

STUDENTEVALUERINGER I HØYERE UTDANNING:

På hvilken måte og til hvilken nytte?¹

JARLE AARSTAD²

Høgskolen i Bergen
Avdeling for ingeniøruddanning
Postboks 7030
Nygårdsgaten 112
5020 Bergen
Telefon: 555 87926
E-post: jarle.aarstad@hib.no

-
- 1 Denne artikkelen bygger på en oppgave fra et kurs i høgskolepedagogikk (15 studiepoeng) høstsemesteret 2005 og vårsemesteret 2006 ved Høgskolen i Bergen, Avdeling for lærerutdanning. Undertegnede står imidlertid alene ansvarlig for den informasjon og de meninger som fremkommer i besvarelsen.
 - 2 Jarle Aarstad er ansatt som førsteamanuensis i organisasjonsfag ved institutt for økonomiskadministrative fag.

Sammendrag

I denne oppgaven retter jeg primært søkelyset mot hvordan formaliserte studentevalueringer i høyere utdanning kan og bør brukes, og i forlengelsen av dette stiller jeg også spørsmål ved nytteverdien av informasjonen som fremkommer. Innledningsvis gjør jeg rede for en del forskningsbidrag som er foretatt på tematikken studentevalueringer. I fortsettelsen gjør jeg noen analyser av en midtevaluering jeg nylig foretok i et kurs, og i avsluttende del kommer jeg med noen oppsummerende refleksjoner. Både internasjonal forskning, samt egne analyser, viser klart at forhold langt utover faglærers didaktiske og pedagogiske kvalifikasjoner er med på å influere studentenes evalueringer av undervisningen. Jeg diskuterer også hvilke mulige negative konsekvenser ukritisk bruk av studentevalueringer kan ha for faglig nivå og faglig integritet i høyere utdanning.

Nøkkelord

Studentevalueringer, høyere utdanning, litteraturstudium, faktoranalyser, korrelasjonsanalyser

Innledning

Studentevalueringer har i løpet av de senere årene blitt et stadig viktigere element i vurdering av undervisningskvalitet og faglig innhold ved de fleste universiteter og høyskoler i Norge. Denne tendensen gjør seg nå også gjeldene i de fleste vestlige land (Moore, Eng, & Daniel, 2003). I vår norske kontekst er studentevalueringer ett av flere tiltak i Kvalitetsreformen for å sikre nivået i høyere utdanning, og bør derfor i all hovedsak hilses velkommen av både studenter så vel som faglærere. Studentevalueringer kan indikere styrker og svakheter i faglig opplegg og didaktisk formidlingsevne hos foreleser (Hand & Rowe, 2001). Det gir også en formalisert arena for tilbakemeldinger fra studentene, og kan derved skape rom for opplevd prosedyrerettferdig hos denne gruppen (Kravitz, Stone-Romero, & Ryer, 1997). Til sist kan det også fungere som et insentivsystem for faglig og pedagogisk kvalitet, noe som i mange tilfeller vil være en positiv motivasjonsfaktor hos undervisningspersonalet (Moore et al., 2003).

I denne oppgaven vil jeg spesielt rette søkelyset mot hvordan formaliserte studentevalueringer i høyere utdanning kan og bør brukes, og i forlengelse av dette stiller jeg spørsmål ved nytteverdien av den informasjonen som fremkommer. Det er viktig å ha et reflektert forholdt til dette, siden datamaterialet som blir tilveiebrakt – enten i kvantitativ eller kvalitativ form – i all hovedsak er en avspeiling av studentenes *individuelle persepsjoner*. Målsettingen for alle parter må være at studentevalueringer fungerer etter hensikten; at informasjonen som fremkommer i best mulig grad gjenspeiler faktiske forhold og at gjennomgående styrker og svakheter i det faglige opplegget avdekkes. Jeg presiserer at jeg ikke vil vektlegge hvordan et instrument for studentevalueringer i detalj bør utformes. I større grad vil jeg redegjøre for de prinsipielle sidene ved verktøyet, og i hvilken grad substansen i informasjonen vi fremskaffer kan komme til nytte.

