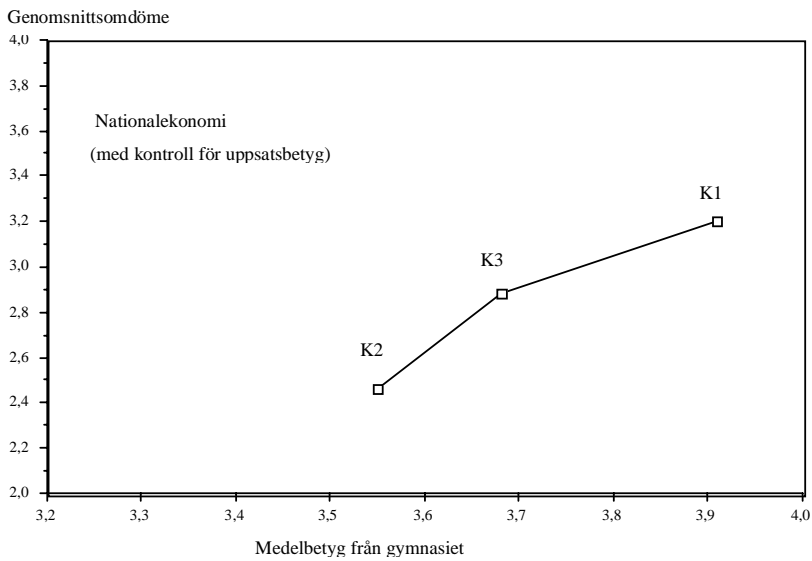
	Kategori 1	Kategori 2	Kategori 3
B-nivå	494	65	120
C-nivå	140	12	8

De minsta högskolorna har anmärkningsvärt många studerande på B-nivå i nationalekonomi, vilket torde sammanhånga med att ekonomlinjen har en stark ställning vid dessa högskolor.

Grundskolebetygen visar en spridning på 0.42 och gymnasiebetygen 0.51. Det är i förhållande till dessa värden som medeltalsskillnaderna skall bedömas, och de är stora nog för att vara signifikanta i flertalet jämförelser.

I figur 6.2 visas genomsnittsomdömena om uppsatserna (efter korrektion för uppsatsbetyget) i relation till medelbetygen från gymnasiet med B- och C-nivåerna tagna tillsammans.

Figur 6.2 Höskolekategoriernas genomsnittsomdömen i förhållande till medelbetygen från gymnasiet

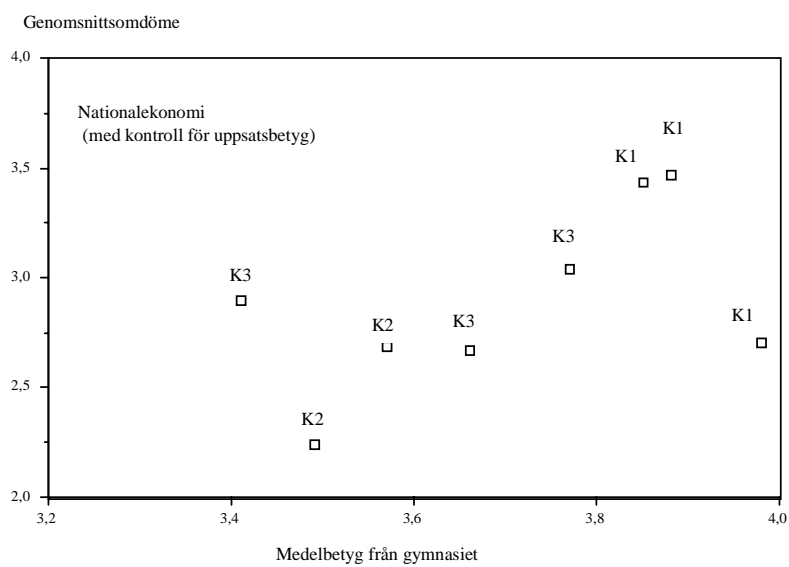


Kategori 2 ligger lägst men ungefär så som gymnasiebetygen gav anledning att förvänta. Skillnaderna i uppsatsomdömen är små.

Med samma reservationer som beträffande historia skall sambandet också visas mellan betygen och omdömena beräknade för de olika högskolorna (figur 6.3). Omdömena är korrigerade för uppsatsbetyg.

För sex av de åtta högskolorna är sambandet mycket markant mellan de båda skalorna. Två av universiteten ligger högst i båda variablerna, medan det tredje ligger avsevärt under det enligt huvudtenden förväntade genomsnittsomdömet på uppsatserna. I gengäld har en av de mindre högskolorna klart överträffat förväntningarna. Över huvud taget har kategori 3 nått bättre placeringar än kategori 2.

Figur 6.3 Högskolornas genomsnittsomdömen i förhållande till medelbetygen från gymnasiet



6.7 Uppsatsernas källmaterial

Klassificeringen av uppsatserna såsom huvudsakligen baserade på primär- eller sekundärmaterial hade ett markant samband med bedömningarna i historia. Som framgår av tabell 6.10 kan man inte vänta sig någon motsvarighet i nationalekonomi där endast två uppsatser enligt båda bedömarna baseras på primärkällor, däremot 19 på sekundärkällor. Nio uppsatser klassificerades olika av de båda bedömarna. Det finns inget samband mellan högskolekategori och val av källmaterial.

Tabell 6.10 Källmaterial för uppsatser i nationalekonomi

	Primärkällor	Sekundärkällor	Bådadera	Samtliga
Kategori 1	1	9	5	15
Kategori 2	1	4	1	6
Kategori 3		6	3	9
Samtliga	2	19	9	30

Några detaljerade medeltalsberäkningar utfördes inte, däremot variansanalyser baserade på medeltalet av samtliga bedömningar. Några signifikanta skillnader förelåg inte.

6.8 Uppsatsämnen

Uppsatsämnen i nationalekonomi kunde enligt titlarna grupperas i tre ämnesdelar på följande sätt:

- Makroekonomi 14 uppsatser
- Mikroekonomi 9 uppsatser
- Finansiell ekonomi 7 uppsatser

Det fanns inga skillnader i grupperingen mellan de tre kategorierna av högskolor. Det enda knappt signifikanta sambandet med omdömena om uppsatsernas kvalitet gällde språkbehandling och formalia, där uppsatserna i finansiell ekonomi i genomsnitt erhöll 3.50, makroekonomi 2.64 och mikroekonomi 2.61 (p-värde 0.10, *eta* 0.39).

6.9 Formella drag hos uppsatserna

Liksom i historia görs här ett försök att karakterisera uppsatserna med mått som relativt lätt kan beräknas utan datorhjälp, nämligen

- uppsatsens längd mätt genom skattning av antalet ord i huvudtexten (för detaljer se bilaga 3)
- svensk eller engelsk text
- antalet underrubriker
- antalet tabeller och figurer
- antalet referenser
- antalet fotnoter
- antalet bilagor.

I tabell 6.11 redovisas medeltal och spridning för dessa formella drag.

Tabell 6.11 Medeltal och spridning för formella drag hos uppsatserna i nationalekonomi

Antal	Medeltal	Spridning	Variationskoeff
Ord	11 383	4 471	0.39
Underrubriker	29.5	15.4	0.52
Tabeller och figurer	10.5	7.5	0.71
Referenser	22.8	12.9	0.57
Fotnoter	35.8	36.3	1.01
Bilagor	2.9	4.0	1.38

Sju av uppsatserna i nationalekonomi var skrivna på engelska, i samtliga fall sådana som var baserade på sekundärmaterial. I övrigt skilde de sig inte från de svenskspråkiga uppsatserna.

Den genomsnittliga C-uppsatsen i nationalekonomi var drygt 11 000 ord lång med spridningen 4 500. Genomgående är variationerna stora i förhållande till medeltalen, vilket innebär att fördelningarna är skeva med en stor andel nollfrekvenser och en uppåt långt utdragen kurva.

Sambandet mellan dessa mått och bedömningarna av uppsatsens kvalitet i olika avseenden redovisas i tabell 6.12.

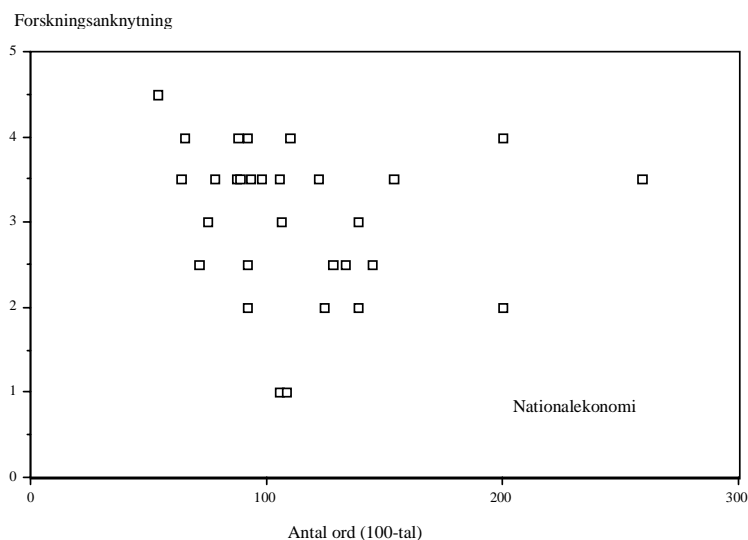
Tabell 6.12 Korrelationen mellan formella drag och uppsatsernas kvalitet enligt genomsnittet för två bedömare

	Ord	Rubrik	Tab/fig	Ref	Noter	Bilagor
Foankn	-.15	.00	.10	.11	.14	-.11
Problem	.00	.12	.24	.00	.08	-.16
Teori	-.15	.10	.42	.10	-.04	-.19
Metod	.04	.29	.43	.19	.12	-.07
Slutsats	.06	.19	.38	.18	.20	-.21
Språk	.38	.21	.22	.28	.38	-.21
Totalomdöme	.02	.09	.30	.09	.17	-.23
Genomsnittsomdöme	.03	.17	.34	.16	.17	-.20
Svårighet	.10	.03	-.28	.43	.44	-.34
Uppsatsbetyg	.17	.01	.27	-.15	.12	.04

Korrelationer på 0.35 och högre är signifikanta på 5 procentsnivån. Endast ett fåtal, framför allt avseende tabeller och figurer samt svårighetsgrad, når denna nivå. Också uppsatsbetygen har sin högsta korrelation med antalet tabeller och figurer, låtvara ej signifikant skild från noll.

