

Konsumtionsuppsatser som ny bedömningspraktik för lärarutbildare

Mats Lundström (A), Karin Stolpe(B), Lars Björklund(B), Maria Åström (C)

(A) Malmö universitet

(B) Linköpings universitet

(C) Göteborgs universitet

ABSTRACT

Vid den senaste lärarutbildningsreformen beslutades att på de flesta lärarutbildningsprogrammen ska lärarstudenter skriva självständiga arbeten omfattande totalt 30 högskolepoäng. Vid många lärosäten har detta lett till en uppdelning i två olika självständiga arbeten: en konsumtions- och en produktionsuppsats. Denna studie undersöker om konsumtionsuppsatser – en typ av kunskapsöversikter – innebär en ny bedömningspraktik för lärarutbildare som examinatorer. Examinatorer vid lärosäten med lärarutbildning har blivit intervjuade och fått besvara en enkät rörande dessa bedömningskriterier. Resultaten indikerar att en majoritet av examinatorerna inte ser stora skillnader mellan att bedöma en konsumtionsuppsats och en mer traditionell produktionsuppsats där egen empiri i form av till exempel intervjuer samlas in och analyseras. Vid intervjuerna framkom dock några skillnader mellan de två uppsatstyperna. Dessa hänger huvudsakligen ihop med antingen hur strukturen byggs upp, vad som är metod eller hur forskningsförankringen är framskrivet. Slutsatsen blir att den nya form av uppsatser som konsumtionsuppsatser innebär inte tycks ha renderat i någon stor förändring av bedömningspraktiken hos examinatorerna.

INTRODUKTION

I den senaste utredningen kring lärarutbildningen - *En hållbar lärarutbildning* (SOU 2008:109) - betonas vikten av att lärarutbildningarna utbildar för ett vetenskapligt och kritiskt förhållningssätt. Utredningen fastställer att

lärarutbildningen ska vila på en vetenskaplig grund samt vara forskningsanknuten, något som redan tidigare betonats av Högskoleverket (Säljö & Södling, 2006). Som ett led i att förstärka kvaliteten i högre utbildning föreslog en utredning (Prop. 2009/10:139) att utvärderingarna skulle lägga stor vikt vid självständiga arbeten (Hjort & Sundkvist, 2010). Som en konsekvens av detta har lärarstudenternas vetenskapliga texter fått en allt större roll i såväl examination av studenterna som bedömning av lärosätets kvalitet (Svärd, 2013). Dessa skärpta krav på att kunna skriva vetenskaplig text medför att det blir viktigt med kunskap om hur examinators bedömer studenters vetenskapliga texter. Eftersom de självständiga arbetena har en viktig roll i högskoleutbildningar är det väsentligt att pedagogisk forskning belyser området, som en del av högskolepedagogisk forskning. Råde (2016) beskriver hur studier om lärarutbildningens examensarbeten ofta är genomförda ur ett studentperspektiv eller som dokumentstudier. Det har också förekommit studier kring handledning (Gustavsson & Eriksson, 2015) och kvaliteten i examensarbeten (Forsberg & Lundgren, 2006). Även om Högskoleverket har initierat en del utredningar som redovisas nedan menar vi att det är viktigt att mer pedagogisk forskning görs inom området. Det är därför av vikt att belysa handledar- och examinatorperspektivet i forskning om examensarbeten på lärarutbildningen. Genom att undersöka vilka bedömningskriterier som betonas som viktiga i en lärarstudents uppsats ges möjlighet att analysera hur viktiga delar av en pedagogisk utbildning bedöms. Redan när de första examensarbetena i den reformerade lärarutbildningen skrevs år 2004, var de föremål för forskning. Gustafsson och Hallström (2005) diskuterar hur lärarstudenternas examensarbeten ska syfta till att utveckla en mängd förmågor som behövs i en kommande karriär. De menade redan då att både läroplansteoretisk och pedagogisk forskning behövs inom området, inte minst avseende bedömning av vetenskaplig förmåga hos studenterna.

Ett vanligt sätt att organisera det självständiga arbetet i enlighet med den nya examensordningen från 2011 är att låta studenterna skriva två självständiga arbeten om vardera 15 högskolepoäng (hp). Därmed har antalet självständiga arbeten på lärarutbildningsprogrammen ökat. Detta medför i sin tur att allt fler disputerade lärare från olika ämnesdiscipliner måste involveras i handledning och examination av dessa självständiga arbeten på lärarutbildningen, vilket kan leda till problem med reliabilitet och samstämmighet i bedömningen av arbetena (Svärd, 2013). I lärarutbildningen kommer examinatorerna från olika ämnesdiscipliner och bedömningstraditioner, till exempel från ett naturvetenskapligt ämne, naturvetenskapernas didaktik eller humaniora. Dessa examinatorer har dessutom olika lång erfarenhet som handledare och examinatorer. Olika uppfattningar om hur god vetenskap ser ut, testas och redovisas (Schwab,

1964; Scriven, 1964) samt olika erfarenhet skulle kunna ge upphov till olikheter i bedömning av självständiga arbeten.

Den nya lärarutbildningsreformen (Bet. 2009/10:UbU16; SOU 2008:109) innebär att studenternas förmåga att producera texter betonas. En följd av detta är att skrivandet av längre vetenskapliga texter introduceras tidigare i utbildningen. En vanlig lösning i de lärarprogram som skriver självständiga arbeten om sammanlagt 30 hp, är att det första skrivs på grundnivå och det andra på avancerad nivå. En variant som förekommer på flera lärosäten är att arbetet på grundnivå utgör en kunskapsöversikt där studenterna använder artiklar, forskningsrapporter och avhandlingar som data. Denna typ av kunskapsöversikter kommer i denna artikel att benämnas *konsumtionsuppsatser*. Till skillnad från konsumtionsuppsatsen, samlar studenten sedan in uppsatsen på avancerad nivå egen primärdata i form av exempelvis enkäter eller intervjuer. Den påminner om de examensarbeten som har funnits under en längre tid på lärarutbildningarna och kallas i fortsättningen för *produktionsuppsats*. Konsumtionsuppsatser tillhör fortfarande genren vetenskaplig text, men den genomförs och skulle kunna skrivas på ett annorlunda sätt än en produktionsuppsats. Därmed uppkommer frågan om bedömningen av konsumtionsuppsatser kommer att göras utifrån andra aspekter än de väl etablerade produktionsuppsatserna. Med andra ord, medför frånvaron av studentinsamlad empiri i form av till exempel intervjuer att kriterierna för bedömningen av konsumtionsuppsatser skiljer sig från kriterier för bedömning av produktionsuppsatser?

Eftersom konsumtionsuppsatser är en relativt ny företeelse på lärarutbildningsprogrammen är de också obeforskade. Denna artikel diskuterar vilka bedömningskriterier som handledare och examinatorer för dessa konsumtionsuppsatser använder samt hur dessa relaterar till kriterier som traditionellt används vid bedömning av produktionsuppsatser. Data har samlats in vid totalt sex lärosäten, där både enskilda intervjuer och en webbaserad enkätstudie genomförts med examinatorer av självständiga arbeten. Samtliga i intervjustudien och majoriteten av informanter i enkätstudien kommer från tre lärosäten med stor volym av lärarutbildning.

