

Likheter och skillnader i bedömning av examensarbeten

En intervjustudie med fyra examinatorer på lärarprogrammet

Ilona Rinne (Göteborgs universitet / Institutionen för didaktik och pedagogisk profession)

Abstract

Syftet med denna studie är att få kunskap om hur examinatorer resonerar kring bedömning och betygsättning av examensarbeten. Studien bygger på en kombination av semistrukturerade intervjuer och så kallad ”think aloud” metod med fyra examinatorer på grundlärarprogrammet. Det empiriska underlaget utgörs av examinatorernas resonemang om bedömning och betygsättning av fem examensarbeten. Resultatet visar att ingen av uppsatserna har fått samma betyg av samtliga examinatorer. Både likheter och skillnader synliggörs i examinatorernas resonemang. Likheterna visar sig i resonemanget kring bedömning och betygsättning medan skillnaderna uppkommer främst vid betygsättning. Med utgångspunkt i examinatorernas resonemang väcker studien frågor kring betygsättningens likvärdighet.

Nyckelord: *bedömning, betygsättning, examensarbeten, likvärdighet*

Inledning

Forskning om hur universitetslärare faktiskt går tillväga när de bedömer är ännu begränsad. Som examinator på lärarprogrammet har jag ofta känt mig osäker när jag har bedömt och betygsatt examensarbeten. Betygskriterierna ger stort tolkningsutrymme och jag är inte säker på att studenterna förstår innebörden av dem. Många av mina kollegor vittnar om liknande upplevelser. Vad vet vi egentligen om hur vi bedömer? Bedömer vi lika? Vilka ställningstaganden gör vi vid bedömning och betygsättning? För att få mer kunskap inom detta ämne har jag intervjuat fyra examinatorer på lärarprogrammet i samband med bedömning av fem examensarbeten.

Syfte och frågeställning:

Syftet med studien är att, *med utgångspunkt i fyra examinatorers resonemang kring bedömning och betygsättning av fem examensarbeten på lärarprogrammet, få kunskap om hur examinatorer resonerar kring bedömning betygsättning.*

Studien utgår ifrån följande forskningsfråga:

Vilka likheter och skillnader kan synliggöras i examinatorernas bedömning och betygsättning av examensarbeten?

Begreppsdefinition

Fokus i denna studie riktas mot *bedömning* och *betygsättning*. Begreppen är nära sammanlänkade, dock inte synonyma. Betygsättning förutsätter bedömning, medan bedömning inte behöver resultera i betygsättning. Bedömning kan dessutom ske i sammanhang som inte är pedagogiska, medan betygsättning är en pedagogisk aktivitet.

Forskning inom området

Forskning kring bedömning och betygsättning är omfattande och har intresserat en rad forskare på både nationell och internationell nivå. En mängd aspekter av betygsättning har fokuserats, däribland tolkningen av betygs-kriterier och gränsdragningar mellan de olika betygsstegen (Hyltegren, 2014; Selghed, 2010; Sadler, 2009; 2005), samt vikten av att involvera studenterna i bedömningsprocessen, exempelvis genom dialog (Bloxham & Campbell, 2010). Inte minst anses tydliga betygs-kriterier samt större transparens i bedömningsprocessen gagna studenters lärande (Bergqvist, 2015; Högskoleverket, 2008). Många forskare är emellertid tveksamma till huruvida det överhuvudtaget är möjligt att göra kriterierna tydliga (Hyltegren; Selghed; Sadler), vilket i sin tur får konsekvenser för betygens likvärdighet och studenternas rättssäkerhet.

Betygens likvärdighet och studenternas rättssäkerhet kan på nationell nivå även äventyras av det faktum att varje enskilt svenskt lärosäte har frihet att bestämma vilken betygsskala som tillämpas i det målrelaterade betygssystemet. Före utbildnings- och examensordningens införande 2007 diskuterades ett förslag om ett nytt sjugradigt poängsystem, ECTS (European Credit Transfer System), som var baserat på en relativ betygsskala (Högskoleverket, 2008).

Skalan antogs dock inte, men diskussionen medförde i sin tur att en mängd betygsskalor¹ introducerades på landets lärosäten. Sju gradiga betygsskalor liknande ECTS har därefter börjat tillämpas vid exempelvis Stockholms universitet, Kungliga Tekniska högskolan samt vid Högskolan i Gävle. Dessa betygsskalor är dock inte relativa, utan målrelaterade. Lärosätena har även frihet att bestämma vilken betygsskala ska vara gällande för varje enskild kurs², vilket ska anges i kursplanen. Tanken är att betygsskalorna ska kombineras med betygskriterier. Enligt högskoleförordningen finns det emellertid inga krav på att betygskriterierna ska anges för varje kurs samt om de ska vara bindande för examinatorerna och studenterna. Lärosätena måste därmed själva ta ställning till vilken ställning samt vilken rättslig status betygskriterierna ska ha (Högskoleverket). Detta är angeläget inte minst ur en rättssäkerhetsaspekt, eftersom betygsättning är en form av myndighetsutövning, som inte får ”överklagas till domstol eller någon annan överklagandeinstans” (ibid., s. 17). Examinatorn har därmed både stor makt och ett stort ansvar när hen använder sin befogenhet att sätta betyg.