Korrelationerna i tabell 6.12 förutsätter en lineär relation mellan formalia och kvalitet. Det finns skäl att kontrollera att så förhåller sig. Därför har även andragskurvor anpassats till de olika relationerna och multipla korrelationer beräknats. Som regel blir dessa något högre, men endast i två fall är den kvadratiske termen signifikant.

Figur 6.4 Sambandet mellan antal ord och omdöme om forskningsanknytning



I figur 6.4 tenderar värdena att gå nedåt till ett bottenvärde runt 14 000 ord och därefter uppåt.

6.10 Kvinnliga och manliga författare

I tabell 6.13 har medeltalen sammanställts för manliga respektive kvinnliga författare. Båda grupperna innefattar såväl ensamförfattare som samkönade författarpar. Två blandade par har ingått i analysen, dock utan att medeltal anges.

Omdömena om männens uppsatser ligger högre i alla variabler, mest så i teorimedvetenhet. Skillnaderna är dock var för sig inte signifikanta, men det är tydligt att de är större i nationalekonomi än i historia. *Eta*-värdena uppgår här i genomsnitt till 0.33 mot 0.06 i historia.

Tabell 6.13 Jämförelse mellan uppsatser i nationalekonomi med manliga och kvinnliga författare

	Män n=19	Kvinnor n=9	Bådadera n=2	p-värde	eta
Foankn	3.11	2.78	.	.509	.22
Problem	3.24	2.72	.	.416	.25
Teori	3.32	2.44	.	.068	.42
Metod	3.11	2.67	.	.499	.22
Slutsats	3.21	2.39	.	.127	.38
Språk	3.03	2.28	.	.085	.41
Totalomdöme	2.97	2.28	.	.147	.35
Genomsnittsomdöme	3.14	2.51	.	.146	.36
Svårighet	3.37	3.06	.	.170	.36

En kontroll genom tvåvägsanalys av författarnas kön mot högskolekategori resp antal författare förändrar inte bilden.

6.11 En eller två författare

I historia hade uppsatserna endast en författare. I nationalekonomi var nära hälften författade av två studerande. Bedömarna hade inte tillgång till uppgift härom. I tabell 6.14 görs en jämförelse av medeltalen för samtliga variabler.

Tabell 6.14 Jämförelse mellan uppsatser i nationalekonomi med en respektive två författare

	En författare n=16	Två författare n=14	p-värde	eta
Foankn	2.84	3.25	.218	.23
Problem	2.81	3.36	.117	.29
Teori	2.66	3.54	.011	.46
Metod	2.62	3.36	.023	.41
Slutsats	2.69	3.25	.127	.28
Språk	2.41	3.32	.006	.49
Totalomdöme	2.38	3.25	.012	.45
Genomsnittsomdöme	2.63	3.33	.017	.43
Svårighet	3.25	3.36	.581	.10
Uppsatsbetyg	1.06	1.43	.017	.43

Uppsatser med två författare har genomgående fått högre genomsnittsomdömen. För teorimedvetenhet, metod, språklig utformning och formalia samt total- och genomsnittsomdöme är skillnaderna klart signifikanta. Sambanden bestyrks av bilagetabell 6.2 där bedömarna från olika typer av högskolor redovisas var för sig.

Examinatorerna har satt signifikant högre betyg på uppsatser med två författare.

En- och tvåförfattaruppsatser kommer i ungefär lika proportioner från olika kategorier av högskolor. Likafullt kan det vara skäl att analysera variablerna författarantal och högskolekategori tillsammans såsom gjorts för källtyp och högskolekategori i historia.

Tabell 6.15 Tvåvägs variansanalys av högskolekategori och antal författare (p-värden)

	Högskole- kategori	Antal författare	Samspel
Forskningsanknytning	.049	.079	.283
Problemformulering	.276	.017	.096
Teoretisk medvetenhet	.028	.002	.306
Metod	.356	.001	.012
Genomförande och slutsatser	.073	.018	.138
Språklig utformning och formalia	.954	.003	.273
Totalomdöme	.102	.000	.016
Genomsnittsomdöme	.094	.001	.046
Svårighetsgrad	.574	.286	.445

Variansanalysen fördelar i någon mån om de olika variablernas samband. Teoretisk medvetenhet har nu samband med båda indelningsgrunderna. Författarantal har fått sina samband förstärkta. Dessutom har tre signifikanta samspel tillkommit, nämligen för metod, totalomdöme och genomsnittsomdöme. De medeltal som dessa grundas på visas i tabell 6.16.

Tabell 6.16 Högskolekategori i kombination med antal författare. Medeltal i variabler med signifikanta samspelseffekter

	Antal författare	Kategori 1	Kategori 2	Kategori 3
Antal uppsatser	1 förf	9	3	4
	2 förf	6	3	5
Metod	1 förf	3.11	1.50	2.38
	2 förf	3.08	3.67	3.50
Totalomdöme	1 förf	2.94	1.17	2.00
	2 förf	3.00	3.17	3.60
Genomsnittsomdöme	1 förf	3.09	1.60	2.38
	2 förf	3.20	3.24	3.54

För alla tre variablerna innebär samspelen att effekten av två författare uppträder bara i högskolekategorierna 2 och 3 och inte vid de gamla universiteten. Särskilt stor är skillnaden mellan en och två författare i kategori 2 (de nya universiteten). Icke signifikanta

differenser som går i samma riktning finns i alla övriga variabler utom bedömningen av uppsatsämnets svårighetsgrad.

7 Diskussion och slutsatser

Inledningsvis angavs följande syften med undersökningen:

- * Att jämföra grundutbildningens resultat vid högskolor av olika storlek och status med C-uppsatser som underlag.
- * Att basera dessa jämförelser på externa bedömningar av uppsatsernas kvalitet som utförts under experimentellt kontrollerade betingelser.
- * Att vid utvärderingen av resultaten beakta ett antal alternativa faktorer som kan antas påverka uppsatsernas kvalitet, däribland studerandeunderlagets sammansättning.
- * Att göra en bedömning av den använda metodikens tillämpbarhet som del av ett program för högskolans kvalitetskontroll.

I vilken utsträckning har dessa syften har uppnåtts? Urvalet av ämnen och högskolor diskuterades i kapitel 4. Det val som gjordes av historia och nationalekonomi förefaller att ha varit lyckat. Ämnena har visat en rik variation som gjort det möjligt att belysa olika faktorer av betydelse för studieresultaten. C-uppsatsernas kvalitet har varierat och kvaliteten har kunnat bedömas med en i dessa sammanhang tämligen tillfredsställande säkerhet.

Det går att mäta kvaliteten

Bedömningarna av uppsatsernas kvalitet omfattade sju olika delskalor. Det mest tillförlitliga måttet visade sig inte oväntat vara genomsnittet av de sju omdömena som hade starka inbördes samband. Detta gör att de relativa omdömena om varje enskild uppsats

inte blir särskilt tillförlitliga. Likafullt har de bidragit till att karakterisera olika undergrupper i materialet. Omdömena om språk och formalia intog en något mer självständig ställning än övriga. De låg högst bland omdömena i historia och relativt lågt i nationalekonomi. Teoretisk medvetenhet låg högst bland omdömena i nationalekonomi och lägst bland omdömena i historia.

Betingelserna för mätningen har kontrollerats genom att slumpen fått avgöra vilka uppsatser som skulle bedömas från olika högskolor. Valet av bedömare för en uppsats har reglerats på ett systematiskt sätt så att varje uppsats fått två bedömare som arbetat ovetande om varandra. Samtliga bedömarpar har bedömt uppsatser från alla medverkande högskolor men utan att veta varifrån en viss uppsats kom. I historia kunde man visserligen gissa sig till ursprunget av några lokalhistoriska uppsatser, men detta tycks inte ha spelat någon roll för resultaten. Även bedömarparens sammansättning med en bedömare från institutioner med fasta resurser för forskarutbildning och en från andra institutioner torde ha bidragit till tillförlitligheten och på en punkt också gett intressanta resultat.