SYFTE OCH FORSKNINGSPRÅG

Syftet med studien är att undersöka om konsumtionsuppsatser - som en relativt ny företeelse på lärarprogrammen jämfört med produktionsuppsatser - innebär en ny bedömningspraktik för lärarutbildare som examinatorer. Följande forskningsfråga ramar in studien:

På vilka sätt anser examinatorer att konsumtionsuppsatser skiljer sig från produktionsuppsatser med avseende på bedömningskriterier?

TIDIGARE FORSKNING

I nedanstående avsnitt kommer inledningsvis studier av självständiga arbeten i en svensk kontext att presenteras. Därpå följer internationell forskning om lärares bedömningspraktiker inom högre utbildning. I begreppet bedömningspraktik lägger vi i denna artikel användningen av bedömningskriterier.

Självständiga arbeten på svenska lärosäten

Utredningar och forskning om självständiga arbeten på svenska lärarprogram har till stor del fokuserat på kvaliteten på uppsatserna. I en utredning på uppdrag från Högskoleverket (HSV) 2006, granskade Forsberg och Lundgren 116 examensarbeten i lärarutbildningen samt jämförde dessa med examensarbeten från sjuksköterskeprogrammet samt ämnena företagsekonomi och psykologi (Forsberg & Lundgren, 2006). Författarna till utredningen visade att det finns indikationer på att det råder en viss oenighet om såväl bedömningsgrunder som betygskriterier både inom bedömargruppen i studien och jämfört med examinatoreer på respektive utbildning. I denna utredning gjordes kvalitetsbedömningen utifrån sju aspekter; forskningsanknytning, problemformulering, teoretisk medvetenhet, metod, teori-empiri, språklig utformning och formalia samt uppsatsämnets svårighetsgrad (Forsberg & Lundgren, 2006). Korrelationen mellan de två bedömare som granskade samma uppsatser var i genomsnitt 0.45 när en helhetsbedömning gjordes på en femgradig skala. Bedömningarna hos de fyra bedömarparen varierade mellan 0.30-0.79 i korrelation. Det var störst samvariation mellan helhetsbedömning och bedömning av teori-empiri samt teorimedvetenhet, vilket antyder att bedömarna lade stor vikt vid hur teori behandlas i examensarbetena. Kriteriernas svårighetsgrad samt språk och formalia ansågs mindre viktiga av bedömarna för en helhetsbedömning. Den språkliga utformningen, formalia och disposition utgjorde sällan något större problem, tvärtom ansågs en majoritet av uppsatserna vara tillfredsställande i detta avseende. Detsamma gäller val av problem som ofta knyter an till den kommande yrkesverksamheten och/eller något aktuellt ämne. Problem som lyftes fram och som kan identifieras även i den typiska uppsatsen handlar till exempel om normativitet, uppsatsens teoretiska förankring och studenters kritiska förmåga (Forsberg & Lundgren, 2006). Resultaten indikerar att olika bedömare inte bedömer på samma sätt. Detta kan grunda sig i oenighet i vad som anses som viktigt eller att bedömare har olika syn på vad som kan betecknas som hög kvalitet. Med samma empiri visar dessutom Forsberg (2012) att det inte är någon större skillnad mellan vilken kvalitet olika bedömningsdelar eller kategorier (exempelvis problemformulering, teorimedvetenhet och metod) av ett examensarbete har. Bedömarna i studien rankar kvaliteten av dessa att vara på ungefär samma nivå (2.6-3.1 på en

femgradig skala). En liknande studie på uppdrag från HSV, där lärarutbildningen stod i fokus (Säljö & Södling, 2006), indikerar att kriterier som anknyter till kritiskt tänkande, analytisk förmåga, källkritik, ifrågasättande av slutsatser och påståenden, nyfikenhet och överblick anses som mest viktiga. Även Säljö och Södlings studie visar att bedömningspraktiken inte är enhetlig och konsistent.

Lärares bedömningspraktik och kriteriers användning vid bedömning

Svennberg, Meckbach och Redelius (2014) och Young (2011) anser att lärares bedömningspraktik i allmänhet uppvisar brist på konsistens och likformighet. Svennberg et al. (2014) beskriver hur lärare i sina bedömningar tenderar att använda sig av en blandning av studenters attityder, ansträngning och prestation. I liknande termer uttrycker Young (2011) lärares betygssättning som lärares syn på kunskaper och färdigheter blandat med uppfattningar om studenters närvaro, attityd och motivation. Även Sadler (2005) anser att examinatorers dominerande utgångspunkt när de ska bedöma studentarbeten är en hybrid av två faktorer: lärarens personliga förväntningar och hur andra studenter har presterat. För att motverka olikheter och öka reliabiliteten i bedömningar arbetar allt fler lärosäten med bedömningskriterier som fastställs i förväg (Bloxham, Boyd & Orr, 2011). Syftet är att dessa kriterier ska vara kända och accepterade hos studenter, handledare och examinatorer. Därmed ska det bli mer explicit på vilken grund betygssättningen sker. Bloxham m.fl. (2011) anser att trots goda intentioner har systemet mött kritik, vilket leder till slutsatsen att förutbestämda kriterier medför en positivistisk syn på bedömning. De menar att en uppgift eller uppsats kan besvaras eller skrivas på väldigt olika sätt men där olika sätt kan vara lika effektiva och bra. Studier har visat att examinatorer många gånger inte heller använder sig av ett analytiskt, kriteriebaserat sätt att göra sin bedömning, utan istället använder sig av ett holistiskt perspektiv. Bloxham m.fl. (2011) fann att trots att det fanns officiella kriterier att bedöma studenters uppsatser utifrån användes dessa inte alls eller först i efterhand för att motivera betyget. Istället för att ta hänsyn till enskilda kriterier gjorde nästan alla 12 bedömarna i studien en helhetsbedömning vid det slutliga betygssättandet av uppsatser.

Forskning visar också att vetenskapligt skrivande är den del av akademiska studier som studenterna anser är den största utmaningen (Tiihonen, 2015). Vidare visar studier som varit mer studentfokuserade att lärarstudenters frågor vid handledning i högre grad rör vetenskaplig tradition och produktion, än det studerade området eller forskningsfrågan (Gustavsson & Eriksson, 2015). Gustavsson och Eriksson observerade handledningssamtal vid lärarhandledningens examensarbeten. I deras studie dominerades handledningstillfällena av val av undersökningsområde, vad som menas med kritiskt förhållningssätt, karaktären av textproduktion samt användning av referenser. Denna handledning kommer att resultera i att dessa områden,

uttryckt i bedömningskriterier kommer att uppfattas som viktiga av studenten.

Karlsudd (2018) visar att problemformuleringen har avgörande betydelse för resultatet av uppsatsen. Genom att jämföra problemformuleringen i 58 examensarbeten vid lärarutbildningen med det betyg examensarbetet fick visade Karlsudd att en välformulerad problemformulering var nödvändig för ett högt betyg. Karlsudd drar slutsatsen att ett välformulerat syfte och problem är särskilt viktigt för att uppsatsen ska bedömas att hålla hög standard.