Varje lärosäte har alltså frihet att bestämma över vilket betygssystem som tillämpas och varje fakultet får ta ställning till huruvida betygskriterier ska utformas. Det förefaller inte finnas någon konsensus vad gäller betygskriteriernas betydelse för betygsättning. Däremot finns det uppfattningar om att skriftliga betygskriterier bidrar till att tydliggöra både för läraren och för studenten vad som är viktigt och vad som bedöms (Bergqvist, 2015). En fråga som uppstår i sammanhanget är hur betygskriterier kan tillämpas. Sadler (2009) beskriver två former av bedömning: *analytisk* och *holistisk*. Kortfattat kan den holistiska beskrivas som en helhetsbedömning av ett arbete. Samtliga kriterier måste vara uppfyllda till en miniminivå. Vissa delar i arbetet kan vara av särdeles god kvalitet, vilket i sin tur tas i beaktning i slutbedömningen (*disjunctive grading rule*). Vid holistisk bedömning kan alltså kompensatoriskt förfarande tillämpas, vilket innebär att vissa tillkortakommanden kan bortses om de övervägs av slående kvaliteter inom arbetets övriga delar. Den analytiska bedömningen utgår ifrån att bedömningen görs först av separata kvalitativa delar och att det slutgiltiga betyget baseras på en sammanslagning av de separata delarna. Varje kriterium väger lika tungt och samtliga kriterier måste uppfyllas för att ett visst betyg ska uppnås (*conjunctive grading rule*). Sadler är tveksam till att mer explicita bedömningsmodeller gagnar

¹ Enligt högskoleförordningen ska betygen underkänd, godkänd, väl godkänd användas om inget annat betygssystem föreskrivs. Sju gradiga betygsskalor Stockholms universitet, Kungliga Tekniska Högskolan samt Högskolan i Gävle tillämpar

² Exempelvis vid Göteborgs universitet tillämpas betygsskalan U/G eller U/G/VG.

betygsättningen, eftersom "fixa" och "tydliga" kriterier riskerar att förbise helheten och komplexiteten i bedömningen. Betygskriteriers möjliga tillämpning är relevant för följande studie och kommer att diskuteras med utgångspunkt i studiens resultat.

Genomförande

Nedan redogörs för studiens metod och genomförande. En kortfattad presentation av examinatorerna samt det bedömningsunderlag som tillämpades i studien kommer att ges.

Metod

Studien bygger på semistrukturerade intervjuer med fyra examinatorer. Intervjuerna utgår ifrån en intervjuguide (bilaga 2) och kombineras med en så kallad "think aloud" metod (Charters, 2003) där examinatorer resonerar högt kring bedömning och betygsättning av fem examensarbeten. Metoden tillämpas inom en mängd områden, såsom medicin, teknik och utbildning (bland andra Holmstrup, et. al. 2015; Jaspers, et. al., 2004). Genom att låta deltagare "tänka högt" kan kunskap erhållas om hur de tänker exempelvis kring problemlösning. Examinatorerna intervjuades enskilt. Varje intervju varade drygt en timme. Samtliga intervjuer spelades in och transkriberades.

Examinatorer

I studien deltar fyra examinatorer. Av forskningsetiska skäl ges ingen beskrivning av examinatorerna. Innan intervjuerna genomfördes var tanken att kategorisera examinatorerna efter tjänstear och ungefärlig ålder. I syfte att skydda examinatorernas identitet beslöt jag dock att utesluta all information om dem, förutom erfarenhet av bedömning av uppsatser, eftersom det har relevans för studiens resultat. Examinator 1, 2 och 3 har minst tio års erfarenhet av bedömning, medan examinator 4 hittills enbart har bedömt enstaka uppsatser, men kommer att examinera i fortsättningen. Examinator 2 har tidigare tjänstgjort vid ett annat lärosäte. Tre av examinatorerna har bidragit med ett examensarbete som de har examinerat och betygsatt vid ett tidigare tillfälle. Eftersom en av examinatorerna ännu inte har examinerat vid institutionen har jag istället bidragit med en uppsats. Ingen information om hur uppsatserna har betygsatts har getts till examinatorerna i förväg. De kände alltså enbart till den uppsats som de själva tidigare har examinerat.

Bedömningsunderlag

Det empiriska bedömningsunderlaget utgörs av fem examensarbeten (uppsatser)³ på grundlärarprogrammet. Dessa benämns här med bokstäver i alfabetisk ordning (A-D). Samtliga uppsatser har avidentifierats vad gäller skribent, handledare och examinator. En av uppsatserna har examinerats vid ett annat lärosäte. Nedan följer en förteckning över hur, samt av vem, uppsatserna har betygsatts:

Uppsats A: Uppsatsen har examinerats av examinator 2 och betygsatts med VG.

Uppsats B: Uppsatsen har examinerats av examinator 3 och betygsatts med G efter kompletteringen som gjordes inom en vecka efter examinationen. Vid examinationstillfället blev uppsatsen underkänd.

Uppsats C: Uppsatsen har examinerats av examinator 1 och betygsatts med U.

Uppsats D: Uppsatsen har examinerats av mig och betygsatts med G.

Uppsats E: Uppsatsen har examinerats vid två tillfällen och har underkänts vid båda. Jag var examinator vid omtentamen. Motiveringen för underkännande har varit att den empiriska undersökningen är bristfällig samt att resonemanget i uppsatsen till stor del bygger på en sammanställning av olika källor och är därför inte tillräckligt självständigt.

Studiens tillförlitlighet

Samtliga examinatorer påpekade att de inte hann läsa alla uppsatser lika noga inför den här studien som de hade gjort om det handlade om ett examinationstillfälle, vilket har betydelse för studiens tillförlitlighet. Examinatorerna gjorde anteckningar i uppsatserna, men framhöll ändå under intervjun att de inte hade alla uppsatser lika färska i minnet.