I neste avsnitt foretar jeg en gjennomgang av en del forskningsbidrag som er gjort på temaet og problemområdet studentevalueringer. I påfølgende avsnitt

presenterer og analyserer jeg en midtevaluering som jeg nylig gjorde i et kurs ved Institutt for økonomisk- administrative fag ved Høgskolen i Bergen. Hensikten er ikke at den empiriske illustrasjonen skal gi en uttømmende redegjørelse for hvordan en kan eller bør gå frem for å analysere et slikt materiale, men snarere peke på noen grep som kan benyttes for i best mulig grad å trekke ut substansiell informasjon av tilgjengelige data. I den siste delen foretar jeg en diskusjon av analyseresultatene, og kommer også med noen oppsummerende refleksjoner.

Litteraturgjennomgang

Studentforutsetninger

Med *studentforutsetninger* tenker jeg spesielt på den faglige motivasjonen eller forutsetningen som hver enkelt student måtte ha. Selv om motivasjon og forutsetninger er to selvstendige begreper, henger de imidlertid sammen og kan være vanskelig å behandle separat. Studenter som har gode forutsetninger for å gjennomføre et studium vil naturlig nok være relativt godt motiverte siden de i stor grad mestrer studiene. Motivasjon kan naturlig nok også påvirke resultatene de har oppnådd både før og under høyere utdanning og derved i tur påvirke forutsetningene.

Seiler, Seiler, & Chiang (1999) fant at studenter som hadde høy poengsum ved opptak til college (GPA – grade point average) ga gode studentevalueringer. Radmacher & Martin (2001) fant videre at studenter som var opptatt på studier med høy prestisje, jevnt over ga positive tilbakemeldinger. Siden prestisjefylte studier normalt har høye opptakskrav kan også dette indikere en positiv sammenheng mellom studentenes generelle kvalifikasjoner og tilbakemeldingene de gir til foreleser. Dyktige studenter forventer videre jevnt over gode karakterer, og en rekke studier har påvist en positiv sammenheng mellom slike forventninger og kursevalueringer (Eiszler, 2002; Greenwald & Gillmore, 1997; Isely & Singh, 2005; Kidd & Latif, 2004; Krautmann & Sander, 1999; McPherson, 1982; Radmacher et al., 2001). I forhold til det sistnevnte momentet er imidlertid kausalitetsproblematikken kompleks, noe jeg vil komme tilbake til senere.

Oppsummert kan vi si at empirien indikerer en klar sammenheng mellom studentenes faglige kvalifikasjoner og gode studentevalueringer. Dette kan forklares på flere måter. For det første er det naturlig å anta at dyktige studenter jevnt over har en positiv holdning til det å lære og å tilegne seg nytt stoff. Naturlig nok vil dette smitte over på opplevelsen de har av det som skjer i undervisningssammenheng. Studenter med dårligere forutsetninger, og som derved kan ha vansker med å tilegne seg faget, vil på den annen side kunne ha en tendens til å skyld

på dårlig undervisning. Dette kan forklares ut i fra attribusjonsteori hvor egen suksess forklares som et resultat av personlige egenskaper, mens nederlag blir tillagt eksterne forhold (Greenwald et al., 1997; Kelly, 1973). Motiverte studenter vil også i større grad arbeide med faget både før og etter forelesning, og dette vil i tur gjøre faget enda mer tilgjengelig for denne gruppen.

McPherson (1982) oppdaget forøvrig at dess høyere prosentandel av masterstudenter på en handelshøyskole som deltok i evalueringene, dess mer negative tenderte tilbakemeldingene å være. Dette kan også forklares ut i fra studentenes generelle motivasjonsgrad. Det fremgår ikke av studien hvordan evalueringene i detalj ble utført, men uavhengig av dette mener jeg at det først og fremst er de mest motiverte studentene som uoppfordret ønsker å gi tilbakemelding til faglærer, mens de mindre motiverte ikke bryr seg i samme grad. Om faglærere eller andre derfor anstrenger seg for å få høyest mulig svarprosent (enten ved å benytte insentiver eller å gjøre evalueringene obligatoriske) vil dette også trekke de mindre motiverte studentene inn i datamaterialet.