Något som kan medföra skillnader i vilka bedömningskriterier som är viktiga för examinatorer är examinatorers ämnesbakgrund. Becher (1994) beskriver disciplin skillnaderna som bestående av olika kulturella karaktärer. Becher bygger sin argumentation på 350 intervjuer med informanter från 12 olika fält (biologi, kemi, ekonomi, ingenjörsvetenskap, geografi, historia, juridik, matematik, moderna språk, farmakologi, fysik och sociologi). Dessa olika fält har gemensamma beröringspunkter kopplade till akademien men har olika traditioner vad gäller hur de ser på kunskap och därmed hur de bedömer. Å andra sidan har en del studier uppvisat små skillnader mellan hur examinatorer bedömer. Bettany-Saltikov, Kilinc och Stow (2009) fann hög interbedömarreliabilitet i bedömningen av masteruppsatser när allmänna, övergripande kriterier användes. Studien var dock begränsad, endast fyra informanter fanns med. På liknande sätt, fann inte heller Kiley och Mullins (2004) någon skillnad som kunde förklaras av ämnesbakgrund.

STUDIENS KONTEXT

De undersökta lärosätenas bedömningskriterier

Samtliga tre lärosäten, där informanterna i vår intervjustudie har sin huvudsakliga sysselsättning, utgår från en modell för bedömning av uppsatser som kan betraktas som målstyrd där förutbestämda kriterier ska utgöra bedömningsgrunden. Syftet är att studenten ska examineras utifrån hur väl de uppfyller kurskraven. Modellen anses som lämplig eftersom bedömningsprocessen blir mer strukturerad samt tydlig för studenterna (Sadler, 2009). Bettany-Saltikova m.fl. (2009) betonar dock att trots att det finns officiella kriterier kommer examinatorer att utveckla egna bedömningskriterier. Dessa kan vara påverkade av kulturen på fakulteten men även av examinatorns egna erfarenheter. Även om målen på de tre lärosätena till viss del uttrycks på olika sätt, finns det tydliga kopplingar mellan prestation och betyg utifrån kursmålen, utan jämförelser med andra studenters prestationer (jmf Sadler, 2005).

De tre lärosätena använder en modell där mål/förväntade studieresultat presenteras i termer av vad studenten ska kunna efter kursen. Målen/de

förväntade studieresultaten är uppbyggda enligt principen att studenterna ska kunna genomföra de olika delarna av en forsknings-/undersökningsprocess. Att formulera problem, planera undersökningen, välja och motivera sitt litteratururval, analysera och presentera resultatet samt att kritiskt kunna granska sitt eget och andras arbeten ingår i samtliga kursplaner kopplade till konsumtionsarbeten. På detta sätt finns det hög överensstämmelse mellan målen/de förväntade studieresultaten jämfört med traditionella examensarbeten.

Det som skiljer mellan de tre lärosätena är att lärosäte 3 betonar att slutsatserna ska dras även utifrån teorier, något som saknas i målen för de övriga två lärosätena. På lärosäte 3 ska resultaten diskuteras i förhållande till yrkesverksamheten, på lärosäte 2 ska relevansen för yrket vara tydligt framskriven medan det på lärosäte 1 ska planeras, motiveras och genomföras enligt praxis för det utbildningsvetenskapliga området. Likheter mellan de tre lärosätenas förväntade studieresultat är dock avsevärt större jämfört med skillnaderna.

Denna studie utgår från antagandet att användningen av kriterier är en form av bedömningspraktik. Vi är medvetna om att begreppet bedömningspraktik kan ses som bredare än denna begränsning. I denna studie finner vi det dock användbart att begränsa bedömningspraktiken till informanternas åsikter om användandet av kriterier. Detta val har gjorts utifrån de metoder vi använder, intervju och enkät.

METOD

Studien har genomförts i två steg, där det första utgör en intervjustudie och det andra en enkätstudie. Svaren i intervjuerna användes för att konstruera den webbaserade enkäten. Data från båda insamlingsmetoderna har använts i den här artikeln.

Som en förberedelse till intervjuerna fick varje informant läsa fem till åtta uppsatser. En majoritet av uppsatserna som lästes var konsumtionsuppsatser, men två informanter fick istället läsa produktionsuppsatser. Orsaken till detta var att få fram alla möjliga kriterier, om dessa eventuellt skulle skilja sig åt mellan de två uppsatstyperna. Vissa av uppsatserna var kända för informanten sedan tidigare, till exempel genom att ha examinerat dem, medan andra var nya. Urvalet av uppsatser skulle ge en variation i uppsatsernas kvalitet, det vill säga skulle ha fått olika betyg vid examination, dock minst godkänt. Motivet till detta urval av uppsatser var att tydligare kunna urskilja vad som skiljer ett bra arbete från ett mindre bra.

Intervju utifrån två metoder

Sexton lärare på tre olika lärosäten intervjuades enskilt med hjälp av en kombination av comparative judgement (CJ) (Pollitt, 2012; Thurstone, 1927) och repertory grid-technique (RGT) (Kelly, 1955). Dessförinnan hade tre pilotintervjuer genomförts. Studiens datainsamlingsmetoder, CJ och RGT, beskrivs närmre i Björklund, Stolpe och Lundström (2016). Urvalet av informanter gjordes med syfte att få informanter som disputerat i olika ämnesdiscipliner samt med olika lång erfarenhet av examination. Lärarna representerade genom sitt disputationssämne ämnesdisciplinerna pedagogik, pedagogiskt arbete, ämnesdidaktik, naturvetenskap, samhällsvetenskap, estetisk verksamhet samt etnologi. Gemensamt för informanterna var att de var aktiva inom lärarutbildningens självständiga arbeten. Intervjuerna genomfördes på informantens arbetsplats och tog 35-90 minuter. Intervjuerna ljudinspelades och transkriberades. Bakgrundsfakta i form av disputationssämne, disputationår, handledarerfarenhet och examinationserfarenhet samlades också in.

I comparative judgement får informanten göra en helhetsbedömning av uppsatsen. Genom att låta informanterna jämföra uppsatserna två och två och besvara frågan: ”vilken är bäst?”, görs en ranking av de olika uppsatserna. Pollitt (2012) anser att det holistiska synsätt som används i CJ ger hög validitet och metoden har visat sig nå hög interbedömarreliabilitet jämfört med då enskilda kriterier används (Kimbell, 2009a,b; Kimbell m.fl., 2009).