Ytterligare en aspekt som beaktas i förhållande till studiens tillförlitlighet är att en intervju, där forskare och respondent möts ansikte mot ansikte, är en intim, mellanmännisklig situation. En viktig uppgift för forskaren är att bidra till ett avspänt klimat så att respondenten inte känner sig obekvämt i situationen. Jag upplevde att det faktum att examinatorerna inte i förväg kände till hur uppsatserna har bedömts kunde leda till viss osäkerhet hos vissa av dem och att

³ Båda benämningarna används i detta arbete och avser samma sak.

de i viss mån kunde känna oro över att inte kunna säga det ”rätta” svaret. Under intervjun betonade jag att det enbart är examinatorernas bedömning som är av betydelse för studien samt att de inte behöver känna någon press eller oro sig över att deras bedömning på något sätt kunde vara ”fel”. Exempelvis var examinator 2 inledningsvis försiktig i sin bedömning, eftersom hen utgick ifrån att samtliga uppsatser var godkända. Efter att ha fått veta att så inte var fallet ändrade hen sin bedömning. Att det fanns sådana tendenser under intervjun är väsentligt att påpeka. Det är skillnad mellan att examinera själv och att tala högt om sin bedömning i en intervju.

Resultat

I det följande redogörs för studiens resultat utifrån examinatorernas utsagor. Det empiriska materialet utgörs av sammanlagt drygt fyra timmars inspelade intervjuer, vilket ger underlag för en omfattande analys. Inom ramen för detta arbete kommer enbart vissa aspekter att fokuseras. Resultatet delas in i två avsnitt. Inledningsvis redogörs för examinatorernas tillvägagångssätt vid bedömning. Därefter presenteras examinatorernas betygsättning av de uppsatser som har ingått i studien. Med utgångspunkt i den beskrivningen redogörs för studiens forskningsfråga, nämligen vilka likheter och skillnader som kan synliggöras i examinatorernas bedömning och betygsättning av examensarbeten.

Del 1

Examinators tillvägagångssätt vid bedömning och betygsättning av examensarbeten

En fråga som behandlades under intervjun gällde hur examinatorerna har lärt sig att bedöma och betygsätta. Eftersom samtliga har lärarbakgrund har de erfarenhet av betygsättning från den tid de var verksamma som lärare i skolan. På universitetet finns det möjlighet att delta i kortare utbildningar, vilket examinator 1 och 2 hade erfarenhet av. I första hand uppgav examinatorerna att bedömning var något som man lär sig i praktiken. Alla talade om att de hade någon de kunde vända sig till om de var osäkra eller om de blev ifrågasatta av någon student. Ifrågasättande av bedömning hade, för övrigt, inträffat enbart vid enstaka tillfällen.

Samtliga examinatorer går tillväga på liknande sätt i bedömningsprocessen. Vid en första övergripande läsning skapar de sig en uppfattning om uppsatsens innehåll i relation till betygsriterier:

Ja, jag brukar börja med... dels så kollar man ju alltid vad det är för kriterier som gäller för ett arbete och sen så brukar jag titta på hur den yttre strukturen ser ut, om det hänger ihop, rubriknivåer, så jag bläddrar fram och tillbaka, så att jag får en uppfattning om ungefär hur det ser ut. Och sen så börjar jag väl att läsa helt enkelt och notera efter hand som jag läser och tittar mycket på övergången från yttre struktur till inre struktur och försöker att se hur det hänger ihop på olika sätt. Och sen så om man har flera stycken som man läser så gör man väl på samma sätt med alla sen går man tillbaka lite grand och kollar (Examinator 2).

Svårigheter kring bedömning

Vad gäller svårigheter kring bedömning ger examinatorerna skiftande svar. Examinator 1 uppger att bedömning är svårt, eftersom det alltid råder osäkerhet kring om bedömningen är rätt. Hen menar att betygsättning är orättvist samt att det aldrig går att veta vad examinatorn påverkas av under bedömningsprocessen. Hen betonar den mänskliga aspekten, att människan är föränderlig och irrationell:

Ja, jag tycker att det är jättesvårt alltihopa. Ja, tycker det är skitsvårt. Jag har hela tiden dubier om jag gör rätt, om jag är för hård, om jag bedömer för hårt, om jag missar viktiga saker. Hela tiden tänker jag på sådant. (...)man kommer aldrig att kunna bedöma rätt, för jag förändras ju varje gång jag läser någonting. Det är oerhört orättvist. Det kan vara orättvist mot den första jag läser, det kan vara orättvist mot den sista jag läser och jag kan inte vara riktigt säker på när det är jag förändras. Därför att det blir ju som en process som bara sker under tiden jag håller på. Kanske till och med min dagsform, något jag läste igår, så händer det saker hela tiden. Det är det som gör det så svårt att forska på människor, för vi är irrationella, vi förändras hela tiden, vad vi än är med om. Det är samma sak med oss lärare, vi förändras hela tiden under tiden vi bedömer. Det är förfärligt att vi gör det (Examinator 1).

Enligt examinator 4 kan det vara svårare att göra en bedömning om uppsatsen handlar om ett ämne som hen inte var insatt i:

Ja, men det är just det där om jag inte känner mig helt trygg i ämnet eller inte alls trygg, utan om jag bara kan läsa den som uppsats som sådan (Examinator 4).