Tillförlitligheten i denna studie skulle ha kunnat ökas, t.ex. genom att ha fler uppsatser från varje högskola och/eller fler bedömare för varje uppsats. I ett avseende hade dock en viss förbättring kunnat göras med jämförelsevis måttliga insatser, nämligen genom att redan vid urvalet ta hänsyn till högskolornas betyg på uppsatserna. En kontroll på denna punkt har nu skett i efterhand vid den statistiska behandlingen av resultaten. Helt tillfredsställande är dock ingendera metoden eftersom betygskraven sannolikt varierar mellan olika högskolor. Bättre hade varit att vid en stratifiering använda en individvariabel utanför högskolornas egen kontroll. Någon sådan fanns inte omedelbart tillgänglig.

Förklaringarna till skillnaderna mellan högskolor och ämnen har huvudsakligen fått sökas inom ramen för uppsatsbedömningen och i någon mån i enkäten till institutionerna. Särskilt lyckat var att låta bedömarna ange vad slags källor som uppsatsen baserats på, något som visade sig ha stor betydelse i historia men inte i nationalekonomi. Lika lyckat var det att välja uppsatser i nationalekonomi med en respektive två författare var för sig – detta utan att

bedömarna visste vilken kategori de tillhörde. De fick inte heller veta om författarna var kvinnor eller män.

Höga antagningskrav främjar god uppsatskvalitet

Den starkaste kontrollvariabeln kom emellertid helt utifrån och utan direkt anknytning till de enskilda uppsatsförfattarna, nämligen den genomsnittliga nivån på ämnens B- och C-studerande enligt deras slutbetyg från grundskola och gymnasium.

Metoden har på det hela taget fungerat väl, även om det skulle varit bättre om fler uppsatser hade bedömts. I ett reguljärt program för kvalitetskontroll eller ranking passar metodiken snarare för punktmässiga kontroller än för rutinmässig användning. Vad denna undersökning kunnat bidra med är snarast att visa vilken komplex operation som krävs för att belysa högskolans nettobidrag till de studerandes prestationer efter avslutad utbildning. Blotta förekomsten av sådana möjligheter kan å andra sidan vara en signal till skärpt uppmärksamhet på de möjligheter institutionerna har att påverka utfallet.

Huvudfrågan gällde om det finns några påtagliga kvalitetskillnader mellan uppsatser från universiteten och de mindre högskolorna.

Vid bearbetningen av kvalitetsbedömningarna prövades först uppdelningen på olika högskolekategorier. Resultaten tycktes bekräfta en hierarkisk ordning med universiteten högst på skalan och de minsta högskolorna lägst med avseende på uppsatsomdömena. Detta gällde historia. I nationalekonomi var skillnaderna mindre och det var de nya universiteten som tenderade att ligga lägst.

Bra uppsatser i historia bygger vanligen på primärkällor

Valet av primärkällor för uppsatsen visade sig vara av stor betydelse i historia. Uppsatser som byggde på primärmaterial var vanligast vid universiteten och fick i genomsnitt såväl högre betyg vid sina institutioner som mer positiva omdömen av bedömarna. Valet

av primärkällor betydde också mer för bedömare som kom från universiteten där man oftare såg på C-uppsatsen som en inledning på forskarbanan och ett urvalsinstrument för forskarutbildning. Dessa bedömare skilde i gengäld mindre på stora, medelstora och små högskolor än de bedömare som kom från de två senare typerna av institutioner. Uppdelningen av bedömarparen gjordes primärt för att gardera resultaten från alltför stor universitetsdominans – en risk som tydligen hade överskattats.

Den enskilde studentens val av källor påverkas rimligen både av de värderingar som råder vid institutionen och den tillgång på källmaterial som finns nära till hands. I båda avseendena torde de gamla universitetsinstitutionerna ha ett försteg. Viktigt för de studerande inom ämnet kan vara att veta att uppsatser med primärkällor tenderar att värderas högre av ämnets främsta forskningsföreträdare.

Bra uppsatser i nationalekonomi har ofta flera författare

Inom nationalekonomi är källsituationen annorlunda än i historia. Endast två uppsatser har bedömts ha primärkällor som huvudsakligt underlag. Detta ligger i ämnets natur där en studerande på C-nivå knappast kan ha tillgång till eller överblick över egentliga primärkällor utan är hänvisad till att göra jämförelser och djupdykningar i redan publicerat material. Inom ämnet ser man inte heller C-uppsatsen som något bidrag till kunskapsutvecklingen utan dess syfte är att erbjuda en kritisk och självständig fördjupning av ämnesstudierna. Därmed gäller också andra kriterier för vad som är god kvalitet. Sådana skillnader som här kunnat iaktas finns sannolikt mellan många ämnen och därmed blir det svårt att upprätthålla något generellt kvalitetsbegrepp när det gäller C-uppsatser.

I nationalekonomi är det vanligt att uppsatserna har mer än en författare och dessa uppsatser skiljer sig från sådana med en författare genom att få högre betyg både i intern och extern bedömning. De externa bedömarna visste ju inte vilka uppsatser som hade två författare och bedömde uppsatserna såsom sådana. De interna

bedömarna tycks inte ha låtit sina betyg påverkas av författarantalet fastän det var känt för dem.

I nationalekonomi är det fler manliga än kvinnliga författare medan fördelningen är mer jämn i historia. Omdömena om kvinnors och mäns uppsatser skiljer sig inte alls i historia, medan det finns en tendens i nationalekonomi att männens uppsatser fått högre omdömen och betyg.

Mera om sambanden mellan skolbetygen och uppsatskvaliteten

Vid undersökningens uppläggning eftersträvades att flera högskolor skulle ingå i varje storleks- och statusgrupp så att generaliteten i kommande resultat skulle kunna prövas. Däremot var det inte avsikten att de enskilda högskolorna skulle bli föremål för närmare granskning. antalet uppsatser inom varje högskola var alltför litet. Men när gruppvärdena sattes i relation till studenternas genomsnittliga betyg från grundskola och gymnasium framträdde sådana oregelbundenheter som gjorde det troligt att de tre kategorierna var långt ifrån homogena i sina resultat. Det gav anledning att trots allt dra in de enskilda högskolornas värden i analysen, låt vara med bibehållen anonymitet.

När så de enskilda högskolornas uppsatsmedelvärden (efter korrektion för betygsnivå) ställdes mot medelbetygsnivån från grundskola och gymnasium hos deras studerande på B- och C-nivå inom ämnena bekräftades denna förmodan att storleks- och statusgrupperna var långtifrån enhetliga i sina studieresultat.

Mest slående var att de flesta högskolornas genomsnittsomdömen var mycket starkt relaterade till genomsnittsnivån hos de studerandes tidigare skolprestationer. I historia följde värdena för sju av åtta högskolor ett mycket regelbundet mönster. De minsta högskolorna inte fullt när upp till den trend som har satts av de gamla och de nya universiteten. En högskola låg långt under de övriga.

I nationalekonomi följde värdena för sex av åtta högskolor mycket nära, medan en högskola låg klart under och en av de minsta högskolorna klart över. Även om man med hänsyn till det ringa antalet uppsatser inte bör fästa alltför stor vikt vid de enskilda hög-

skolornas placering i förhållande till studentnivån, så spräcker resultatet antagandet att högskolor tillhörande olika storleks- och statusgrupper konsekvent skiljer sig åt i studieresultat.

Det finns anledning att diskutera trenderna, som i stort sett är desamma, i båda ämnena. På kollektiv nivå förutsäger studenternas medelbetyg från gymnasiet omdömen om uppsatser på C-nivå satta av neutrala bedömare. Det är många levnadsår och många studieår som ligger emellan och ändå kan man för flertalet högskolor räkna ut vad man kan förvänta sig för omdömen om ett fåtal uppsatser. De första åren inom högskolan har uppenbarligen inte rått på de initialskillnader som studenterna förde med sig från skolan.

När man ser på de högskolor som följer den regelbundna trenden så finner man i historia att både gymnasiebetygen och uppsatsomdömena rangordnar de olika högskolorna enligt den storleks- och statushierarki som råder. Först kommer två av de gamla universiteten, sedan kommer de nya universiteten och slutligen de mindre högskolorna. Det tredje gamla universitetet låg klart lägre.

I nationalekonomi är den allmänna trenden likartad, dvs. ett starkt samband mellan studentnivå och uppsatsomdömen. Där kan man emellertid se intressanta omkastningar mellan kategorierna av högskolor. De mindre högskolorna som låg sist i historia ligger här mellan de gamla och de nya universiteten både i fråga om attraktivitet och uppsatsomdömen. En av de mindre högskolorna ligger till och med klart över förväntan när det gäller uppsatsomdömena. Förhållandena i nationalekonomi tycks dock på det hela taget uppvisa mindre variationer mellan högskolorna än i historia. Så finns det t.ex. åtskilligt fler professorer i ämnet, de är jämnare fördelade över olika typer av högskolor och de deltar oftare i uppsatshandledningen. Det förefaller också vara så att olika högskolor använder ungefär samma, ofta internationellt vedertagna kurslitteratur

Vad ligger bakom huvudtrenden i dessa resultat – ett samband på kollektiv nivå mellan skolresultat och studieresultat på C-nivå mätta genom uppsatsomdömen? En viktig faktor torde vara högskolornas och ämnenas anseende och därmed sammanhängande söktryck vid spärrad antagning. Såväl högskolornas urvalsförfarande som en högst betydande självselektion leder till att högskolor

och ämnen med olika anseende erhåller studerande med olika kvalifikationer vid inträdet. Studentgruppens sammansättning påverkar i sin tur lärarnas förväntningar och krav på de studerande. Och de studerande själva hamnar i kamratmiljöer som kan variera i intellektuell stimulans och inbördes konkurrens om betyg och andra belöningar.