Generelle studentforutsetninger og motivasjon kan også være en funksjon av modningsnivå. Det kan derfor godt hende at eldre studenter eller studenter med lang yrkeserfaring har en annen opplevelse av undervisningskvaliteten enn sine yngre og mindre erfarne medstudenter. Seiler et al. (1999) oppdaget at deltidsstudenter jevnt over var mer fornøyd med undervisningen enn fulltidsstudenter. Generelt sett er deltidsstudenter ofte både eldre enn fulltidsstudenter og har i tillegg gjerne lang yrkeserfaring. Funnet jeg har referert til kan derfor indikere sammenheng mellom modningsnivå og opplevelse av undervisningssituasjonen. En annen forklaring kan selvsagt være at deltidsstudenter opplever det som et privilegium å studere mens unge fulltidsstudenter mer tar det som en selvfølge. Radmacher et al. (2001) fant imidlertid ingen sammenheng mellom studentenes alder og evalueringene de ga, men utvalget deres var fulltidsstudenter på bachelor-nivå, og her er naturlig nok aldersspredningen liten.

Utover generelle forutsetninger (og generell motivasjon) kan naturlig nok også disse begrepene brytes ned i spesifikke karakteristika hos studenten. Med

dette mener jeg spesifikke forutsetninger (og motivasjon) for enkelte fag, mens andre ikke vekker samme interesse (Yount, 1996). Siden empirien indikerer en klar sammenheng mellom generelle forutsetninger og opplevelse av kurskvalitet, er det da også naturlig å forvente at studenter som har spesiell interesse for et fag også tenderer til å gi gode tilbakemeldinger i det eller de kursene som omhandler dette faget. Jeg har ikke funnet studier som på en eksplisitt måte studerer dette, men Seiler et al. (1999) fant at obligatoriske kurs jevnt over ble evaluert mer negativt i form av kursinnhold enn valgfrie kurs. Siden valgfrie kurs naturlig nok tiltrekker seg studenter som nettopp har interesse for disse fagene, gir derfor dette en klar indikasjon på at spesifikke interesser (og forutsetninger) påvirker studentenes evalueringer. I tråd med dette fant også McPherson (1982) at den gjennomsnittelige evalueringen var høyest i kurs hvor en stor andel av studentene hadde faget som fordypningsemne.

Faglærers forutsetninger

Med *lærerforutsetninger* mener jeg undervisningserfaring, formelt utdanningsnivå, faglig profil (inkludert eventuell pedagogisk formalkompetanse) og andre personlig egenskaper som for eksempel personlighet. McPherson (1982) fant en klar positiv sammenheng mellom antall semestre med undervisningserfaring og tilbakemelding på undervisning for lavere grads college studenter. For høyere grads studenter var det imidlertid ingen slik sammenheng. En annen studie har faktisk påvist negative (men svake) sammenhenger (Crader & Butler, 1996). Seiler et al. (1999) fant videre at amerikanske universitetsprofessorer med fast ansettelse (som er en indikator for ansiennitet) jevnt over fikk gode evalueringer på personlige egenskaper som foreleser. Den samme studien viste imidlertid en negativ sammenheng mellom fast ansettelse og opplevd kursinnhold fra studentenes side.

Oppsummert kan vi derfor si at empirien gir et noe blandet bilde av sammenhengen mellom undervisningserfaring og hvordan studentene opplever den faglige kvaliteten. Etter min mening er dette overraskende, men en mulig forklaring kan være at mindre erfarne forelesere legger seg mer i selen enn sine mer erfarne kolleger, siden det tross alt er mer som står på spill for den førstnevnte

gruppen (for eksempel kvalifisering for fast ansettelse, opprykk, dokumentasjon på «god» undervisning ved eventuell ny jobbsøknad, etc.). En annen forklaring kan faktisk være den naturlige sammenhengen mellom undervisningserfaring og alder. Studier har vist en klar negativ sammenheng mellom faglæreres alder og studentevalueringer (Arbuckle & Williams, 2003; Renaud & Murray, 1996). Ingen av bidragene jeg har kommentert ovenfor har imidlertid kontrollert for denne faktoren, og dette kan ha maskert en genuin sammenheng mellom faglæreres undervisningserfaring og de evalueringene han eller hun måtte få.