RGT som utvecklades av Kelly (1955) har använts inom flera forskningsfält för att undersöka individers personliga konstrukt (Svennberg m.fl., 2014). Ett konstrukt är ett mentalt objekt eller en föreställning som skapar mening. Dessa konstrukt och system av konstrukt används för att se skillnader i händelser, objekt och personer och därmed förstå och ge mening åt omvärlden (Kelly, 1955). Kellys definition av konstrukt som något som används för att se skillnader och ge mening, medför att ett konstrukt kan ses som ett kriterium. Eftersom ett kriterium är mer känt för våra informanter än konstrukt har kriterium använts som begrepp vid intervjuer och enkät. Kelly introducerade difference-modellen för att ta fram bipolära personliga konstrukt. I denna modell beskriver informanten hur det tredje elementet i en triad skiljer sig från de två övriga. Detta är den vanligaste modellen för RGT. RGT har använts även i Sverige (Björklund, 2008; Svennberg m.fl., 2014). Svennberg m.fl. använde RGT för att fastställa vilka kriterier idrottslärare använder vid sin betygssättning. De ser RGT som en möjlig teknik för att få examinatoreer att verbalisera sina individuella, ibland implicita, kriterier. RGT utgör därmed den teoretiska basen för studien. RGT försöker förklara varför individer har olika syn på och attityder till händelser i världen. Under uppväxten kommer individer att använda olika sätt att konstruera en förståelig värld. Ett konstrukt används alltid i form av motpoler, något som är bra ställs i jämförelse med något som är dåligt (Kelly, 1955).

Vid intervjun gjorde informanten först en ranking av de olika uppsatserna utifrån en helhetsbedömning och jämförelse mellan olika arbeten (CJ). Informanten fick sedan betygsätta arbetena på en relativ skala 1-9. Därefter jämfördes tre uppsatser i taget (RGT) där informanten skulle ange vilken av uppsatserna som avvek på något sätt ifrån de andra två, positivt eller negativt. Informanten angav detta, förklarade varför och därefter förhandlades ett namn på det konstruktet fram mellan intervjuare och informant, varefter ett kriterium fastställdes. Kriteriets motsatspar fastställdes, till exempel gott språk-slarvigt språk, väl underbyggda slutsatser-dåligt underbyggda slutsatser. Sedan fick informanten placera ut de lästa uppsatserna på en relativ skala 1-9 efter hur väl de olika uppsatserna uppfyllde kriteriet. Därefter upprepades processen 5-10 gånger, beroende på hur många kriterier informanten kunde ta fram. I vissa fall fick informanten ta ställning med hjälp av kriterier som fastslagits i tidigare intervjuer. Därmed genererade intervjuerna både en inledande helhetsbedömning och en bedömning av olika kriterier som informanten angett utifrån de lästa uppsatserna (Lundström, Åström, Stolpe & Björklund, 2016; Stolpe, Björklund, Lundström & Åström, i arbete). I citaten nedan betecknas informanterna vid intervjuerna med en bokstav och tre siffror (exempel M202).

Webbaserad enkät

Den webbaserade enkäten som konstruerats utifrån de konstrukt som framkom vid intervjuerna, besvarades av 66 examinatorer från sex lärosäten med lärarutbildning (svarsfrekvens 36,9 %). Av de 66 som besvarade enkäten var 36 kvinnor och 30 män. En bortfallsanalys visade att ingen variabel som kön, lärosäte, ämnesbakgrund eller antal examinerade uppsatser var över- eller underrepresenterad i bortfallet. Totalt var det 12 informanter som både blev intervjuade och besvarade enkäten.

Webbenkäten utformades som en Q-sortenkät (Brown, 1997; Stephenson, 1953). Q-sortenkäter bygger på att informanten får ta ställning till ett antal kriterier. Kriterierna hämtades från RGT-intervjuerna. Ursprungligen framkom sammanlagt 92 kriterier vid intervjuerna. Flera av dessa påminde om varandra även om de inte var exakt lika i ordalydelse. Därefter kategoriserade och reducerade två av författarna till denna artikel tillsammans de 92 kriterierna. Stephenson (1953) betonar att en Q-sortenkät inte ska innehålla för många kriterier eller kriterier som är väldigt lika, därför minskades dessa till 52. Kriterier togs bort eller slogs samman när det fanns två kriterier som var lika varandra även om de inte var exakt likadant formulerade. Exempelvis slogs *vetenskaplig förankring* och *god forskningsförankring* samman till *forskningsförankring*. Kriteriet *bra litteratur* gick in under *relevant forskningslitteratur*. Ytterligare 7 av de 52 kriterierna som framkommit vid intervjuerna reducerades för att göra enkäten något kortare och mer lätthanterlig för dem som besvarade den. Exempelvis togs *diskussion om*

forskningsetik bort eftersom *forskningsetik* fanns med. Efter denna gallring fanns 45 kriterier kvar.

En pilotenkät tillverkades och testades av nio informanter. Enkäten var internetbaserad och bestod av en Q-sortdel och en frågedel där uppgifter om kön, disputationssämne, lärarexamen samt erfarenhet som handledare och examinator samlades in. Frågan som ställdes var: *Vad tycker du är viktigast för att beskriva en god uppsats/examensarbete?* Informanterna fick i Q-sortdelen ta ställning till ett av kriterierna i taget och göra en grovsortering av kriterierna i mycket viktiga, viktiga och mindre viktiga. Efter denna första grovsortering finsorterade informanterna kriterierna på en 11-gradig normalfördelad skala från 0-10. Det var inte möjligt att sätta samtliga kriterier som viktiga eller mycket viktiga, utan det var nödvändigt att använda hela skalan. I slutet av enkäten fanns en fråga om det fanns något eller några av kriterierna som är viktigare att använda sig av vid bedömning av en konsumtionsuppsats.

Kriterierna delades sedan in i 10 kategorier, först av en av författarna och sedan av de övriga författarna till denna artikel. Fem av kriterierna kategoriserades på ett annorlunda sätt av en i forskargruppen jämfört med den första kategoriseringen. Dessa fem kriterier diskuterades sedan i hela forskargruppen där man till slut enades om kategorisering. Kriterierna kategoriserades under följande rubriker: *forskningsanknytning, problemformulering, teorimedvetenhet, metod, teori-empiri, språk och formalia, svårighetsgrad, genomförande och slutsatser, helhet* samt *övrigt*. Samtliga kategorier förutom övrigt är hämtade från tidigare undersökningar av examensarbeten (Forsberg & Lundgren, 2006; Råde, 2014). Övrigt-rubriken tillkom när det inte tydligt gick att koppla till någon av de övriga kategorierna. Som exempel kan nämnas nya *fynd, ambitiös* och *god motivering*.

Data analyserades genom att förekomsten av de olika kriterierna i respektive kategori som fastställdes i intervjuerna beräknades. Därefter analyserades intervjuerna med syfte att upptäcka informanternas uppfattningar om likheter och skillnader mellan bedömning av konsumtionsuppsatser och produktionsuppsatser, i enlighet med studiens syfte. För att ytterligare kunna uppnå syftet med studien, ställdes sist i enkäten en fråga angående om det fanns något eller några av kriterierna som var extra viktiga att beakta vid bedömning av en konsumtionsuppsats. Enkätdata analyserades med hjälp av Standard Packages for Social Sciences (SPSS) där medelvärde och standardavvikelse för varje kriterium och kategori beräknades. Övriga resultat från studien redovisas i Lundström m.fl. (2016) (2016) och Stolpe m.fl. (i arbete).