Vad ligger då bakom avvikelserna från trenden? Ett universitet ligger både i historia och nationalekonomi avsevärt under den trend som gäller flertalet högskolor. Det är svårt att utan ingående fältstudier hitta förklaringar till detta. Om man även beaktar valet av källor och antalet författare sker mycket små förändringar. Skillnaderna kunde inte med framgång uttraderas med hjälp av undersökningens egna variabler.

Högskoleverkets utvärderingar av förutsättningar för magisterutbildning inom olika institutioner kan här ge vissa uppslag. De sakkunniga granskarna uppehåller sig främst vid lärarkompetens och materiella resurser. Men de hänvisar också i vissa fall till brister i organisationen av studierna, t.ex. i samband med val av uppsatsämne, handledning och examination. Sådana svårigheter uppkommer antagligen lättare vid institutioner med mycket stort antal studerande på C-nivå och därvidlag är skillnaderna mellan berörda institutioner enorma. I en institution med stark forskning och forskarutbildning kan då C-nivån och de krav man där ställer få stå tillbaka för andra och mer prestigefyllda uppgifter. Detta förklarar dock inte varför en viss sådan institution avviker från de övriga i negativ riktning. Även principiella skillnader i värderingen av olika slags uppsatser kan ligga bakom.

För den mindre högskola som avviker i positiv riktning i nationalekonomi måste andra förklaringar sökas. Olika högskolor kan avsiktligt profilera sig så att vissa ämnen uppfattas som viktigare än andra. Den avvikande högskolan ligger lågt på "förkunskaps-skalan" men har fått uppsatsomdömen i nivå med högskolor rätt högt upp på skalan. Den rimligaste förklaringen är att man där har prioriterat nationalekonomi och särskilt uppsatshandledningen som inte heller har att konkurrera med forskarutbildning om resurser och engagemang. Över huvud taget har ju de mindre högskolorna klarat

sig bättre i nationalekonomi än i historia och särskilt i förhållande till högskolorna inom mellankategorin.

Referenser

- Berg, M. (1999) Konsten att göra en ranking. Hela receptet. *Moderna Tider*, 102, 8–38.
- Bjereld, U. (red) (1996) Den goda C-uppsatsen. Statsvetenskapliga institutionerna vid Göteborgs och Uppsala universitet.
- Dracup, C. (1997) The reliability of marking on a psychology degree. *British Journal of Psychology*, 88:4, 691–709.
- Gerrevall, P. (1992) *Högskolestuderandes erfarenheter av självständigt arbete*. Pedagogiska institutionen, Lunds universitet.
- Gustafsson, J.-E., Andersson, A., & Hansen, M. (1999) Manuskript om UGUR.
- Higher Education Funding Council for England (1999) *Performance indicators in higher education*. First report of the Performance Indicators Steering Group, 99/11.
- Hounsell, D. (1984) Learning and essay writing. In Marton F., Hounsell, D., & Entwistle, N. (eds.) *The experience of learning*. Edinburgh: Scottish Academic Press.
- Högskoleverkets rapportserie:
- 1997:8 Kvalitet och förändring
 - 1997:37 Rätt att inrätta professurer
 - 1997:38 Examensrättsprövning vid elva högskolor
 - 1997:40 Tillväxt och växtverk
 - 1997:45 Magisterexamen söker identitet
 - 1998:11 Högskola i dynamisk utveckling
 - 1998:27 Vetenskapsområden
 - 1999:3 Magisterexamen söker identitet Del II

- Landauer, T.K. & Dumais, S.T. (1997) A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104:2, 211–240.
- Löfgren, C. & Ohlsson, H. (1999) What determines when undergraduates complete their theses? Evidence from two economics departments. *Economics of Education Review*, 18, 341–204.
- Marsh H.W. & Ball, S. (1981) Interjudgmental reliability of reviews for the Journal of Educational Psychology. *Journal of Educational Psychology*, 73:6, 872–880.
- Marsh H.W. & Ball, S. (1989) The peer review process used to evaluate manuscripts submitted to academic journals: Interjudgmental reliability. *Journal of Experimental Education*, 57:3, 151–169.
- Nilsson, K.-A. (1997) Extern medverkan i examinationen. Lunds universitet: Utvärderingsenheten 97:198.
- Norton, L.S. (1990) Essay writing: what really counts? *Higher Education*, 20, 411–442.
- Page, E..B (1966) The imminence of grading eassys by computer. *Phi Delta Kappan*, 238–243.
- Page, E. B. & Petersen, N.S. (1995) The computer moves into essay grading. *Phi Delta Kappan*, 76:7, 561–566.
- Penny, A.J. & Grover, C. (1996) An analysis of student grade expectations and marker consistency. *Assessment & Evaluation in Higher Education* , 21:2, 173–184.
- SOU 1992:1: Frihet, ansvar, kompetens.
- Yorke, M. m.fl. (1999) Benchmarking academic standards in the UK. *Tertiary Education and Management*, 5:1, 81–96.

Bilaga 1 Instruktion för bedömning av C-uppsatser

Inledningsvis ber vi Dig ange vilken typ av källor, antingen primära eller sekundära, som uppsatsen bygger på. Till de sekundära källorna räknar vi användningen av rent teoretiska arbeten som underlag för uppsatsen.

Bedömningar i skalan 1–5 skall göras i fråga om dels olika aspekter på uppsatsens kvalitet, dels uppsatsämnets svårighetsgrad. Dessutom skall ett helhetsomdöme avges i mera absoluta termer – från *Underkänd* till *Mycket väl godkänd*. De olika aspekterna på kvalitet kan förtydligas på följande sätt:

Forskningsanknytning

I vilken utsträckning relaterar författaren sin uppsats och sin problematik till tidigare forskning? Knyter författaren an till andra studier inom ämnesområdet? Eller är uppsatsen mer att betrakta som en isolerad produkt som inte sätts in i ett allmänvetenskapligt sammanhang ?

Problemformulering

Hur skarp är den problemformulering som ligger till grund för uppsatsen? Är författaren medveten om vilken generell problematik som de i uppsatsen ställda frågorna behandlar? Lyckas författaren precisera sitt problem på ett sådant sätt att det utmynnar i en fruktbar undersökningsidé?

Teoretisk medvetenhet

Har författaren visat teoretisk medvetenhet vid studiens genomförande?

Metod

Uppvisar författaren insikter om de fördelar och nackdelar som den valda metoden innebär? Är den valda metoden ändamålsenlig utifrån uppsatsens syfte och frågeställningar? I vilken utsträckning använder författaren den valda metoden på ett kompetent sätt?

Genomförande och slutsatser

Hur väl genomför författaren sin undersökning? Uppfylls syftet och besvaras frågeställningarna? Lyckas författaren formulera slutsatser som känns relevanta och välgrundade samt relatera slutsatserna till den generella problematik som legat till grund för uppsatsen?

Språklig utformning och formalia

Är dispositionen av uppsatsen klar, tydlig och logisk? Är språket klart, tydligt och korrekt? Är hänvisningar och belägg tillräckliga och systematiska?

Slutligen är vi tacksamma för en kortfattad kommentar och/eller karakteristik i anslutning till varje uppsats.

Bilaga 2 Formulär för bedömning av C-uppsats

Bedömare:

Uppsatsens titel:

Uppsatsen bygger huvudsakligen på primärkällor, t. ex. arkivmaterial och särskilt insamlade data

Uppsatsen bygger huvudsakligen på sekundärkällor, t.ex. forskningslitteratur och publicerad statistik.

För följande aspekter på uppsatsens kvalitet används skalan 1–5, där 1= mycket dålig och 5= mycket bra. Ringa in det värde som gäller!

Forskningsanknytning	1	2	3	4	5
Problemformulering	1	2	3	4	5
Teoretisk medvetenhet	1	2	3	4	5
Metod	1	2	3	4	5
Genomförande och slutsatser	1	2	3	4	5
Språklig utformning och formalia	1	2	3	4	5

Dessutom bedöms i skalan 1–5, där 1= mycket lätt och 5=mycket svårt,

Uppsatsämnets svårighetsgrad 1 2 3 4 5

Slutligen avges ett helhetsomdöme om uppsatsen i skalan 1–5 med graderna

- | | |
|---|--------------------|
| 1 | Underkänd |
| 2 | Med tvekan godkänd |
| 3 | Godkänd |
| 4 | Väl godkänd |
| 5 | Mycket väl godkänd |

Kommentar/karakteristik (fortsätt gärna på baksidan):

.....

.....

.....

.....

.....

.....

Bilaga 3 Beräkningen av antal ord

Uppsatserna varierade kraftigt i typografi och utformning i övrigt vilket gjorde att ett enkelt mått på t ex sidantal inte skulle ha gett någon tillförlitlig bild av uppsatsens omfång. I stället bilades ett antal mått, nämligen

- antalet ord
- antalet rubriker och underrubriker
- antalet tabeller och figurer
- antalet referenser
- antalet fotnoter
- antalet bilagor.