Andre studier har videre indikert at faglæreres kjønn, personlighet, rase og utseende kan påvirke hvordan man blir evaluert. I et eksperiment lot Arbuckle & Williams (2003) studentene få høre en forelesning hvor stemmen til foreleser var gjort kjønnsnøytral og aldersnøytral. Studenter som ble forespeilt at foreleser var en ung mann, anga at vedkommende både var mer entusiastisk og hadde mer meningsfull stemmebruk enn tilbakemeldingene fra studenter som ble forespeilt andre karakteristika hos foreleser (ung kvinne, gammel mann eller gammel kvinne). Andre studier har også påvist at mannlige lærere blir vurdert til å være mer entusiastiske enn sine kvinnelige kolleger (Leone-Perkins, Schnuth, & Kantner, 1999), og spesielt er det mannlige studenter som gjør negative vurderinger av kvinnelige faglærere (Basow, 1995). Med andre ord, det er ikke bare alder som spiller inn, men også kjønn påvirker hvordan studentene opplever foreleser. Radmacher et al. (2001) har videre avdekket en meget sterk positiv sammenheng mellom utadvent personlighet hos foreleser og gode studentevalueringer, noe som også finner støtte i en rekke tidligere studier (Ambady & Rosenthal, 1993; Inglehart, 1997; Murray, Rushton, & Paunonen, 1990; Waters, Kemp, & Pucci, 1988). Renaud & Murray (1996) har også påvist at flere personlighetstrekk influerer evalueringer.

Hva gjelder rasemessig bakgrunn så har jeg ikke funnet bidrag som på en eksplisitt måte har studert hvordan dette påvirker studentevalueringer. En undersøkelse av kvinnelige fysioterapistudenter viste imidlertid at etnisk bakgrunn hos studentene påvirket evalueringene av et foredrag de holdt for randomiserte utvalg av ferdigutdannede fysioterapeuter. Selv om foredragene var identiske

i innhold, fikk den afroamerikanske studenten generelt dårligst evalueringer (Haskins, RoseStPrix, & Elbaum, 1997). En annen studie har til og med påvist at opplevd fysisk attraktive forelesere (både kvinner og menn) jevnt over skåret høyere på studentevalueringer enn sine mindre attraktive kolleger (Riniolo, Johnson, Sherman, & Misso, 2006). Campbell, Gerdes, & Steiner (2005) greide imidlertid ikke å påvise en slik sammenheng.

Jeg må tilsist medgi at jeg overraskende nok ikke har funnet bidrag som eksplisitt tester hvorvidt faglæreres utdanningsnivå eller om vedkommende har formalkompetanse i pedagogikk i noen grad influerer studentenes evalueringer. Seiler et al. (1999) fant imidlertid at forskningsorienterte forelesere fikk gode tilbakemeldinger på personlige egenskaper, og dette kan indikere en positiv sammenheng mellom faglig ekspertise og opplevd kvalitet. Fast ansettelse i det amerikanske universitetssystemet er også en indikator på faglig dyktighet (i tillegg til ansiennitet), og her har jeg allerede påpekt at Seiler et al oppdaget en positiv sammenheng med personlige egenskaper hos foreleser. Effekten på kursinnhold var imidlertid negativ, men som nevnt; dette kan forklares ved at fast ansettelse gir lavere insentiver for å skape opplevd gode kurs.

Målsettinger og faginnhold

Studentenes målsettinger i høyere utdanning vil naturlig nok variere fra individ til individ, og på samme måte vil *faglæreres målsetting* også kunne avvike fra både andre faglæreres målsettinger og mange av studentenes målsettinger. Til sist kan det godt tenkes at både faglærers og studentenes målsettinger slett ikke stemmer overens med verken fagplanens (eller kursinnholdets) målsettinger. Problemstillinger som er relaterte til målsettinger i høyere utdanning er med andre ord meget komplekse, og ikke minst bør vi derfor ha et reflektert forhold til dette når vi ber studentene evaluere både faglærere og kursinnhold. Som vi skal se kan også studentevalueringer påvirke det faktiske *faginnholdet* både i form av arbeidsmengde og kurskrav.

Som jeg har påpekt tidligere blir studentevalueringer i Norge og internasjonalt nå (nærmest ukritisk) benyttet som et insentivsystem for undervisningsperso-