RESULTAT

I huvudsak ser informanterna i både intervjustudien och enkätstudien små skillnader mellan att bedöma en produktionsuppsats och en konsumtions-

uppsats. Kriterier och diskussioner kring dessa rör sig allt som oftast på ett generellt plan där resonemangen går att tillämpa oavsett vilken typ av uppsats eller på vilken nivå (grund-, eller avancerad) det gäller. De klassiska kriterierna kopplade till *forskningsförankring, syfte, frågeställningar, metod, slutsatser* och *formalia* framkommer också i de flesta intervjuer och betonas oavsett vilken typ av uppsats det rör sig om. *Röd tråd, god struktur* och *tydlighet* är de helhetsbegrepp som oftast poängteras.

I enkäten valde 18 av de 66 informanterna (27 %) att ange minst ett kriterium som de ansåg vara viktigare vid en konsumtionsuppsats. Det är således en minoritet som väljer att ange minst ett särskiljande kriterium, vilket indikerar att många examinatorer inte ser någon skillnad i att granska konsumtionsuppsatser jämfört med produktionsuppsatser, vilket också framkom i intervjuerna.

I intervjuerna framkommer dock en del skillnader i uppfattningar om de två uppsatstyperna och vilka konsekvenser detta får för bedömningen av uppsatserna. Dessa hänger huvudsakligen ihop med antingen hur strukturen byggs upp, vad som är metod eller hur forskningsförankringen är framskriven. Skillnaderna som anges är analyserade utifrån ett bedömningsperspektiv men är också ett uttryck för att det är skillnad i uppsatstyperna och sättet att skriva dem på. Under nedanstående rubriker redovisas resultat från intervju och webenkät. Utgångspunkten har varit att redovisa sådant som framkommit frekvent i intervjuerna och/eller i webenkäten. Fokus ligger på de skillnader som framkommit, även om resultaten visar att informanterna huvudsakligen ser små skillnader mellan att bedöma konsumtionsuppsats och produktionsuppsats.

Uppsatsens struktur

Ett kriterium som återkommer i flera intervjuer är uppsatsens struktur. Informanterna återkommer till att det blir skillnad på strukturen i de båda uppsatstyperna och att konsumtionsuppsatserna ofta är lättare att strukturera. M222 ser konsumtionsuppsatserna som enklare att skriva eftersom mer är givet i en konsumtionsuppsats och det blir mindre risk att något går fel.

På sätt och vis skulle man kunna säga att det är enklare med en konsumtionsuppsats. För ofta är det ju när de ska systematisera sitt material som det går snett, eller att de inte lyckas hitta någon form av princip eller teman så blir det ingenting. Här tycker jag att det är ganska prydligt. Man har hittat någon typ av struktur i de här texterna. Men sedan har de blivit producerade, reproducerande och sedan har man inte kommit längre.

För M222 är systematisering och struktur viktigt, hen menar att det finns en tydligare mall att följa i ett konsumtionsarbete men att detta också kan

medföra att texten endast blir en reproduktion av tidigare texter, vilket får konsekvenserna för bedömningen.

På ett liknande sätt lyfter L003 upp en risk som hen anser finns vid konsumtionsuppsatser, nämligen att uppräkningsen av de olika studierna blir långrandig.

Jenny och Laura har fått en trevlig struktur på sin uppsats. Dels tydliggör de i sin konsumtionsuppsats de enskilda studierna de har analyserat på ett föredömligt sätt bättre än vad jag sett andra. Ja, de har hittat ett exempel på att skriva en konsumtionsuppsats där de enskilda bidragen lyfts fram utan att det blir stackato-mässigt.

För både M222 och L003 blir undvikande av uppräkningsen med risk för långrandighet en viktig aspekt vid deras bedömning av en konsumtionsuppsats. L003 har varit handledare eller examinator till Jenny och Laura och poängterar att de lyckats bättre än andra arbeten hen tidigare granskat i det avseendet.

I enkätsvaren förekommer endast två svar som kan kopplas till uppsatsens struktur. Det ena svaret betonar att det ska vara lätt att få översikt över uppsatsen, vilket kan kopplas till uppsatsens struktur. Det andra enkätsvaret rör tematisering, som också hänger samman med struktur. Struktur nämns inte explicit i något av enkätsvaren.

Metod

I en konsumtionsuppsats samlas inte egen empiri i form av intervjuer, enkäter eller liknande utan det är insamlandet och analys av texter som är i fokus. Detta får konsekvenser för kategorin metod. Metod är den kategori som oftast tas upp i intervjuerna. Informant L079 vill att studenterna ska förstå skillnaden mellan det som vanligtvis ses som metod i en produktionsuppsats och den metod som används i en konsumtionsuppsats.

Jag har ju löst problemet genom att inte kalla, när jag skriver konsumtionsuppsats så kallar vi det inte metod. I min föreläsning säger jag att vi kallar det inte för metod bara för att vi inte ska förväxla med produktionsuppsatser utan vi kallar det för tillvägagångssätt eller genomförande.

L079 har alltså löst dilemmat med att empirin och därmed även metoden är annorlunda i en konsumtionsuppsats genom att undvika begreppet i samband med handledning av konsumtionsuppsatser. Ett enskilt kriterium som lyfts fram gällande metod görs av M222 som visar på svårigheten att i en konsumtionsuppsats synliggöra metoden.

Uppsats E beskriver bara sökvägen, det finns ingen metod. Den saknar, metodmedvetenhet handlar inte bara om att beskriva sökvägen utan dels också som lärandemål 2 beskriver, handlar om att det framgår på vilka grunder man har granskat och värderat källorna, inte bara redogjort.

De två citaten ovan visar på olikheter i vad som räknas som metod. Medan L079 inte alls vill kalla något i en konsumtionsuppsats för metod, efterlyser M222 en tydligare framskrivning i metod i den granskade konsumtionsuppsatsen. Detta visar på två olika sätt att angripa metodproblematiken i en konsumtionsuppsats. Bedömningskriterier kopplade till metod kommer därmed att skilja både mellan uppsatstyperna och mellan olika examinatorer.

Även L088 tar upp metod och problematiken med begreppet. Fast i detta fall är det en problematik som förekommer även i en produktionsuppsats.

Sedan kan man gå lite djupare att det ska finnas en insamlings och en analysmetod. // Men det är hur man skriver fram sin metod. // alltså att man är medveten om att det finns en insamlingsmetod och en analysmetod och att de två inte är samma.

L088s uppdelning i insamlings- och analysmetod används för att synliggöra att begreppet metod är väldigt brett och kan användas i olika skeden av uppsatsskrivandet.

I enkätsvaren är metod en av de kategorier som oftast anges. Kriterier som är kopplade till kategorin metod nämns i fem av enkäterna när informanterna ska ange något som är viktigare att använda sig av vid bedömning av en konsumtionsuppsats. I enkätsvaren nämns som exempel på metod att sökvägar måste redovisas tydligt och att systematiken är extra viktig i en konsumtionsuppsats.

Forskningsförankring

Nära kopplat till bedömningen av kategorin metod är *forskningsförankring* eftersom metoden i en konsumtionsuppsats är sökning och analys av tidigare forskning. Det är inte helt lätt att skilja vad som tillhör vilken kategori när metod och forskningsförankring jämförs. Ett sätt är att se insamlingen och ett första urval som metod och en djupare analys och hur denna skrivs fram i uppsatsen som forskningsförankring. Forskningsförankringen förväntas att ske genom en *systematisk genomgång* och *analys av relevanta källor*. Förutom litteraturkriteriet tillhör även kriterier som *vetenskaplig förankring*, *kännedom om tidigare forskning*, *användning av primärkällor* och *systematisk användning av källor* forskningsförankringskategorin.