Av dessa var det egentligen endast antalet ord som krävde speciella regler för beräkningen. Följande förfarande tillämpades:

På var tionde sida med början på sid. 5 togs var tionde rad med början på rad 5 ut för beräkning av antalet ord på raden. Genomsnittet av antalet ord på dessa vanligen tre eller fyra rader beräknades och multiplicerades med antalet rader på sidan, vilket gav ett ordantal för sidan.

I nästa steg beräknades medelantalet ord per sida för de 3–5 sidor som uttagits för beräkning. Detta medeltal multiplicerades slutligen med antalet sidor i uppsatsen. Det sålunda skattade ordantalet i hela uppsatsen avrundades till närmaste 100-tal ord och användes i de beräkningar som redovisas i avsnitten 4.9 och 5.9.

När såsom i nationalekonomi åtskilliga tabeller och figurer ingick i texten undantogs sådana sidor från beräkningen av ord per sida och närmaste fyllda textsida medtogs i stället. Däremot ingick dessa sidor med tabeller och figurer i beräkningen av det totala sidantalet, vilket snarare gav ett mått på informationsmängden i

uppsatsen än antalet ord. Eftersom antalet tabeller och figurer bildar ett särskilt mått uppkommer en viss överlap mellan antalet ord och tabeller och figurer.

Bilagetabell 4.1 Slumpvis dragna nummer för urval av uppsatser

Historia, en förf	Antal	Startnr	Följande nummer			
Uppsala	184	3	39	75	111	147
Lund	275	47	102	157	212	267
Stockholm	256	43	94	145	196	247
Växjö	58	19	38	57		
Örebro	71	12	36	60	(extra 7, 25)	
Härnösand	38	9	22	35	(extra 6)	
Halmstad		36	9	21	33	(extra 8, 17)
Jönköping	8	2	4	6		
Nationalekon, en förf	Antal	Startnr	Följande nummer			
Uppsala	25	6	18			
Lund	59	7	26	45		
Stockholm	112	5	33	61	89	
Växjö	19	3	12			
Örebro	10	7				
Östersund	8	3	7			
Skövde	2	2				
Jönköping	2	2				
Nationalekon, två förf	Antal	Startnr	Följande nummer			
Uppsala	45	13	28	43		
Lund	22	7	18			
Stockholm	21	16				
Växjö	7	7				
Örebro	11	4	9			
Östersund	7	7				
Skövde	4	2	4			
Jönköping	6	4	7			

Bilagetabell 5.1 Jämförelse mellan olika kategorier av högskolor i historia med uppdelning på bedömare 1 och 2 i varje bedömarpar (efter korrektion för betygsskillnader)

Bedömare 1

	Kategori 1 n=15	Kategori 2 n=8	Kategori 3 n=12	p	eta
Foankn	3.32	3.50	2.51	.130	.38
Problem	3.32	3.12	2.43	.273	.28
Teori	2.64	2.60	2.05	.531	.20
Metod	2.70	3.08	2.32	.381	.24
Slutsats	3.01	2.74	2.25	.279	.28
Språk	2.62	3.22	2.83	.173	.33
Totalomdöme	2.93	3.09	2.28	.364	.25
Genomsnittsomdöme	2.94	3.05	2.38	.280	.28
Svårighet	3.40	3.36	3.01	.639	.17

Bedömare 2

	Kategori 1 n=15	Kategori 2 n=8	Kategori 3 n=12	p	eta
Foankn	3.33	3.50	2.08	.012	.50
Problem	3.48	3.36	1.73	.002	.57
Teori	3.07	3.11	1.84	.039	.43
Metod	2.60	3.10	1.93	.103	.37
Slutsats	3.71	3.73	2.21	.006	.53
Språk	3.40	3.74	2.59	.033	.44
Totalomdöme	3.54	3.61	2.08	.013	.50
Genomsnittsomdöme	3.30	3.45	2.07	.005	.54
Svårighet	3.13	3.11	3.35	.788	.12

Bilagetablell 5.2 Jämförelse mellan uppsatser i historia baserade på olika slags källor med uppdelning på bedömare 1 och 2 i varje bedömarpar

Bedömare 1

	Primär n=22	Sekundär n=12	Bådadera n=1	p-värde	eta
Foankn	3.36	2.58	.	.110	.36
Problem	3.36	2.25	.	.021	.46
Teori	2.77	1.67	.	.010	.50
Metod	3.23	1.50	.	.000	.72
Slutsats	3.14	1.75	.	.000	.71
Språk	2.91	2.58	.	.134	.34
Totalomdöme	3.32	1.58	.	.000	.68
Genomsnittsomdöme	3.16	1.99	.	.000	.62
Svårighet	3.50	2.92	.	.167	.32

Bedömare 2

	Primär n=18	Sekundär n=16	Bådadera n=1	p-värde	eta
Foankn	3.32	2.44	.	.071	.39
Problem	3.39	2.07	.	.010	.51
Teori	3.06	2.27	.	.069	.40
Metod	2.78	2.19	.	.322	.26
Slutsats	3.61	2.75	.	.134	.34
Språk	3.44	2.94	.	.312	.26
Totalomdöme	3.50	2.50	.	.062	.40
Genomsnittsomdöme	3.32	2.49	.	.063	.40
Svårighet	3.06	3.31	.	.317	.26

Bilagetabell 6.1 Jämförelse mellan olika kategorier av högskolor i nationalekonomi med uppdelning på bedömare 1 och 2 i varje bedömarpar (efter korrektion för betygsskillnader)

Bedömare 1

	Kategori 1 n=15	Kategori 2 n=6	Kategori 3 n=9	p	eta
Foankn	3.28	2.17	3.20	.075	.43
Problem	3.67	3.05	2.96	.173	.35
Teori	3.30	2.19	3.38	.071	.43
Metod	3.32	2.71	2.88	.293	.30
Slutsats	3.24	2.19	3.02	.125	.38
Språk	2.89	2.56	2.48	.745	.15
Totalomdöme	3.09	2.35	2.85	.363	.27
Genomsnittsomdöme	3.26	2.46	2.97	.148	.37
Svårighet	3.26	3.00	2.89	.532	.22

Bedömare 2

	Kategori 1 n=15	Kategori 2 n=6	Kategori 3 n=9	p	eta
Foankn	3.18	2.36	3.24	.231	.33
Problem	3.06	2.25	2.74	.395	.26
Teori	3.40	2.38	2.87	.137	.38
Metod	3.05	2.58	2.75	.639	.18
Slutsats	3.42	2.39	2.48	.084	.42
Språk	2.93	3.21	2.86	.800	.13
Totalomdöme	2.96	2.06	2.68	.321	.29
Genomsnittsomdöme	3.14	2.46	2.80	.334	.28
Svårighet	3.56	3.49	3.46	.968	.05

Bilagetabell 6.2 Jämförelse mellan uppsatser i nationalekonomi med en eller två författare med uppdelning på bedömare 1 och 2 i varje bedömarpar

Bedömare 1

	En författare n=16	Två författare n=14	p-värde	eta
Foankn	2.88	3.21	.379	.17
Problem	3.19	3.50	.383	.16
Teori	2.75	3.50	.060	.35
Metod	2.75	3.43	.039	.38
Slutsats	2.75	3.21	.241	.22
Språk	2.19	3.29	.017	.43
Totalomdöme	2.56	3.21	.087	.32
Genomsnittsomdöme	2.72	3.33	.046	.37
Svårighet	3.00	3.21	.450	.14

Bedömare 2

	En författare n=16	Två författare n=14	p-värde	eta
Foankn	2.81	3.29	.231	.22
Problem	2.44	3.21	.103	.30
Teori	2.56	3.57	.009	.47
Metod	2.50	3.29	.072	.33
Slutsats	2.62	3.29	.135	.28
Språk	2.62	3.37	.043	.37
Totalomdöme	2.19	3.29	.015	.44
Genomsnittsomdöme	2.54	3.33	.029	.40
Svårighet	3.53	3.50	.920	.02

An Academic Question - A Summary

Background

The Swedish system of higher education is composed of different categories of institutions for post-secondary education. The main demarcation line goes between research universities with both undergraduate and graduate programs and regional colleges with undergraduate programs. During recent years, a relatively large proportion of new government grants has gone to regional colleges. They have also received resources for research in selected fields. Three of the regional colleges were upgraded to universities in 1999.

In several studies, the Swedish National Agency for Higher Education has evaluated aspects of the system, including the introduction of master level programs at some regional colleges and the prerequisites for upgrading colleges to universities. Most studies have been based on expert evaluations on the *input* side, mainly the competence of the teaching staff, as well as library and physical resources. Comparisons of *output* measures showing the quality of student performances at the time of graduation are on the whole lacking to date. However, there is a lot of speculation in the public debate about possible quality differences between universities and colleges. This is the topic of the present study.

The first academic degree in Sweden (Fil.kand.) requires three years of study, equal to 120 credit points. An additional requirement is that 60 of these credits should be obtained in one of the subjects studied. In the humanities and social sciences, the third semester in a subject, i.e. the C-level study up to 60 points, includes a small thesis which is credited with 10 to 20 points out of 20. The thesis

normally reports an in-depth study of a limited problem, based on primary or secondary sources.