Del 2

Betygsättning av examensarbeten

Ingen av uppsatserna i studien har betygsatts med samma betyg av samtliga examinatorer. Störst samstämmighet råder vid bedömningen av uppsatserna A, D och E, där tre av fyra examinatorer har satt samma betyg. *Uppsats A* har bedömts med betyget VG, vilket även stämmer överens med bedömningen vid examinationstillfället. Här var det examinator 1 som

hade en avvikande bedömning. Tre examinatorer har bedömt *uppsats D* med samma betyg som vid examinationstillfället medan examinator 3 hade satt ett högre betyg. Även *uppsats E* har underkänts, vilket var fallet även vid examinationstillfället. Här gör examinator 3 en bedömning som avviker från de övriga. Den största spridningen i betygsättning gäller *uppsats B*, som har bedömts med samtliga betyg på betygsskalan (U-VG). *Uppsats C* har underkänts av två examinatorer (examinator 1 och 2) och godkänts av två (examinator 3 och 4). Uppsatsen underkändes vid examinationstillfället.

	Uppsats A	Uppsats B	Uppsats C	Uppsats D	Uppsats E
Examinator 1	G om komplettering görs	U	U	G	U (möjlighet till komplettering)
Examinator 2	VG	U	U	G	U
Examinator 3	VG	G	G	VG	G
Examinator 4	VG	VG	G om bearbetning görs. Examinatorn tvekar.	G Examinatorn tvekar.	U

När det gäller betygsättningen råder störst samstämmighet mellan examinator 1 och 2. Deras bedömning avviker enbart av uppsats A. Dessa examinatorer har även flest underkända betyg jämfört med de övriga. Examinator 3 och 4 har betygsatt två uppsatser med VG. Examinator 3 har inte underkänt någon av uppsatserna.

Likheter i examinatorernas bedömning

Tydligast likheter i examinatorers resonemang kring betygsättning framkommer vid betyget U. Här ges exempel från två examinatorers betygsättning av uppsats E. Båda examinatorerna pekar på att uppsatsen har alltför stora brister för att bli godkänd. Examinator 4 tar fasta på bristande redovisning av metod och teori samt studentens språkbehandling. Examinatorn gör bedömningen att uppsatsen har sådana brister att den inte kan färdigställas inom den tid som ges till komplettering:

- Ja, den har jag satt U på då, för jag tycker att det är brister både i metod och teori och förståelsen av innehållet som studenten har jobbat med. Sen är det strukturella frågor.

Återigen är det det här med syfte och frågeställningar. Jag vet inte, ibland är det en lärare ibland flera, så det är lite rörigt och jag undrar varför är inte det här korrigerat? Språkbehandlingen är väldigt dålig Den här skulle jag underkänna för språket, för att det inte är ”i stort sett korrekt”. Nej, det är inte i stort sett korrekt, utan det är för många fel här. (...) Här hade det behövts fler pekningar med hela handen. Formalia funkar inte heller. Referenser och så i alla lägen. Så. Sen var det också rörigt. Hela det här, tidigare forskning. Läsförståelsestrategier. (...) Det är lite rörigt. Jag tycker inte att studenten har förstått hur saker och ting hör ihop.

- Bearbetning en vecka eller?

Nej. Det är inte bearbetning, utan det är mer omarbetning. Själva resultatet, det är ju... visst är det lite tunt. Den här studenten verkar tro att i resultat presenterar du hårda fakta och sen så gör du en analys. Det är ofullständigt tunt (Examinator 4).

Handledaren ansågs ha en viktig roll av samtliga examinators och uppgiften om handledaren kunde i viss mån även påverka examinatorns förväntningar på uppsatsen. I likhet med examinator 2 är examinator 1 kritisk mot att vissa handledare inte hjälper studenterna tillräckligt med att upptäcka grundläggande brister i uppsatserna:

Varför har inte handledaren sagt det? (...). Nej, den var dålig. (...) Det här tycker jag är ett handledarfel. Det är inte ett studentfel. Det är ett handledarfel (Examinator 1).

När det gäller betyget VG handlar likheterna i betygsättning om att uppsatsens delar hänger ihop samt att det är synligt hur teorin har tillämpats. Även språkliga kvaliteter lyfter uppsatsen:

Jag tyckte att det här var en väldigt bra uppsats. Det var en av de här fem som jag tyckte var tydligast vad gäller teori och metod, en fenomenografisk klassisk ansats, som jag tyckte att de hade gjort ganska bra. Kanske lite mager empiri, men jag tycker att de har skrivit ihop det ganska bra. Men jag tyckte att den var väldigt rak och bra, inget tjafs. Ganska gedigen när det gäller allting. Bakgrunden var tydlig och snygg struktur. Det är en av de fem som jag tyckte var tydligaste när det gäller metod och teori och skriven på ett väldigt bra språk som jag tyckte flöt, så jag skulle sätta ett VG på den här (Examinator 3).

I de fall då examinatorerna har varit eniga kring betyget G visar deras resonemang kring betygsättningen att de kan vara overseende med vissa tillkortakommanden om uppsatsen i stort uppfyller de grundläggande kraven i betygsriterierna. Det som examinatorerna hade störst overseende med var den teoretiska förankringen, vilket samtliga påpekade var en svår uppgift för studenterna att hantera. Examinator 3 anser till och med att det kravet skulle kunna strykas:

Jag har satt ett litet frågetecken kring teori. Där försöker de göra något av sociokulturell teori som blir lite halvhjärtat... och det är nog det här kravet att vi ska ha teori. Man skulle kunna stryka det också (Examinator 3).