När det diskuteras litteratur är det även kopplat till sökprocessen. Sökprocessen kan sägas ligga någonstans i gränslandet mellan

forskningsförankring och metod. Men det behöver inte vara själva sökningen som är svår utan hur litteraturen sedan används. L003 ser litteraturdelen som ämnesdjupet som ska innehålla ett kritiskt förhållningssätt.

Alltså vi har kört lite hårdare med dom och dom har ju då istället visat en stark självständighet och hittat litteratur själva och börjat bedöma litteratur. // Men i det här ämnesdjupet så har dom kommit häpnadsväckande långt // jag skulle vilja säga med att lämna den här Studentlitteraturambitionen.

Även M205 ser vissa skillnader i bedömningen av en konsumtionsuppsats jämfört med en produktionsuppsats. De flesta av de intervjuade lyfte fram de delar av en uppsats som tidigare nämnts som de stora skillnaderna. Men för M205 är det dessutom viktigt att studenterna är medvetna om att en kunskapsöversikt kommer att påverka hur forskningsfrågorna skrivs fram.

När det gäller SAG-arbetena (konsumtionsuppsatser) så är det två saker minst som inte kommit fram. En väldigt kritisk aspekt är systematiken i sökningarna. Men också medvetenheten om att det är en kunskapsöversikt. Den här medvetenheten över att det är en kunskapsöversikt påverkar också hur man ställer, vilket syfte man har när man ställer frågor. //

Men det handlar också om källorna, man ska söka och sedan göra ett urval. Och både själva sökningsprocessen och urvalsprocessen är viktiga delar. Och ur urvalet bör det komma en lista på vilka man valt, vilket ingen gör explicit.

Utifrån detta resonemang kommer det redan från början vara viktigt att ta hänsyn till att det är en kunskapsöversikt/konsumtionsuppsats både för studenten som skriver uppsatsen och examinatorn som ska bedöma den.

Liksom i kategorin metod är det fem informanter som i enkätsvaren angett något kriterium som extra viktigt vid en konsumtionsuppsats som tillhör kategorin forskningsförankring. Här betonas nödvändigheten att använda internationell forskning, att använda sig av primärkällor, men också att kunna vara kritisk mot tidigare forskning.

Sammanfattning av resultat

De flesta intervjuade tar upp kriterier som berör delarna som är mitt i processen, inte de i början som syfte och forskningsfrågor, eller de i slutet som slutsatser och diskussion. Samma resultat finns i enkäterna där det endast är enstaka informanter som har betonat kriterier eller kategorier som tillhör

delar i början eller i slutet av uppsatsprocessen som viktigare vid en konsumtionsuppsats.

Huvudfokus i denna artikel är inte skillnader mellan olika examinatorer. Men vi konstaterar ändå att enkätsvaren visade små skillnader mellan examinatorer från de olika ämnesområdena. Att ha disputerat inom ett visst ämnesområde tycks inte påverka vilka kriterier som examinatorn anser som viktigare. Det finns inte heller något samband mellan vilket lärosäte examinatorn arbetar på och vilka kriterier som betonas.

Sammanfattningsvis indikerar resultaten i både intervjuer och enkät att det som informanterna anser skilja mellan bedömning av en konsumtionsuppsats och en produktionsuppsats huvudsakligen är förknippat med de delar som har att göra med sökningen och redovisningen av tidigare gjord forskning. Detta kan vara både vilken metod som används och hur sökningen analyseras och redovisas. Kriterierna som framkommer vid intervjuerna påminner till stor del om de kriterier som brukar användas vid bedömning av produktionsuppsatser. Att endast 27 procent av informanterna väljer att nämna något kriterium som är viktigare vid en konsumtionsuppsats antyder också att examinatorer anser att de vid bedömningen av de två uppsatstyperna använder sig av samma bedömningskriterier.

DISKUSSION

Resultaten visar att många examinatorer i vår studie inte ser stora, klara skillnader mellan vilka kriterier de använder vid bedömning av de båda uppsatstyperna. Detta kan tolkas som att uppgiften för studenterna är att producera en vetenskaplig uppsats och en sådan ska uppfylla vissa kriterier oavsett nivå eller uppsatstyp. Dessa bedömningskriterier är vedertagna och ser ut på ett liknande sätt inom stora delar av akademien (Sadler, 2009). De är kopplade till forskningsanknytning, metod, språk och formalia, genomförande och slutsatser samt helhet. Det rör sig om att ha undersökbara frågor, relevant metod, redovisa resultat tydligt samt dra rätt slutsatser med hjälp av korrekt akribi. Detta är tidigare rapporterat av Bettany-Saltikov m.fl. (2009), som fann hög interbedömarreliabilitet i bedömningen av masteruppsatser när allmänna, övergripande kriterier användes. De skillnader som lyfts fram i vår studie är ofta kopplade till hur tidigare forskning behandlas i metod eller forskningsförankring eftersom denna är central i en kunskapsöversikt, vilken en konsumtionsuppsats räknas till. Det är huvudsakligen bedömningen av de mellersta delarna i skrivandet av en konsumtionsuppsats som mest skiljer sig från en produktionsuppsats, inte inledningens syfte och forskningsfrågor eller avslutningens slutsatser och diskussion. I relation till Gustavsson och Eriksson (2015), som analyserade handledningssamtal, visar vår studie på vissa likheter jämfört med deras.

Liksom i deras studie, rör sig frågor som kommer upp mest kring vetenskaplig tradition och produktion.

Kopplingen teori-empiri var sparsamt representerade i intervjuerna, liksom teorimedvetenhet. Att enbart använda tidigare forskning ses inte tydligt som empiri av informantgruppen. I konsumtionsuppsatsen antyder informanterna att bedömningen av kopplingen mellan empiri och teori uppfattas annorlunda jämfört med produktionsuppsatsen. Detta kan jämföras med den granskning HSV gjorde 2006 (Forsberg & Lundgren, 2006) där produktion av forskning i form av empiri var kännetecknande för lärarutbildningens examensarbeten. Men förklaringen skulle också kunna vara att informanterna istället har beskrivit empiridelarna i de delar som vi kategoriserat till genomförande och slutsatser. Här finns kriterier som bearbetning och analys av empirin. I så fall kan empirin i form av tidigare forskning ses som viktig men till viss del som fristående från teoridelen. Den mest troliga tolkningen till att teori inte nämns speciellt ofta i våra intervjuer och i vår enkät är att detta inte ses som speciellt viktigt vid konsumtionsuppsatser. En anledning skulle kunna vara att konsumtionsuppsatsen på de undersökta lärosätena skrivs så tidigt i utbildningen så studenterna inte kan förväntas kunna föra teoretiska resonemang samt att teorikriterier inte är så tydligt framskrivna i samtliga lärosätes betygskriterier. En alternativ tolkning är att eftersom teorikriterier knappt nämns i enkäterna anses de som lika viktiga i båda typerna av uppsatser. Men i så fall skulle de kommit upp fler gånger i intervjuerna eftersom de granskade uppsatserna bör innehålla teoriresonemang med olika kvalitet och därmed nämnts som något som skiljer granskade uppsatser åt.