Design

In the present study, the C-level thesis is taken as a quality indicator of the outcome of the studies at undergraduate level, which can be compared between different categories of higher education institutions.

More specifically, the study aims (1) to compare C-level theses from different categories of higher education institutions as evaluated by external examiners, (2) to perform these comparisons under strictly controlled conditions, (3) to relate the results to factors which, in addition to teaching, may influence the quality of the theses, and (4) to judge the applicability of the methodology if used as a component in a regular program for quality assurance of instruction. Instruction is taken here in a broad sense. In addition to teaching, it includes such tasks as the organization of thesis work, assistance with the choice of topics, guidance during work and feedback on preliminary manuscripts.

Two disciplines were selected for investigation. These were history from the humanities and economics from the social sciences. Both disciplines have a central position within their fields; they have a long academic tradition and are regarded as having a well-established practice of evaluation.

For each discipline, eight departments were selected: three from the old and large universities, two from the regional colleges which were upgraded to universities in 1999, and three from minor regional colleges.

Lists were obtained from the departments of the C-level theses passed during the period spring term 1996 to spring term 1998. Five theses were randomly drawn from each university department and three to five from each of the other institutions. In total, 35 history theses and 30 economics theses were evaluated.

Each thesis was rated by two external examiners who did not know whom the other rater was. Three pairs of examiners were

engaged for each discipline. In each pair, one examiner came from a department with a doctoral program and one from a department with only undergraduate instruction. All in all, there were twelve examiners, and each examiner rated 10 to 12 theses.

Ratings on a scale of 1 to 5 were obtained for the following six aspects of quality: relation to prior research, definition of problem, understanding of theory, methodology, procedure and conclusions, and finally, quality of language and formalities. In addition, a summative judgement was given about the overall quality of the thesis on a scale from “fail” through “pass” to “pass with distinction”.

In addition, the examiners rated the difficulty of the topic chosen and judged whether the thesis was based mostly on primary or secondary sources.

These nine ratings from each examiner were used in the statistical analysis of the results.

The ratings

In both subjects, the average rating was close to the midpoint 3 with a standard deviation of about one step on the rating scale. The history theses got their highest ratings for relation to prior research and quality of language formalities and the lowest ratings were awarded for understanding of theory and the methodology. The economics theses were rated highest for problem definition and theory and lowest for quality of language – a profile of ratings almost opposite to that for the history theses.

The size of the inter-rater correlation varied between aspects. The average rating over all aspects was the most reliable one with an inter-rater correlation of 0.56 in history and 0.58 in economics. This corresponds to a reliability for the two raters taken together of 0.72 and 0.73 and compares fairly well with other research on the reliability of ratings. The average external rating in history correlated 0.48 with the grade (on a two-point scale) that the theses obtained in their own department. In economics, this correlation

was only 0.31 but somewhat higher for two of the separate aspects; problem definition and methodology (0.38).

Two out of 35 history theses were unanimously rated “fail”. In economics, both raters placed three out of 30 theses in the “fail” category.

Comparisons between categories of institutions

As mentioned above, three categories of institutions were compared:

- old and large universities (category 1 with 15 theses both in history and economics);
- new universities which at the time of data collection were large regional colleges (category 2 with 8 theses in history and 6 in economics);
- minor colleges (category 3 with 12 theses in history and 9 in economics).

In history, the average ratings for categories 1 and 2 were rather equal but clearly higher than those for category 3. With the exception of the language rating, the differences between categories 1 + 2 and 3 amounted to about one unit on the 5-point scale, which corresponds to one unit or more of the standard deviation and is clearly significant. In the difficulty ratings, there were no differences at all between the categories, which supports the interpretation of the other differences as being real differences in quality.

In economics, the differences were smaller than in history and the rank order between categories was different. First came the old and large universities (category 1), then the minor regional colleges (category 3) and last the new universities (category 2). Also here, the smallest differences were found for language and difficulty.

In trying to explain the differences, we were able to use a large data base comprising all individuals in Sweden born in the years

1972–1979. It contained, among other things, information on school marks from upper-secondary education and on participation in higher education. This made it possible to compute for each department a set of student input measures, more specifically the average upper-secondary school marks for those registered at B- and C-level during the academic years 1995/96 and 1996/97. This is a measure at the collective level and not attributable to the individual authors of the theses evaluated. But as a measure of the competitiveness of the departments, it turned out to function quite well.

Figure 1 Relation between the mean upper-secondary school marks and the average ratings

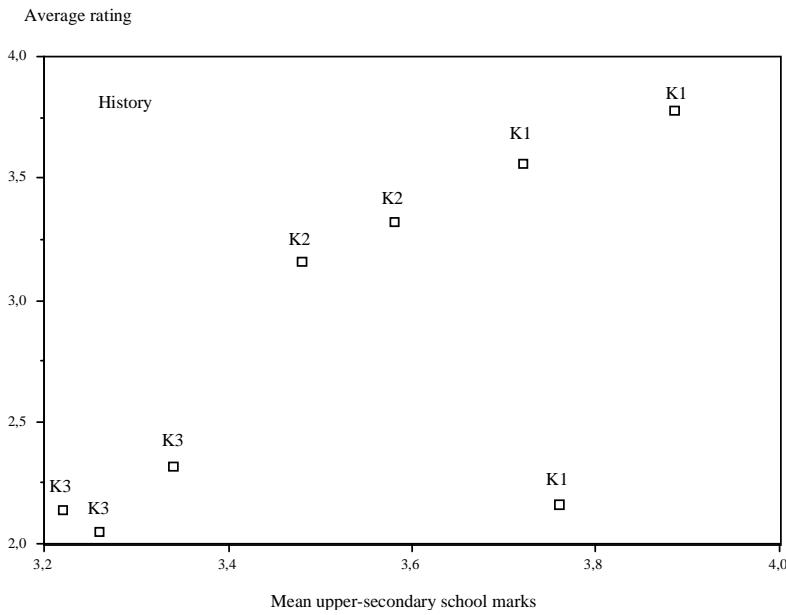


Figure 1 shows the relation between the average upper-secondary school marks for each department and the average rating of the history theses from the departments. K1–K3 refer to the different categories. Seven of the eight departments are found on an imagined curve with two of the university departments on top in both variables, then followed by the category 2 departments and finally the minor regional colleges at the bottom. For one of the university departments, however, the average rating is found at the same level as the category 3 ratings in spite of the department getting high-achieving students from upper-secondary school.

Figure 2 Relation between the mean upper-secondary school marks and the average ratings

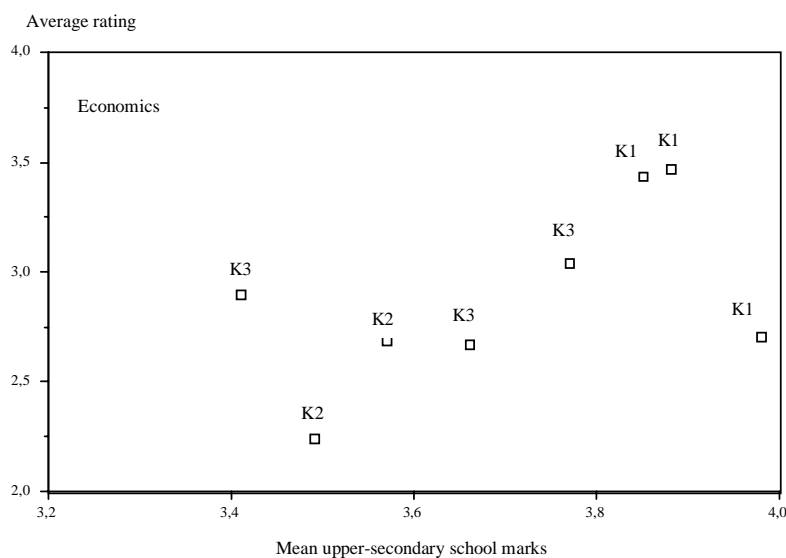


Figure 2 shows the same kind of relation for economics. Six departments are found on an imagined straight line. One K1 department – the highest according to school marks – lies clearly below the expected ratings and one K3 department – with low school marks – lies above the expected ratings.

The collective measures of the quality of student input reflect variations in reputation and competition among the institutions. Both self-selection and selection based on school marks or aptitude tests create differences in student capability between institutions. These differences do not only indicate the learning capacity of the students but they also influence the teachers' expectations and demands and the intellectual stimulation and competition in the peer group. These in turn affect the quality of outcomes.

Such a sequence of influences would explain the covariation between the collective measure of student input and the quality

rating of the theses. But what about the deviations from the general trend? Are the negative deviations an indication of low quality of instruction, more lenient evaluation practices, or a different concept of quality from that of the external examiners? Is the positive deviation of one economics program a sign of particularly high quality of instruction or motivation? Or are the deviations just sampling accidents due to few observations? This technical explanation, however, is not likely for history since four out of the five theses had obtained a “pass with distinction”.

Other sources of variation

In a questionnaire study prior to the collection of ratings, it was found that history and economics departments had somewhat different views on the objectives of C-level theses. History tended to regard the thesis more as an introduction – or even a contribution – to research, while economics viewed it as an in-depth continuation of the basic study of the discipline. This difference between the subjects is reflected in two variables that, besides student input, turned out to be powerful sources of variation in the ratings.