Examinator 1 ställde sig undrande till hela idén kring examensarbeten och kravet på att lärarstudenterna ska kunna skriva ett vetenskapligt arbete för att få ut sin lärarexamen. Hen ifrågasätter alltså varför blivande lärare ska sätta sig in i en massa teori och vetenskaplig metod samt vad de ska ha det till. Examinatorn påpekar att studenterna under sin utbildning inte får möjlighet att lära sig det som krävs för att de ska kunna skriva ett vetenskapligt arbete. Av den anledningen vill denna examinator helst slippa examinera:

- Återigen, varför ska hon göra något sådant här? Hon kan väl göra något annat än att visa att hon har förstått en massa teorier, det är ju det viktigaste. (...) Alltså det är så här att jag värjer mig så mycket jag kan från att vara examinator. Jag säger att jag inte vill, samtidigt så säger de att jag måste. Jag vill inte vara examinator. Jag tycker att det är förfärligt. Därför att det här... jag tycker inte att det här är rätt. Jag tycker att det här är fel och så måste jag sitta där... Ska jag vara examinator så vill jag bara ha sådana som blir godkända minst.
- Du tycker att det är fel att de ska skriva sådana här arbeten?
- Nej, inte egentligen, men då måste de också göras på rätt sätt. Jag tycker inte att de får grunden för att kunna göra ett sådant här arbete (Examinator 1).

Examinatorerna visar alltså på en medvetenhet kring uppgiftens svårighetsgrad och förefaller anpassa sina krav efter det som kan vara rimligt att begära av studenterna utifrån de förutsättningar som ges.

Skillnader i examinatorernas bedömning

Den största spridningen i betygsättning gäller uppsats B, som har underkänts av två examinatorer (1 och 2), godkänts av examinator 3 och fått det högsta betyget av examinator 4. Examinator 2 utgick ifrån att samtliga uppsatser var godkända, vilket förefaller inledningsvis påverka hans bedömning:

- Eh... jag satte G med ett litet frågetecken. Och det som jag framförallt satte frågetecken kring, det var metod och teori. Det var väl i den här som jag inte uppfattade om det fanns någon teori överhuvudtaget.
- Nej, just det. Det är metod. Det finns bara metod.
- Och där blev jag lite konfunderad faktiskt, för det är väldigt mycket metod och ingen teori. Då blev det lite så här att jag tyckte genomgående på de uppsatserna som var härifrån GU, så tyckte jag att i tre av fyra, så tyckte jag att det var jättestora brister i teori och analys, så att jag hade nog inte godkänt någon egentligen och då var väl den här lite bättre. Och då vet jag inte riktigt hur man har lagt ribban här, alltså. Jag blev förvånad att de har gått igenom. Jag antar att alla de här är godkända (examinator 2).

Examinator 1 är, till skillnad från examinator 2, inte tveksam i sin bedömning, utan anser till och med att uppsatsen inte skulle ha examinerats i nuvarande skick:

- Ja, jag sammanfattar. Den får U. Upplever inte att studenten har fått handledning. Borde ha stoppats från att gå upp. Jag tycker att syftet är okej, metod, teori – knappt, resultatet, analysen – nej. Ingen egentlig analys görs. Det är bara en ren beskrivning. Källorna är bra. Diskussionen – nej. Struktur – enkel, formalia – svag. Referenser – dåliga, refererandet är dåligt. Språk – värderingar, enkelt, tillkrånglat. Sammanhang – knappt. Skriver rätt dåligt, får en känsla av att det är något mer än vad jag förstår som vill vara sagt. Många fel i referenser, både i text och i referenslistan. Ganska svag metodlitteratur.
- Skulle den här få komplettera?
- Nej.
- Den skulle behöva göras om?
- Ja, hen skulle behöva göra om (Examinator 1).

Examinator 3 var den som examinerade uppsats C. Den blev underkänd vid examinationstillfället, men godkänd efter komplettering:

- Det var den jag var examinator på. Den fick U av mig med ett antal krav på kompletteringar. Syfte och frågor måste justeras, metoden måste tydligare kopplas till vad som gjorts. Resultat måste förtydliga. Ja, jag hade rätt mycket krav. Och sen så jobbade hon en vecka på detta och kom tillbaka med en förbättrad version, som jag tyckte med ett nödrop kom över gränsen till G. (...)
- Men här är fortfarande en hel del saker som...
- Som inte är perfekta. Jajamän! Så kan det nog vara.
- Men den blev godkänd i alla fall.
- Den blev godkänd. Ja (Examinator 3).

Examinatorn påpekar att studenten har gjort en avsevärd förbättring i relation till den första versionen, vilket i sin tur kom att påverka betygsättningen. Trots att uppsatsen fortfarande uppvisade en mängd brister enligt examinatorn valde hen att godkänna den. Studentens arbetsinsats inkluderas i den slutgiltiga bedömningen:

Det var ganska stora brister och jag skrev att den var tveksam och sen gjorde hon en avsevärd förbättring. (...) Hon jobbade som tusan med det och möjligtvis tog det lite längre tid än en vecka också. Eh... men det blev i alla fall i förhållande till den första versionen en väldigt stor förbättring. Så veknade examinatorn (Examinator 3).

Av det empiriska underlaget framgår att examinatorerna i grunden har en välvillig inställning till studenterna och att det är besvärligt att behöva underkänna examensarbeten. Det är emellertid viktigt att inte ge vika om ett godkänt betyg inte kan motiveras:

Grundprincipen är att hellre fria än fälla. Det tycker jag. Men det får inte heller vara så att nivån sjunker (examinator 2).

Sammanfattningsvis kan konstateras att examinatorerna gör olika överväganden vid betygsättning. Det går emellertid inte finna någon systematik i hur dessa överväganden görs

mer än på ett övergripande plan. Av studiens resultat framgår inte heller hur stor betydelse eventuella tillkortakommanden ges vid den slutgiltiga betygsättningen.