nalet i høyere utdanning. Et av problemene er imidlertid at det nettopp kan føre til både liberal karaktergivning, samt reduksjon i arbeidskrav (for lærere så vel som studenter). Studenter som blir forespeilt gode karakterer vil, som jeg har påpekt, jevnt over gi gode studentevalueringer. I tur kan også dette være positivt for faglæreren. Han eller hun kan derfor føle seg fristet til å være raus med karakterene, noe som kan forklare sammenhengen mellom gode evalueringer og studentenes forventninger til karakterer (Eiszler, 2002; Greenwald et al., 1997; Isely et al., 2005; Kidd et al., 2004; Krautmann et al., 1999; McPherson, 1982; Radmacher et al., 2001). Greenwald et al. (1997) fant i denne forbindelse en klar negativ(!) sammenheng mellom forventet karakter og arbeidsinnsats, noe de mener bare kan forklares ut ifra at studentene forventer gode karakterer i kurs med «snille» professorer; både i form av liberal karaktergivning og lette arbeidskrav (som studentene i tur evaluerte positivt som «betaling»). Om forventet karakter ene og alene skyldtes hardt arbeid ville naturlig nok sammenhengen mellom forventet karakter og arbeidsinnsats vært positiv og ikke negativ. Andre studier hevder også å kunne påvise en kausal sammenheng som Greenwald et al. argumenterer for (Eiszler, 2002; Isely et al., 2005).

I vår norske kontekst virker imidlertid kanskje problematikken jeg har diskutert ovenfor noe fremmed. Økt bruk av både studentevalueringer samt at mange studenter nå i større grad enn før blir evaluert underveis i kursmoduler (eksempelvis mappeevalueringer og obligatoriske innleveringer), har likevel også aktualisert debatten her hjemme. Liberal karaktergivning fører videre til færre klager og færre henvendelser om karakterbegrunnelse, noe som naturlig nok letter arbeidsbyrden til faglærer. Dette kan derfor forklare inflasjonen i gode karakterer som vi har observert de senere årene ved amerikanske læresteder, og som vi nå også ser tendenser til her til lands (spesielt i form av høyere produksjon av studiepoeng og lavere strykandel). Redusert bruk av ekstern sensor kan også spille inn her. Sistnevnte forhold innebærer at faglærer i større grad vurderer studentene alene og derved kan føle seg fristet eller presset til liberal karaktersetting. Et annet forsterkende moment som kommer inn her er at finansieringssystemet innenfor høgskolesektoren nå i større grad enn før er basert på produksjon av studiepoeng. Dette kan danne et ekstra insentiv for å gi studenter ståkarakter, til tross for marginale kunnskaper.

Midtevaluering

Jeg vil nå presentere noen analyser fra en midtevaluering jeg foretok i et kurs på bachelorutdanningen i økonomisk-administrative fag ved Høgskolen i Bergen. Studentene hadde i tiden frem til midtevalueringen hatt 12 dobbelforelesninger med tradisjonell plenumsundervisning. Siktemålet for denne undervisningsblokken hadde vært å gi en oversikt og innføring i sentrale begrep og teorier i faget. På dette tidspunktet hadde jeg også forespeilet studentene at de senere i kurset skulle jobbe med en obligatorisk gruppebasert innlevering (vurdert til bestått/ikke bestått). Til sammen 38 studenter deltok i midtevalueringen. 91 studenter har meddelt at de er deltagere på den obligatoriske innleveringen, og ut i fra dette tallet har jeg en svarandel på i underkant av 42%.

Resultater og analyser

I Tabell 1 har jeg gjengitt en del univariate analyser fra evalueringen. Spørsmål 1-5 og 12-17 berører mer eller mindre direkte hvordan studentenes opplever undertegnede som pedagog. Vi observerer at snittet for disse spørsmålene er rundt 3, noe som kan regnes for sånn omtrent midt på tre. Av standardavvikene ser vi imidlertid at det er stor variasjoner i utvalget, noe som også bekreftes av store ulikeheter i minimums- og maksimumsverdier (lave verdier angir gode evalueringer, mens høye angir dårlige). Sagt på en annen måte, det er ingen konsistens i hvordan studentene vurderer meg som faglærer. På samme måte observerer vi at det er stor variasjon i hvordan studentene svarer på de øvrige spørsmålene. For å vinne best mulig innsikt i mulige sammenhenger mellom evalueringer og andre forhold er det derfor nødvendig å foreta bivariate og multivariate analyser.