Strukturen på en konsumtionsuppsats ses som viktig att bedöma men också svår för studenten att få rätsida på enligt några av informanterna. Bristen på egen insamlad empiri i form av intervjuer eller observationer medför att de data som finns är i form av olika texter. Några av informanterna beskriver skillnader i metod mellan de två uppsatstyperna. Detta görs explicit av L079 som anser att det är viktigt för studenterna att se skillnaderna redan under perioden de skriver konsumtionsuppsatsen. Det som sedan under skrivandet av en produktionsuppsats huvudsakligen kommer att tillhöra en litteraturbakgrund och där sättet denna kommit till på inte betonas, blir enligt L003 och M222 delar som är viktiga för strukturen och hur studenten har analyserat sitt material. På detta sätt förändras litteraturstudier, genom att under en konsumtionsuppsats ses och bedömas som viktig data som det gäller att kunna redogöra för hur denna valts ut och bearbetats, till att under en produktionsuppsats mer ses och bedömas som en bakgrundsinformation och förankring för arbetet.

Dessutom kan omfattningen av ett kriterium vara väldigt olika. Ett kriterium som god struktur kan dels vara ett eget kriterium men kan också spilla över på andra kriterier. I vårt resultat som redovisas ovan syns till exempel hur struktur får konsekvenser för analysen. Även i enkätsvaren finns

metod bland de mest förekommande svaren hos de 27 % som angett något som är viktigare i en konsumtionsuppsats jämfört med en produktionsuppsats.

Resultaten visar också att vissa informanter anser att en konsumtionsuppsats är mindre komplex för studenten. Arbetsgången betonas som tydligare och lättare att följa än när olika typer av data är inblandade. Samtidigt riskerar detta att göra konsumtionsuppsatserna reproducerande där inget nytt kommer fram. Sökningen och vilka frågor som ställs i samband med denna är avgörande för kvaliteten och därmed bedömningen av uppsatsen. Därmed kommer forskningsfrågor, sökning och analys av källorna att hänga samman och kan ses som en viktig röd tråd i en konsumtionsuppsats. Våra resultat visar att införandet av konsumtionsuppsatser på lärarutbildningsprogrammen inte verkar ha medfört att någon ny bedömningspraktik uppkommit när den här studien genomfördes. Konsumtionsuppsatserna är en relativt ny företeelse på lärarutbildningarna och bedömningspraktiken kan komma att förändras när det gått längre tid.

Det finns inte något samband mellan vilket lärosäte man arbetar på och vilka kriterier man lyfter fram som centrala (Stolpe m.fl., i arbete). Detta tyder på att examinatorerna lutar sig mot sin tidigare erfarenhet av bedömning av uppsatser, något som skulle kunna tolkas som att examinatorerna har individuella bedömningspraktiker som väger tyngre än de formella kriterierna. Varje examinator har utvecklat en egen praktik som är svår att förklara endast med hjälp av erfarenhet i form av ämnesbakgrund eller antal examinerade självständiga arbeten. En annan förklaring skulle kunna vara att examinatorerna huvudsakligen använder sig av helhetsbedömningar istället för bedömningskriterier. Användningen av ett holistiskt perspektiv vid bedömningar har tidigare visats av Bloxham m.fl. (2011). En djupare analys av användningen av bedömningskriterierna detta redovisas i Stolpe m.fl. (i arbete) där vi undersökt erfarenhet av handledning och examination i relation till kriteriernas komplexitet.

Metodologisk diskussion och implikationer

Vi har i denna studie använt oss av RGT (Kelly, 1995) som metod. Det innebär att vi ser på individers bedömningspraktik ur ett psykologiskt perspektiv, där examinatorer gör sin bedömning utifrån motsatser och jämförelser. Däremot diskuteras inte huruvida de konstrukt som används är statiska eller föränderliga. Tidigare forskning (Björklund, 2008) visar att det skiljer i bedömning mellan noviser och experter, vilket innebär att de kan vara föränderliga över tid. Trots detta menar vi att resonemang kring konstrukt är intressanta då dessa har god överensstämmelse med bedömningskriterier enligt den modell som användes på lärosätena.

Det som undersökts i denna artikel är huvudsakligen vad examinatorerna uttrycker om bedömning av olika uppsatstyper. Vid intervjuerna hade

informanterna läst och bedömt de uppsatser som diskuteras vid intervjuerna. Det gör att vi anser att det finns en ganska hög överensstämmelse mellan vad de bedömer och vad de säger att de bedömer.

Avslutningsvis vill vi diskutera vilka slutsatser som kan dras om användning av bedömningskriterier utifrån vår studie och tidigare forskning. Det har idag etablerat sig detaljerade bedömningskriterier. Vid våra intervjuer framkom 92 stycken kriterier och vid granskning av våra lärosätes respektive bedömningskriterier var även dessa många, vanligtvis mellan 25-35/lärosäte. Ett sådant högt antal syftar till att kunna täcka bedömningen av alla delar av uppsatsen, men framför allt att på ett explicit sätt visa studenterna vad som kännetecknar en god uppsats. Risken finns att både uppsatsskrivandet och bedömningen förvandlas till en checklista att pricka av. De stora likheterna mellan vad informanterna anser om bedömning av en konsumtionsuppsats jämfört med en produktionsuppsats antyder att en sådan checklista verkar finnas. I en sådan bedömningspraktik riskerar kriterier eller aspekter som kan ses som viktiga men svåra att förklara eller pricka av att försvinna. Några sådana är röd tråd, svårighetsgrad och originalitet. Resultatet i denna studie tyder på att examinatorerna bara i viss omfattning anser att bedömningen av de två uppsatstyperna skiljer sig åt. Kanske kan det vara bra att så som Bloxham m. fl. (2011) föreslår att använda sig av övergripande helhetsbedömningar i kombination av ranking av uppsatserna. Användningen av helhetsbedömningar och ranking av uppsatserna har visat sig att ge större interbedömarreliabilitet. Dock måste detta sätt att bedöma ställas mot tydligheten gentemot studenterna om vad en bra uppsats innebär.

Vi anser att området idag är underbeforskat. Införandet av bedömningskriterier har på en del lärosäten skett eftersom en *constructive alignment*-modell (Biggs & Tang, 2011) har införts. Vi anser att en sådan modell riskerar att dela upp bedömningen och i förlängningen kunskaperna i små enskilda delar. Detta har skett utan att ta hänsyn till forskning som visar att det är svårt att dela upp bedömningen i avgränsade kriterier (Bloxham m fl., 2011).