Diskussion

Ingen av de fem uppsatserna som bedömdes i studien har fått samma betyg av samtliga examinatorer. Frågan är om spridningen i betygsättningen hade minskat om det empiriska underlaget hade utgjorts av fler examinatorer eller om ett större underlag hade bidragit till mindre spridning. I denna studie har faktorer, såsom bakgrund, ämnesinriktning utelämnats medan examinatorers erfarenhet av bedömning kan vara av viss betydelse och kommer följaktligen att diskuteras här. Om fler faktorer hade inkluderats i ett större underlag hade möjligtvis vissa mönster kunnat urskiljas tydligare och mer generella slutsatser kunnat göras. Måhända hade examinatorer kunnat betraktas som representanter för en viss enhet eller ämnesinriktning. I denna studie representerar varje examinator sig själv och de slutsatser som görs baseras således på den enskilda examinatorns utsagor.

Diskrepansen i examinatorernas betygsättning kan dels förstås i relation till hur de tillägnar sig denna färdighet dels till den erfarenhet de har inom området. Vid Göteborgs universitet finns det möjlighet att delta i kortare kurser och work shops där betygsättning av uppsatser diskuteras. Det ligger dock i den enskilde examinatorns intresse att delta i sådana kurser. Ett rimligt antagande är att betygsättande lärare utvecklar kunskap om betygsättning i praktiken och att de med tiden även en känsla för vad som är en god eller en acceptabel studentprestation. Denna kunskap utvecklas emellertid ofta i samspel med närmaste kollegor, vilket i sin tur kan innebära att kunskapen om hur bedömningar görs vid andra institutioner eller lärosäten blir begränsad. Att examinator 2 inledningsvis var försiktig i sin bedömning är ett exempel på detta. Hen utgick ifrån att samtliga uppsatser var godkända och fick därmed uppfattningen att kraven är lägre vid Göteborgs universitet än vid det lärosäte hen tidigare tjänstgjort.

Olikheter i bedömning som synliggörs i betygsättningen av uppsats B kan ha sin förklaring i att examinator 4 ännu inte har examinerat i någon större utsträckning och att hen har kortare erfarenhet än övriga examinatorer som har deltagit i studien. Jämförelse ligger i betygsättningens natur. Ju längre erfarenhet en examinator har desto större underlag har hen för jämförelser. Både bedömning och betygsättning bygger dessutom till stor del på tyst

kunskap och särskilt betygsättningen är därför inte alltid enkel att explicitgöra. Ett uttalande som ”Jag känner igen ett VG när jag ser det” (jfr. Eldelin, 2010, s. 43) är inte ovanligt i betygsättnings-sammanhang. Rådande ”bedömningskultur” vid en viss institution kan därför vara avgörande för tillägnande av erfarenhet inom detta område.

De likheter i examinatorernas betygsättning som har kunnat synliggöras i denna studie handlar om att samtliga tillämpar holistisk bedömning när slutbetyget ska sättas, det vill säga att de kan bortse från vissa tillkortakommanden om uppsatsen i stort motsvarar förväntade kvaliteter (jfr. Sadler, 2009; 2005). Däremot tillämpar examinatorerna en analytisk bedömning under bedömningsprocessen när de går igenom i vilken utsträckning de olika delarna i uppsatsen motsvarar kraven i betygskriterierna. Det första intrycket kan emellertid vara avgörande även om det är svårt att veta i vilken utsträckning det påverkar det slutgiltiga betyget (jfr. examinator 3). Dessutom kan examinatorerna, vid holistisk bedömning, ha något lägre krav på de kriterier som de upplever att studenterna inte har någon möjlighet att uppnå. Det kan exempelvis handla om uppsatsens teoretiska förankring eller att resultatet inte är särskilt omfattande.

Studien visar alltså på såväl likheter som skillnader i examinatorernas resonemang. Vilka kvaliteter eller tillkortakommanden som examinatorerna tar i beaktning eller bortser ifrån varierar således. Samtliga betonade emellertid språkets betydelse, men kunde samtidigt bortse ifrån vissa språkliga brister i slutversionen.

Samtliga examinatorer i studien uppger att de utgår från betygskriterierna i sin betygsättning. Examinator 1 och 3 pekar på svårigheter både vad gäller tolkning och tillämpning av kriterierna och ifrågasätter dessutom hela idén med att lärarstudenter ska skriva ett vetenskapligt arbete. Detta förhållningssätt kan relateras till pågående utvecklingsarbete vid landets lärosäten (Bergqvist, 2015; Bjuremark, 2010) där inte minst sammanlänkningen av kursmål och examination, inklusive betygsättning diskuteras⁴. Betygskriteriernas betydelse och tillämpning måste enligt Högskoleverket (2008) vara tydlig om examinationen ska betraktas som rättssäker. Kravet på att kriterierna ska kunna göras begripliga för studenterna för med sig diverse svårigheter, inte minst vad gäller gradering (Hillström, 2010). Att examinatorerna i studien i vissa fall bortser från vissa tillkortakommanden kan alltså förstås som att de bedömer att dessa krav inte är realistiskt ställda. En fråga som uppkommer i

⁴ Jfr *Constructive – alignment* (Biggs & Tang, 2007).

sammanhanget är om det ska vara upp till varje lärare att avgöra vad som är rimligt att begära eller om kravet att lärarstudenterna ska kunna genomföra ett vetenskapligt arbete möjligtvis är för högt ställda. Att ingen av uppsatserna har bedömts på samma sätt av samtliga examinatorer i denna studie visar att de gör olika överväganden vid bedömning, även om samtliga examinatorer är verksamma vid samma institution.