Min	Maks	Gj snitt	Std avvik		
2	5	3,08	,94	Engasjerende	1
1	5	2,92	1,02	Skaper interesse	2
2	4	2,97	,79	Stimul egentenkning	3
1	4	2,09	,67	Godt planlagt	4
1	5	3,18	1,04	Stort læringsutbytte	5
1	5	3,11	1,3	God lærebok	6
1	5	2,89	,8	Lettere enn forventet	7
2	5	3,05	,57	Mindre arbkrev enn forv	8
1	5	2,11	,95	Innlev fagl rel	9
1	5	1,97	,97	Vil lære mye av innlev	10
1	4	2,82	,56	Innlev for arbkrev	11
1	5	2,79	1,14	Muligheter for notater	12
1	5	2,08	,82	Invit til spm	13
1	5	2,37	,91	God resp på henv	14
1	5	2,13	,93	Eksempler som belyser	15
1	5	2,76	1,00	Presenterer klart	16
1	5	2,79	,93	Gir god helhetsforst	17
1	5	2,89	,76	Fornøyd med egeninnsats	18
2	5	3,42	,86	Forb meg mye til undervisn	19
1	5	3,47	1,08	Jeg er aktiv i undervisn	20
2	5	3,37	,79	Arb mye etter unvern	21
1	5	3,05	1,61	Samarb med andre	22
1	5	1,71	,98	Nesten alltid på forelesn	23

Tabell 1. Univariate analyser i fra midtevalueringen.

Det første jeg gjør er å kjøre eksploratorisk faktoranalyse hvor spørsmål 1-5 og 1217 inngår.³ Bakgrunnen for dette er å forenkle senere analyser samt avdekke eventuelle underliggende dimensjoner av hvordan studentene evaluerer undertegete. Tabell 2 gjenspeiler to hoveddimensjoner. Jeg skal ikke legge for mye i å redegjøre for disse, men den førstnevnte berører sannsynligvis hvordan studentene opplever faglig engasjement og innhold, mens den andre mer går på vurderingen av strukturelle forhold i forelesningssituasjonen. I fortsettelsen vil jeg imidlertid omtale de to dimensjonene som hhv. *opplevd kvalitet I* og *opp-*

3 Extraction method: Principal Axis Factoring. Rotation Method: Varimax with Kaiser Normalization.

levd kvalitet II. Tabell 4 viser for øvrig at korrelasjonen mellom de to begrepene naturlig nok er meget høy.

	Faktor I	Faktor II
Stort læringsutbytte	,859	
Skaper interesse for faget	,815	
Engasjerende	,694	
Stimulerer egeninteresse	,693	
Gir god helhetsforståelse	,652	
Godt planlagt	,589	
God respons på henvendelser		,841
Inviterer til spørsmål		,830
Presenterer klart		,646
Eksempler som belyser faget		,607
Muligheter for notater		,591
<i>Cronbach's</i>	,916	,867

Tabell 2. Eksplorativ faktoranalyse av spørsmål som vedrører opplevd kvalitet.

Videre foretar jeg eksplorativ faktoranalyse av de øvrige spørsmålene i evalueringsskjemaet hvor resultatene er gjengitt i Tabell 3.⁴ Vi observerer at Faktor I klart berører forhold som har med studentenes *egeninnsats* å gjøre. Faktor II sier noe om hvordan studentene opplever vanskelighetsgrad og arbeidsbyrde og er muligens indikatorer på *studentforutsetninger*. Faktor III forteller oss noe om *forventningene* til studentene. Studenter som har positive forventninger er sannsynligvis mest tilbøyelige til å være tilstede på forelesningene. På samme måte er også forventningene høye til den obligatoriske prosjektoppgaven (som de praktisk talt ikke hadde begynt på da jeg foretok midtevalueringen). Til sist registrerer vi at opplevelsen av lærebok og grad av samarbeid med andre i faget (utover den obligatoriske gruppeoppgaven) faktisk utgjør en og samme dimensjon (Faktor IV). Jeg har ikke funnet noen god forklaring på dette, men det kan hende at studenter som samarbeider om faget i større makter å skjønne stoffet i læreboka (og derved har mer positivt forhold til den) enn studenter som jobber mye på egenhånd. Jeg velger derfor å kalle denne dimensjonen *kollektiv delaktighet*.⁵

4 Samme framgangsmåte som i Tabell 2.

5 Spørsmål 11 er tatt ut siden det ikke inngår i noen av de identifiserte dimensjonene.

	Faktor I	Faktor II	Faktor III	Faktor IV
Fornøyd med egeninnsats	,837			
Forbereder meg mye til undervisning	,816			
Jeg er aktiv i undervisningen	,677			
Arbeider mye etter undervisningen	,667			
Lettere å forstå enn forventet				
Mindre arbeidskrevende enn forventet		,758		
Vil lære mye av innleveringen		,724		
Innleveringen faglig relevant				
Er nesten alltid tilstede på forelesning			,724	
God lærebok			,678	
Samarbeider med andre			,664	
				,733
				,596
<i>Cronbach's</i>	,843	,731	,749	,631