In history (but not in economics), the choice between primary and secondary sources for the theses was important. Primary sources were more often used in the university departments. Coincidentally, theses based on primary sources were on average rated more than the one standard deviation higher than those based on secondary sources such as books and articles. This difference was particularly large among the examiners from departments with doctoral programs, who tended to see the C-level thesis as an instrument for selection to graduate studies.

In economics (but not in history), a large proportion of the theses had two authors. The external examiners did not know which ones. Co-authored theses were on average rated about 0.8 standard deviation higher than those with a single author. They also had obtained higher marks in their departments. It seems that individual performance was less important than in history.

In history, no gender differences were found in the ratings. In economics, there was a tendency for male authors to get higher ratings than females but the differences did not reach statistical significance.

In an analysis of formal traits of the theses, it was found that history theses were longer than economics theses – on average about 15,000 vs. 11,000 words. In history, the number of references had a substantial correlation with the ratings (average 0.64). In economics, the highest correlation was for the number of tables and figures, at most 0.43 with the rating of methodology.

Conclusions

The main objective of the study was to compare the quality of output from undergraduate instruction in higher education institutions of different categories. Such differences were found, and on the surface they followed the traditional order of the institutions, ranking old and large universities higher than regional colleges. However, upon closer study, it was necessary to make a number of reservations.

Firstly, the rank order differed to a large extent between the two disciplines. In history, the large regional colleges, which in 1999 became universities (category 2), performed better than the smaller ones (category 3). In economics, this order was reversed. This might be due to a narrower program of the minor institutions, putting priority on economics and other application-oriented subjects, in contrast to the university type of ambitions of the larger ones.

Secondly, in both subjects, one of the universities deviated markedly from the other two in getting low ratings compared to what could have been expected from the average capability of the students. In economics, one of the minor colleges, on the other hand, performed better than expected. Taken together, these findings indicate that the categorization according to size and status is problematic. The outcome is likely to differ between disciplines and a large part of the deviation depends on differences in reputation

and student input rather than on resources and quality of instruction. It is the deviation from expected output – the value added – that should be taken as a measure of quality rather than the gross output.

In addition to the differences in student output, the ratings were influenced by the choice of source materials for the theses (history) and the number of authors (economics). Behind these factors, differences in the objectives as viewed in the two disciplines could be traced. Where research preparation is highly valued and the thesis is an instrument for selection, individual skills in treating primary data are important. Where in-depth understanding of theory is a priority, it does not matter so much if one or two students are responsible for the product. However, it is uncertain whether the departments and their students are aware of the influence of such mechanisms in the evaluation of their performance. Maybe they should be.

A secondary objective of the study was to develop and try out a methodology for studies of the quality of output and value added of instruction. It seems that the rating of C-theses by external examiners is a feasible method for studying student performance at the end of undergraduate instruction. But it requires at least two examiners that independently rate representative samples of theses, preferably a somewhat larger number of theses than used here.

In order to study value added, it is necessary to have a measure of student input. In this study, a collective measure of upper-secondary school marks functioned relatively well, but there are few occasions where one can find such a measure in the form of pre-collected information and so this also has to be collected, which would make the task more cumbersome and time-consuming. Therefore, a routine application of the method as part of a system of quality assurance is hardly likely, but it can be used when other signals indicate that there may be a large inequalities in the evaluation of student performance.

Förteckning över ESO:s rapporter

Expertgruppen för studier i offentlig ekonomi (ESO) är en kommitté (B 1981:03) under Finansdepartementet.

I Ds-serien har ESO tidigare utgivit följande rapporter

1982

Perspektiv på besparingspolitiken (Ds B 1982:3)

Inkomstfördelningseffekter av livsmedelssubventioner (Ds B 1982:7)

Perspektiv på budgetunderskottet, del 1. Budgetunderskottens teori och politik. Statens budgetfinansiering och penningpolitiken (Ds B 1982:9)

Offentliga tjänster på fritids-, idrotts- och kulturområdena (Ds B 1982:10)

Ökad produktivitet i offentlig sektor – en studie av de allmänna domstolarna (Ds B 1982:11)

1983

Staten och kommunernas expansion – några olika styrmedel (Ds Fi 1983:3)

Enhetligt barnstöd? (Ds Fi 1983:6)

Perspektiv på budgetunderskottet, del 2. Fördelningseffekter av budgetunderskott. Hushållsekonomi och budgetunderskott (Ds Fi 1983:7)

Minskad produktivitet i offentlig sektor – en studie av PRV (Ds Fi 1983:18)

Driver subventioner upp kostnader? – prisbildningseffekter av statligt stöd (Ds Fi 1983:19)

Administrationskostnader för några transfereringar (Ds Fi 1983:22)

Generellt statsbidrag till kommuner – modellskisser (Ds Fi 1983:26)

Produktivitet i privat och offentlig tandvård (Ds Fi 1983:27)

Perspektiv på budgetunderskottet, del 3. Budgetunderskott, portföljval och tillgångsmarknader. Modellsimuleringar av offentliga besparingar m.m. (Ds Fi 1983:29)

Fördelningseffekter av kommunal barnomsorg (Ds Fi 1983:30)

Administrationskostnader för våra skatter (Ds Fi 1983:32)

1984

Vem utnyttjar den offentliga sektorns tjänster? (Ds Fi 1984:2)

Perspektiv på budgetunderskottet, del 4. Budgetunderskott, utlandsupplåning och framtida konsumtionsmöjligheter. Budgetunderskott, efterfrågan och inflation (Ds Fi 1984:3)

Konstitutionella begränsningar i riksdagens finansmakt – behov och tänkbara utformningar (Ds Fi 1984:7)

Är subventioner effektiva? (Ds Fi 1984:8)

Marginella expansionsstöd – ekonomiska och administrativa effekter (Ds Fi 1984:12)

Transfereringar och inkomstskatt samt hushållens materiella standard (Ds Fi 1984:17)

Parlamentet och statsutgifterna – hur finansmakten utövas i nio länder
(Ds Fi 1984:18)

Återkommande kostnads- och prestationsjämförelser – en metod att främja effektiviteten i offentlig tjänsteproduktion (Ds Fi 1984:19)

1985

Statsskuldräntorna och ekonomin – effekter på inkomst- och förmögenhetsfördelningen samt på den samlade efterfrågan i samhället (Ds Fi 1985:2)

Produktions-, kostnads- och produktivitetens utveckling inom offentligt bedrivna hälso- och sjukvård 1960-1980 (Ds Fi 1985:3)

Produktions-, kostnads- och produktivitetens utveckling inom den sociala sektorn 1970 - 1980 (Ds Fi 1985:4)

Transfereringar mellan den förvärvsarbetsande och den äldre generationen (Ds Fi 1985:5)

Frivilligorganisationer – alternativ till den offentliga sektorn? (Ds Fi 1985:6)

Organisationer på gränsen mellan privat och offentlig sektor – förstudie (Ds Fi 1985:7)

Produktions-, kostnads- och produktivitetens utveckling inom vägsektorn (Ds Fi 1985:9)

Skatter och arbetsutbud (Ds Fi 1985:10)

Sociala avgifter – problem och möjligheter inom färdtjänst och hemtjänst (Ds Fi 1985:11)

Egen regi eller entreprenad i kommunal verksamhet – möjligheter, problem och erfarenheter (Ds Fi 1985:12)

1986

Produktions-, kostnads- och produktivitetens utveckling inom armén och flygvapnet (Ds Fi 1986:1)

Samhällsekonomiskt beslutsunderlag – en hjälp att fatta beslut (Ds Fi 1986:2)

Effektivare sjukvård genom bättre ekonomistyrning (Ds Fi 1986:3)

Effekter av statsbidrag till kommuner (Ds Fi 1986:7)

Byråkratiseringstendenser i Sverige (Ds Fi 1986:8)

Svensk inkomstfördelning i internationell jämförelse (Ds Fi 1986:12)

Offentliga tjänster – sökarljus mot produktivitet och användare (Ds Fi 1986:13)

Kostnader och resultat i grundskolan – en jämförelse av kommuner (Ds Fi 1986:14)

Regleringar och teknisk utveckling (Ds Fi 1986:15)

Socialbidrag. Bidragsmottagarna: antal och inkomster. Socialbidragen i bidragssystemet (Ds Fi 1986:16)

Produktions-, kostnads- och produktivitetens utveckling inom den offentligt finansierade utbildningssektorn 1960 - 1980 (Ds Fi 1986:17)

Offentliga utgifter och sysselsättning (Ds Fi 1986:29)

1987

Att leva på avgifter – vad innebär en övergång till avgiftsfinansiering?
(Ds Fi 1987:2)

Vägar ut ur jordbruksprisregleringen – några idéskisser (Ds Fi 1987:4)

Kvalitetsutvecklingen inom den kommunala äldreomsorgen 1970 - 1980
(Ds Fi 1987:6)

Produktkostnader för offentliga tjänster – med tillämpningar på kulturområdet
(Ds Fi 1987:10)

Integrering av sjukvård och sjukförsäkring (Ds Fi 1987:11)

1988

Kvalitetsutvecklingen inom den kommunala barnomsorgen (Ds 1988:1)

Från patriark till part – spelregler och lönepolitik för staten som arbetsgivare (Ds 1988:4)

Produktivitetsutvecklingen i kommunal barnomsorg 1981-1985 (Ds 1988:5)

Prestationer och belöningar i offentlig förvaltning (Ds 1988:18)