Vid bedömning och betygsättning kan den mänskliga faktorn inte bortses, vilket påpekas av examinatorerna i studien. Total likvärdighet och rättssäkerhet är därför omöjlig att uppnå. Detta synsätt finner stöd i de rundabordssamtal som hölls i samband med ett samarbetsprojekt mellan universiteten i Linköping och Umeå (Marteinsdottir, 2010). Av samtalen framkom bland annat att bedömning är ”mycket upp till enskilda lärare” (s. 99) samt att lärarna upplevde att det är svårt att göra rättvisa bedömningar. En fråga som diskuterades i detta sammanhang var hur kvaliteten på examinationen kunde säkerställas på bästa sätt, eftersom examensarbeten kan utgöra ett bedömningsunderlag för granskningen av ett lärosäte och dessutom fungera som en indikator på utbildningens kvalitet. Detta kan i sin tur få konsekvenser för resursfördelningen (Marteinsdottir, 2010)⁵. Med utgångspunkt i studiens resultat är frågan om vilken roll ett examensarbete ska ha, samt om det överhuvudtaget kan betraktas som en kvalitetssäkring på utbildningen, angelägen att beakta i den fortsatta diskussionen.

Sammanfattning och vidare forskning

Examinatorerna i studien förefaller utgå ifrån samma grundprinciper vad gäller godtagbar kvalitet, men i den slutgiltiga helhetsbedömningen uppstår det ändå skillnader. En slutsats som kan dras av studiens resultat är att det knappast går att uppnå likvärdighet i betygsättning, vilket examinatorerna har visat sig vara medvetna om.

Som nämnts inledningsvis är forskningen om hur lärare resonerar när de bedömer och betygsätter studentarbeten i allmänhet och examensarbeten i synnerhet ännu inte särskilt omfattande. Med tanke på att betygsättning ofta diskuteras utifrån en rättssäkerhetsaspekt är det ett angeläget forskningsområde, inte minst med tanke på att olika betygsskalor tillämpas

⁵ Marteinsdottir (2010) hänvisar till den utvärdering av högskoleutbildningar som genomfördes av Högskoleverket på regeringsuppdrag år 2009. Istället för den tidigare fokuseringen på förutsättningar och processer skulle kvaliteten värderas utifrån utbildningsresultaten. Lärosäten med hög utbildningskvalitet skulle premieras vid resursfördelning och examenstillståndet skulle kunna bli ifrågasatt eller återkallat om en utbildning bedömdes ha allvarliga brister (s. 97).

vid landets lärosäten. Det skulle därför vara av intresse att i ett komparativt syfte följa upp studien både vid andra institutioner samt vid andra lärosäten.

Referenser

Bergqvist, J., & Lunds universitet. (2015). *Att sätta praxis på pränt: En handbok i att skriva betygskriterier*. Lund: Centre for Educational Development (CED), Lunds universitet.

Biggs, J & Tang, C (2007). *Teaching for Quality Learning at the University*. New York, NY: Maidenhead.

Bjuremark, A. (red.) (2010). *Att bedöma och sätta betyg: en utvecklingskonferens om traditioner. krav och nytänkande vid LiU, 11 mars 2010*, Centrum för undervisning och lärande, Linköpings universitet: Linköping.

Bloxham, S., Campbell, L. (2010). Generating dialogue in assessment feedback: exploring the use of interactive cover sheets. *Assessment & Education in Higher Education*, 35(3), 291-300.

Charters, E. (2003). The use of Think-aloud Methods in Qualitative Research. An introduction to Think-aloud Methods. *Brock Education*, 12(2), 68-82.

Eldelin, E. (2010). Att skilja G från VG i kvalitativa bedömningar – inledningsanförande. I Bjuremark, A. (red.). *Att bedöma och sätta betyg: en utvecklingskonferens om traditioner. krav och nytänkande vid LiU, 11 mars 2010*, Centrum för undervisning och lärande, Linköpings universitet: Linköping, (43-48).

Hillström, M. (2010). Att bedöma och sätta betyg utifrån betygskriterier. I Bjuremark, A. (red.). *Att bedöma och sätta betyg: en utvecklingskonferens om traditioner. krav och nytänkande vid LiU, 11 mars 2010*, Centrum för undervisning och lärande, Linköpings universitet: Linköping, (35-43).

Holmstrup, M., E., Verba, S., D., Lynn, J., S. (2015). Developing best practices teaching procedures for skinfold assessment: observational examination using Think Aloud method. *Advances in Physiology Education* 39:4, 283–287. DOI: 10.1152/advan.00044.2015

Hyltegren, G. (2014). *Vaghet och vanmakt: 20 år med kunskapskrav i den svenska skolan*. Göteborg: Acta universitatis Gothoburgensis.

Högskoleverket. (2008). *Rättssäker examination. Andra omarbetade upplagan*. Rapport 2008:36 R. Stockholm: Högskoleverket.

Jaspers, M., Steen, T., van den Bos, C., Greenen, M. (2004). The think aloud method: a guide to user interface design. *International Journal of Medical Informatics* 73, 781—795.

Marteinsdottir, I. (2010). Bedömning av uppsatser och examensarbeten. I Bjuremark, A. (red.). *Att bedöma och sätta betyg: en utvecklingskonferens om traditioner, krav och nytänkande vid LiU, 11 mars 2010*, Centrum för undervisning och lärande, Linköpings universitet: Linköping, (97-103).

Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment and Evaluation in Higher Education*, 34:2, 159-179, DOI: 10.1080/02602930801956059.