Tabell 3. Eksplorativ faktoranalyse av øvrige spørsmål

Tabell 4 gjengir korrelasjoner (sammenhenger) mellom de ulike begrepene/dimensjonene som jeg har identifisert. Vi observerer sterke og signifikante sammenhenger mellom opplevd kvalitet (både I og II) og henholdsvis egeninnsats, forutsetninger, forventninger, og kollektiv delaktighet. Jeg har også foretatt ulike multivariate regresjonsanalyser som generelt sett bekrefter det vi observerer i Tabell 4. Av plasshensyn har jeg imidlertid ikke tatt disse med.

	OK1	OK2	Egen	Forutsetn	Forv
Opplevd kvalitet 1 (OK1)					
Opplevd kvalitet 2 (OK2)	,731				
Egeninnsats	,427	,357			
Forutsetninger	,624	,570	,342		
Forventninger	,524	,470	,228	,342	
Kollektiv delaktighet	,393	,441	,088	,367	,233

Tabell 4. Korrelasjonsanalyse (N=38). Aller korrelasjoner er signifikante ($p < ,05$), unntatt verdier gjengitt i kursiv.

Oppsummert kan vi derfor med rimelig sikkerhet anta at egenskaper hos studentene i stor grad influerer deres opplevelse av både forelesningssituasjonen og det faglig opplegget for øvrig. Siden dette er tverrsnittsdata bør vi imid-

lertid være forsiktige med å trekke konklusjoner om kausal retning. Det kan for eksempel hende at studenter som føler at undervisningen er interessant i tur blir inspirert og derfor jobber mer med faget, enten alene eller i samarbeid med andre. Til sist kan det også tenkes at de ulike identifiserte begrepene har sitt utspring i en og samme underliggende dimensjon, for eksempel en generell holdning/innstilling til det å i det hele tatt studere.

Diskusjon og refleksjoner

Analysene ovenfor har på ingen måte tatt mål av seg å ha et vitenskaplig formål (og dette var heller ikke bakgrunnen for å utarbeide spørreskjemaet samt innhente dataene). Snarere har motivasjonen vært å gjøre en helt enkel kartlegging av mulige faktorer som kan influere studentevalueringene i det aktuelle kurset. Like fullt mener jeg at resultatene er i tråd med internasjonal forskning som jeg har referert til tidligere. I Tabell 4 ser vi klare positive sammenhenger mellom det jeg henholdsvis har valgt å omtale som egeninnsats, forutsetninger, forventninger, kollektiv delaktighet og opplevd kvalitet på undervisningen. I det store og det hele ser vi at studenter som enten jobber med faget, er tilstede på forelesningene og på annen måte uttrykker at de både er positive til å lese og lære faget, også tenderer å ha et godt inntrykk av underteignede som faglærer.⁶ Analysene viser med andre ord at det slett ikke er noe konsensus hos studentgruppen i forhold til hvordan de opplever underteignede som faglærere. Sagt noe forenklet; dyktige og motiverte studenter er jevnt over mer fornøyde enn mindre dyktige og motiverte studenter.

Som pedagog i høyere utdanning mener jeg at en av mine sentrale oppgaver er å gi studentene en sentral innføring i og anvendelse av fagterminologien. Videre er jeg også ansvarlig for å gi faglig veiledning og tilbakemelding på studentarbeider. Jeg kan selvfølgelig stille en rekke provoserende spørsmål på forelesningene, men å vinne innsikt i selve faget samt skaffe seg et reflektert og kritisk forhold til sentrale antagelser, påhviler imidlertid den enkelte student.