Subventioner i kritisk belysning (Ds 1988:28)

Hur stor blev tvåprocentaren? Erfarenheter från en besparingsteknik
(Ds 1988:34)

Effektiv realkapitalanvändning i kommuner och landsting (Ds 1988:51)

Alternativ i jordbrukspolitiken (Ds 1988:54)

Kvalitet och kostnader i offentlig tjänsteproduktion (Ds 1988:60)

Vad kan vi lära av grannen? Det svenska pensionssystemet i nordisk belysning
(Ds 1988:68)

1989

Hur man mäter sjukvård – exempel på kvalitets- och effektivitetsmätning
(Ds 1989:4)

Lönestrukturen och den "dubbla obalansen" – en empirisk studie av löneskillnader mellan privat och offentlig sektor (Ds 1989:8)

Beställare-utförare – ett alternativ till entreprenad i kommuner (Ds 1989:10)

Vad ska staten äga? De statliga företagen inför 90-talet (Ds 1989:23)

Statsbidrag till kommuner: allt på en check eller lite av varje?
En jämförelse mellan Norge och Sverige (Ds 1989:26)

Produktivitetsmätning av folkbibliotekens utlåningsverksamhet (Ds 1989:42)

Bostadsstödet – alternativ och konsekvenser (Ds 1989:47)

Kommunal förmögenhetsförvaltning i förändring: citykommunerna Stockholm, Göteborg och Malmö (Ds 1989:56)

Hur ska vi få råd att bli gamla? (Ds 1989:59)

Arbetsmarknadsförsäkringar (Ds 1989:68)

1990

Bostadskarriären som förmögenhetsmaskin (Ds 1990:29)

Skola? Förskola? Barnskola! (Ds 1990:31)

Statens dolda kapital. Aktivt ägande: exemplet Vattenfall (Ds 1990:36)
Sjukvårdskostnader i framtiden – vad betyder åldersfaktorn? (Ds 1990:39)
Läkemedelsförmånen (Ds 1990:81)

1991

Målstyrning och resultatuppföljning i offentlig förvaltning (Ds 1991:19)
Metoder i forskning om produktivitet och effektivitet med tillämpningar på offentlig sektor (Ds 1991:20)
Vad kostar det? Prislista för statliga tjänster (Ds 1991:26)
Det framtida pensionssystemet – två alternativ (Ds 1991:27)
Skogspolitik för ett nytt sekel (Ds 1993:31)
Prestationsbaserad ersättning i hälso- och sjukvården – vad blir effekterna? (Ds 1991:49)
Ostyriga projekt – att styra och avstyra stora kommunala satsningar (Ds 1991:50)
Marginal effekter och tröskeffekter – barnfamiljerna och barnomsorgen (Ds 1991:66)
SJ, Televerket och Posten – bättre som bolag? (Ds 1991:77)

1992

Skatteförmåner och särregler i inkomst- och mervärdesskatten (Ds 1992:6)
Frihandeln ett hot mot miljöpolitiken – eller tvärtom? (Ds 1992:12)
Växthuseffekten – slutsatser för jordbruks-, energi- och skattepolitiken (Ds 1992:15)
Fattighetsfällor (Ds 1992:25)
Vad vill vi med socialförsäkringarna? (Ds 1992:26)
Statliga bidrag – motiv, kostnader, effekter? (Ds 1992:46)
Hur bra är vi? Den svenska arbetskraftens kompetens i internationell belysning (Ds 1992:83)
Slutbudsmetoden – ett sätt att lösa tvister på arbetsmarknaden utan konflikt (Ds 1992:88)
Kommunerna som företagsägare – aktiv koncernledning i kommunal regi (Ds 1992:111)
Press och ekonomisk politik – tre fallstudier (Ds 1992:124)
Statsskulden och budgetprocessen (Ds 1992:126)

1993

Presstödet effekter – en utvärdering (Ds 1993:20)
Hur välja rätt investeringar i transportinfrastrukturen? (Ds 1993:22)
Lönar sig förebyggande åtgärder? Exempel från hälso- och sjukvården och trafiken (Ds 1993:37)
Social Security in Sweden and Other European Countries – Three Essays (Ds 1993:51)
Idrott åt alla? – Kartläggning och analys av idrottsstödet (Ds 1993:58)

1994

Att rädda liv – Kostnader och effekter (Ds 1994:14)

Varför kulturstöd? – Ekonomisk teori och svensk verklighet (Ds 1994:16)

Kvalitets- och produktivetsutvecklingen i sjukvården 1960 - 1992

(Ds 1994:22)

Kvalitet och produktivitet – Teori och metod för kvalitetsjusterade produktivetsmått (Ds 1994:23)

Den offentliga sektorns produktivetsutveckling 1980-1992 (Ds 1994:24)

Det offentliga stödet till partierna – Inriktning och omfattning (Ds 1994:31)

Den svenska insolvensrätten – Några förslag till förbättringar inom konkurshandlingen m.m. (Ds 1994:37)

Budgetunderskott och statsskuld – Hur farliga är de? (Ds 1994:38)

Bensinskatteförändringars effekter (Ds 1994:55)

Skolans kostnader, effektivitet och resultat – En branschstudie (Ds 1994:56)

Den offentliga sektorns produktivetsutveckling 1980 - 1992. Bilagor

(Ds 1994:71)

Valfrihet inom skolan – Konsekvenser för kostnader, resultat och segregation (Ds 1994:72)

En Social Försäkring (Ds 1994:81)

Fördelningseffekter av offentliga tjänster (Ds 1994:86)

Nettokostnader för transfereringar i Sverige och några andra länder

(Ds 1994:133)

Skatter och socialförsäkringar över livscykeln – En simuleringsmodell

(Ds 1994:135)

En effektiv försvarspolitik? – Fredsvinst, beredskap och återtagning

(Ds 1994:138)

1995

Försvarets kostnader och produktivitet (Ds 1995:10)

Företagsstödet – Vad kostar det egentligen? (Ds 1995:14)

Hushållning med knappa naturresurser – Exempler allemansrätten, fjällen och skotertrafik i naturen (Ds 1995:15)

Vad blev det av de enskilda alternativen? En kartläggning av verksamheten inom skolan, vården och omsorgen (Ds 1995:25)

Kostnader, produktivitet och måluppfyllelse för Sveriges Television AB

(Ds 1995:31)

Hushållning med knappa naturresurser – Exemplet sportfiske (Ds 1995:47)

Invandring, sysselsättning och ekonomiska effekter (Ds 1995:68)

Generationsräkenskaper (Ds 1995:70)

Kapitalets rörlighet – Den svenska skatte- och utgiftsstrukturen i ett integrerat Europa (Ds 1995:74)

1996

Hur effektivt är EU:s stöd till forskning och utveckling? –

En principdiskussion (Ds 1996:8)

Reglering som spel – Universiteten som förebild för offentliga sektorn?

(Ds 1996:18)

Nästa steg i telepolitiken (Ds 1996:29)

Kan myndigheter utvärdera sig själva? (Ds 1996:36)

Novemberrevolutionen – Om rationalitet och makt i beslutet att avreglera kreditmarknaden 1985 (Ds 1996:37)

Samhällets stöd till barnfamiljerna i Europa (Ds 1996:49)

Kommunerna och decentraliseringen – Tre fallstudier (Ds 1996:68)

1997

Jordbruksstödet – efter Sveriges EU-inträde (Ds 1997:46)

Egenföretagande och manna från himlen (Ds 1997:71)

Lönar sig arbete? (Ds 1997:73)

Ramar, regler, resultat - vem bestämmer över statens budget (Ds 1997:79)

Fisk och Fusk - Mål, medel och makt i fiskeripolitiken (Ds 1997:81)

1998

Vad kostar en ren? En ekonomisk och politisk analys (Ds 1998:8)

Kommuner Kan! Kanske !- Om kommunal välfärd i framtiden (Ds 1998:15)

Arbetsförmedlingarna - Mål och drivkrafter (Ds 1998:16)

Att se till eller titta på - om tillsynen inom miljöområdet (Ds 1998:50)

Regeringskansliet inför 2000-talet - rapport från ett ESO-seminarium (Ds 1998:56)

Kommittéerna och Bofinken - Kan en kommitté se ut hur som helst?

(Ds 1998:57)

Staten och bolagskapitalet - om aktiv styrning av statliga bolag (Ds 1998:64)

1999

Med backspegeln som kompass - om stabiliseringspolitiken som läroprocess (Ds 1999:9)

Rapport från ett ESO-seminarium - Med backspegeln som kompass (Ds 1999:27)

Att ta sig ton - om svensk musikexport 1974-1999 (Ds 1999:28)

Bostad sökes - en ESO-rapport om de hemlösa i folkhemmet (Ds 1999:46)

Att reda sig själv - en ESO-rapport om rederier och subventioner (Ds 1999:47)

Att snubbla in i framtiden - en ESO-rapport om statliga utveckling och avveckling (Ds 1999:49)

Regionalpolitiken - en ESO-rapport om tro och vetande (Ds 1999:50)

Samhällets stöd till de äldre i Europa (Ds 1999:61)

Hederlighetens pris - en ESO-rapport om korruption (Ds 1999:62)

En akademisk fråga - en ESO-rapport om ranking av C-uppsatser (Ds 1999:65)