Sadler, D. R., (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment & Evaluation in Higher Education*, 30(2), 175-194.

Selghed, B. (2010). Ett omöjligt uppdrag? – om lärares bedömningar och betygsättning. I Brante, G., och Hjort, K. (red.). *Dilemman i skolan. Aktuella utmaningar och professionella omställningar*. 82–95. Kristianstad: Högskolan Kristianstad.

Bilaga 1

Betygskriterier till det självständiga arbete på avancerad nivå som skrivs inom L3XA1A och L6XA1A

Efter avslutad kurs förväntas studenten, enligt kursplanen, kunna:

1. identifiera och vetenskapligt beskriva ett problemområde av relevans för yrkesutövningen,
2. visa kännedom om den vetenskapliga processens delar och helhet, från problemformulering till slutsats,
3. visa fördjupad förmåga att kritiskt och självständigt tillvarata relevanta forskningsresultat för att därigenom kunna bidra till kunskapsutvecklingen inom yrkesområdet,
4. visa förmåga att självständigt författa en vetenskapligt sammanhållen text på formell svenska eller engelska och muntligt argumentera för brister och förtjänster i ett vetenskapligt arbete på ett konstruktivt sätt,
5. visa förmåga att muntligt redogöra för och diskutera sina slutsatser,
6. visa förmåga att identifiera behov av ytterligare kunskap inom det valda problemområdet,
7. visa förmåga att identifiera och diskutera forskningsetiska aspekter inom det valda problemområdet

När en uppsats betygsätts gör examinatorn en helhetsbedömning mot bakgrund av de mål som är formulerade i kursplanen. I nedanstående mall har vi delat upp denna helhetsbedömning i olika områden. Kriterierna bör tolkas kompensatoriskt. Ett högre betyg avseende ett kriterium kan alltså uppväga ett lägre betyg vad gäller ett annat. Avsikten med mallen är att ge ett stöd för studenter, handledare och examinatorer. Punkt 4 ovan bedöms till viss del vid ett avslutande examinations-seminarium. Punkt 5 ovan bedöms till viss del vid den populärvetenskapliga presentationen.

Bedömningsområde	För Godkänt krävs att	För Väl godkänt krävs dessutom att
Syfte och problemformulering	Syftet är precist och väl avgränsat och relevant för	Syfte och frågeställningar är självständigt valda och väl

	<p>lärarprofessionen.</p> <p>Uppsatsens frågeställningar har viss förankring i tidigare forskning.</p>	<p>motiverade utifrån tidigare forskning.</p>
Metod och teori	<p>Redogörelsen för den metod och de teorier som används är tydlig och motiverad med för- och nackdelar.</p>	<p>Redogörelsen för den metod och de teorier som används är reflekterad och väl integrerad i uppsatsen.</p>
Resultat och analys	<p>Resultatredovisning, analys och tolkning i enlighet med forskningsfrågorna och de angivna metoderna.</p>	<p>Grundlig och självständig redovisning, analys och tolkning.</p>
Forskningsförankring och källmaterial	<p>Aktuellt forskningsmaterial används. Valet av källor är relevant och motiverat. Källorna täcker forskningsområdet.</p>	<p>Aktuell forskning används och relateras till den egna studien. Valet av källor diskuteras och motiveras väl. Författaren visar i texten en medveten källkritisk hållning.</p>
Diskussion och slutsatser	<p>Diskussionen är relevant för det aktuella området. Forskningsfrågorna är besvarade.</p>	<p>Analys och diskussion är självständig och väl presenterad. Uppsatsen innehåller en fördjupad diskussion.</p>
Struktur	<p>Uppsatsen har en genomtänkt och begriplig struktur. Formalia är korrekt och uppvisar konsekvens.</p>	<p>God integrering mellan problemformulering, metod, teori och analys.</p>
Språk	<p>Språket i uppsatsen är värderingsfritt, formellt och i</p>	<p>Språket är nyanserat, stilistiskt konsekvent och</p>

	stort sett korrekt i enlighet med svenska eller engelska språkregler.	innehåller få fel.
Sammanhang	Det egna arbetet sätts in i ett vidare sammanhang.	Väl valda förslag på relevant vidare forskning diskuteras.

Bilaga 2

Intervjufrågor

1. Berätta om din bakgrund. Hur länge har du arbetat som universitetslärare? Inom vilka kurser undervisar du? Hur ofta examinerar du varje år?
2. Hur har du lärt dig att bedöma och betygsätta? Har du någon formell utbildning inom detta område?
3. Hur arbetar ni med bedömning- och betygsättning på din institution? Förekommer sambedömning? Hur gör du om du är osäker på din bedömning?
4. Hur går du tillväga när du bedömer ett examensarbete?
5. Hur påverkas du av uppgiften om handledaren? Har handledarens forskningsintresse någon betydelse för dig? Har din relation till handledaren, dvs. om du känner hen väl/inte känner hen alls, någon betydelse?
6. Påverkas du av själva examinationsförfarandet, exempelvis hur studenterna reagerar på dina synpunkter under examinationen eller hur de beter sig?
7. Är det något du upplever som svårt med bedömning? Vad i så fall?
8. Har du någon gång blivit ifrågasatt i din bedömning? Av vem?
9. Hur använder du betygskriterier?
10. Hur har du gått tillväga när du har bedömt uppsats A, B, C, D, E?
11. Vilka förtjänster/brister har uppsats A, B, C, D, E?
10. Hur motiverar du din bedömning av A, B, C, D, E? Finns det kvaliteter som du kan bortse ifrån/ absolut inte kan tänka sig att bortse ifrån?