REFERENSER

- Becher, Tony. (1994). The significance of disciplinary differences. *Studies in Higher Education*, 19(2), 151-162, doi.org/10.1080/03075079412331382007
- Bet. 2009/10:UbU16. *Ny lärarutbildning. Utbildningsutskottets betänkande 2009/10:UbU16.*
- Bettany-Saltikova, Josette., Kilinc, Stephanie., & Stowc, Karen. (2009). Bones, boys, bombs and booze: an exploratory study of the reliability of marking dissertations

- across disciplines. *Assessment & Evaluation in Higher Education*, 34(6), 621–639, doi.org/10.1080/02602930802302196
- Biggs, John., & Tang, Catherine. (2011). *Teaching for quality learning at university*. Maidenhead: Open University Press.
- Björklund, L. (2008). The repertory grid technique: Making tacit knowledge explicit: Assessing creative work and problem solving skills. In H. Middleton (Ed.), *Researching Technology Education: Methods and techniques*. Netherlands; Sense Publishers.
- Björklund, L., Stolpe, K., & Lundström, M. (2016). Making tacit knowledge explicit: Three methods to assess attitudes and beliefs. In J. Lavonen, K. Juuti, J. Lampiselkä, A. Uitto & K. Hahl (Eds.), *Electronic Proceedings of the ESERA 2015 Conference. Science education research: Engaging learners for a sustainable future*, 1733-1741. Helsinki, Finland: University of Helsinki. ISBN 978-951-51-1541-6.
- Bloxham, Sue., Boyd, Peter., & Orr, Susan. (2011). Mark my words: the role of assessment criteria in UK higher education grading practices. *Studies in Higher Education*, 36(6), 655-670, doi.org/10.1080/03075071003777716
- Brown, Steven. R. (1997). *The History and Principles of Q Methodology in Psychology and the Social Sciences*. Hämtad på <http://facstaff.uww.edu/cottlec/QArchive/Bps.htm> 150209.
- Forsberg, Eva. (2012). Formation of links to research through thesis writing: The case of the Swedish teacher education. *Paideia*, 50 (71-102).
- Forsberg, Eva., & Lundgren, Ulf.P. (2006). *Examensarbetet inom den nya lärutbildningen*. Högskoleverkets rapportserie 2006:47 R. Högskoleverket.
- Gustafsson, Christina., & Hallström Monica. (2005). Examensarbetet inom lärutbildningen – En analys i relation till högskolelagens mål. *Studies in Educational Policy and Educational Philosophy* 2, 1-21, doi.org/10.1080/16522729.2005.11803904
- Gustavsson, Susanne., & Eriksson, Anita. (2015). Blivande lärares frågor vid handledning - Gör jag en kvalitativ studie med kvantitativa inslag? *Pedagogisk Forskning i Sverige*, 20(1-2), 79-99.
- Hjort, Magnus., & Sundkvist, Maria. (2010). *Högskoleverkets system för kvalitetsutvärdering 2011–2014*. Högskoleverkets rapportserie 2010:22 R. Högskoleverket.
- Karlsudd, Peter. (2018). Att problematisera ”problemet”: Bedömning och utveckling av problemformuleringar i lärutbildningens självständiga arbeten.

Nordic Journal of Vocational Education and Training, 8(1), 1–22, doi:
10.3384/njvet.2242-458X.18811

Kelly, Georg. A. (1955). *The psychology of personal constructs*. New York: Routledge.

Kiley, Margaret., & Mullins, Gerry. (2004). Examining the examiners: How inexperienced examiners approach the assessment of research theses. *International Journal of Educational Research*, 41, 121-135, DOI: 10.1016/j.ijer.2005.04.009

Kimbell, Richard. (2009a). e-assessment in project e-scape. *Design and Technology Education: An International Journal*, 12(2), 66-76.

Kimbell, Richard. (2009b). Holism and the Challenge of Teachers' Judgement. *Design and Technology Education: An International Journal*, 14(1), 5-6.

Kimbell, Richard., Wheeler, Tony., Stables, Kay., Shepard, Tristram., Martin, Fred., Davies, Dan., & Whitehouse, Gillian. (2009). *e-scape portfolio assessment: phase 3 report*. London: Goldsmiths, University of London.

Lundström, M., Åström, M., Stolpe, K. & Björklund, L. (2016). Assessing student theses: Differences and similarities between examiners from different academic disciplines. *Practitioners Research in Higher Education*, 10 (1), 217-226.

Pollitt, Alistair. (2012). Comparative judgement for assessment. *International Journal of Technology and Design Education*, 22, 157-170.

Prop. 2009/10:139. *Fokus på kunskap – kvalitet i den högre utbildningen*.

Råde, Anders. (2014). Ett examensarbete för både yrke och akademi – En utmaning för lärarutbildningen. *Högre utbildning*, 4(1), 19-34.

Råde, Anders. (2016). Fågel, fisk eller mittemellan? –Handledares uppfattningar om lärarutbildningens examensarbete. *Högre utbildning*, (2), 139-155.

Sadler, D.Royce. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment & Evaluation in Higher Education*, 30(2), 175–194, doi.org/10.1080/0260293042000264262

Sadler, D.Royce. 2009. Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education* 34(2), 159–79, doi.org/10.1080/02602930801956059

Schwab, Joseph. J. (1964). The structure of the Natural Sciences. In J. C. Parker (Ed.), *The structure of Knowledge and the Curriculum* (Second ed., pp. 31-49). U.S.A: Rand McNally & Company.

Scriven, Michael. (1964). The structure of the Social Studies. In J. C. Parker (Ed.), *The structure of Knowledge and the Curriculum* (Second ed., pp. 87-105). U.S.A: Rand McNally & Company.

- SOU 2008:109. *En hållbar lärarutbildning*.
- Stephenson, William. (1953). *The study of behavior: Q-technique and its methodology*. Chicago: University of Chicago Press.
- Stolpe, K., Björklund, L., Lundström., & Åström, M. (I arbete). Different profiles for assessment of student thesis in teacher education.
- Svennberg, Lena., Meckbach, Jane., & Redelius, Karin. (2014). Exploring PE teachers' 'gut feeling': An attempt to verbalise and discuss teachers' internalized grading criteria. *European Physical Education Review*, 20(2), 199-214, DOI:10.1177/1356336X13517437
- Svärd, Ola. (2013). *Examensarbetet- en kvalitetsindikator inom högre utbildning?* Doktorsavhandling vid Uppsala universitet.
- Säljö, Roger. & Södling, Maria. (2006) *Utbildning på vetenskaplig grund: röster från fältet*. Högskoleverkets rapportserie 2006:46 R. Högskoleverket.
- Thurstone, Louis. (1927). A law of Comparative judgement. *Psychological Review*, 34, 273-286.
- Tiihonen, Susanne. (2015). ”Studera effektivare” – Om att stöda studenters självreglering och studiefärdigheter. I: Monica Londen, Åsa Mickwitz & Susanne Tiihonen. (Red.). *Hur svårt kan det vara? En antologi om hur lärare inom den högre utbildningen förnyar undervisningen* (s. 133-153). Nordica Helsingiensia. Helsingfors: Unigrafia.
- Young, Suzanne. (2011). A Survey of Student Assessment Practice in Physical Education: Recommendations for Grading. Strategies: *A Journal for Physical and Sport Educators*, 24(6), 24-26, doi.org/10.1080/08924562.2011.10590959