⁶ Vi kan videre selvfølgelig også stille spørsmål ved kausaliteten i datamaterialet som jeg har analysert. Det kan tenkes at studenter som opplever underteignede som en dyktig faglærer av den grunn blir bedre motivert og samtidig får bedre fagforståelse. Studier jeg har referert til tidligere viser imidlertid at faktorer som poengsum ved opptak til college påvirker studentevalueringer på en positiv måte, og dette indikerer at den kausale retningen går fra forutsetninger til opplevd kvalitet (men kausaliteten kan selvfølgelig også gå i begge retninger). Om jeg i mitt eget materiale ønsket å få bedre rede på kausale forhold, kunne jeg for eksempel be studentene informere om opptakssum til studiet eller gjennomsnittskaraktter så langt. Hva gjelder motivasjon kunne jeg samlet inn data om dette innledningsvis i semesteret (før undervisningen var kommet i gang) og sammenlignet med resultater fra midtevalueringen (dette ville imidlertid forutsette at jeg var i stand til å identifisere den samme studenten på to ulike tidspunkter).

Om jeg i jakten på gode studentevalueringer skulle bli fristet til å senke arbeidskravene i et kurs – for eksempel ved å korte ned på pensum eller bytte ut en engelskspråklig lærebok – ville dette sannsynligvis blidgjøre en del studenter. Tidligere har jeg også referert til forskning som nettopp viser at «slepphendte» universitetsprofessorer tenderer til å få gode evalueringer. Det kan også godt hende at de svakeste og minst motiverte studentene på denne måten ville lære mer siden de ikke ville falle fullstendig gjennom pga. krevende kurskrav. Men på den annen side: Hva skal premissene være for høyere utdanning? Er målet at flest mulig studenter skal være fornøyde, at flest mulig (gjennom nitidig undervisning og moderate arbeidskrav) tilegner seg minimumskravene? Eller ønsker vi at stramme kurskrav skal fungere som en naturlig seleksjonsmekanisme hvor motiverte kandidater med tilfredsstillende kvalifikasjoner slipper gjennom? Økende bruk av både studentevalueringer, studenttilfredshetsundersøkelser, samt at studiestedene får betaling for produksjon av studiepoeng, har i høy grad vært med på å aktualisere disse spørsmålene. En student som både strever med motivasjonen og fagtilegnelse vil naturlig nok ønske at fagnivå og fagformidling er av en slik art at han er i stand til å gjennomføre studiet. Men om dette kan forsvares ut i fra arbeidslivets behov for kompetent arbeidskraft finner jeg imidlertid grunn til å stille spørsmål ved. Å beskjære kurskrav og kursinnhold vil videre berøve dyktige og motiverte studenter til å bevege seg mot optimal læring og fagforståelse.

Jeg vil selvsagt være forsiktig med å kategorisk hevde at utstrakt bruk av studentevalueringer (og også andre studenttilfredsundersøkelser) vil føre til utglidning av både fagkrav og integritet i høyere utdanning. Like fullt mener jeg at denne redegjørelsen inneholder gode argumenter for at en slik tendens kan forekomme, og forskningsbidrag som jeg tidligere har referert til indikerer også at slike tendenser faktisk finner sted.

Avsluttende kommentarer

Jeg har i denne oppgaven rettet søkelyset på hvordan formaliserte studentevalueringer i høyere utdanning kan og bør brukes, og i forlengelse av dette har jeg også stilt spørsmål ved nytteverdien av den informasjonen som fremkom-

mer. Oppsummert viser internasjonal forskning, til dels støttet av egne analyser, at det finnes en rekke forhold utenfor faglærers didaktiske kvalifikasjoner som kan influere de tilbakemeldingene som studentene gir. Tilbakemeldinger som jeg og andre faglærere mottar fra studenter bør vi derfor innta et meget edruelig forhold til. Spesielt bør det anspore en viss kritisk holdning når vi vet at faktorer som kjønn, alder, rase og personlighet kan være med på å påvirke evalueringene. Videre har også studentforutsetninger som motivasjon og kvalifikasjoner stor betydning for tilbakemeldingene de gir.

Når dette er sagt betyr det imidlertid ikke at studentevalueringer er verdiløse. Poenget er at vi inntar et reflektert forhold til det innsamlede materialet. I denne oppgaven har jeg derfor belyst hvordan tidligere forskning, samt en viss systematikk i analyser av egne evalueringer, kan hjelpe oss på vei i så måte. Jeg vil på ingen måte hevde av vi skal forholde oss til forskningsresultater på en ukritisk måte (hvor også resultatene er til dels sprikende), men det gir oss uansett et faglig fundament til økt forståelse av den informasjonen vi har tilveiebrakt. Ignorering av kunnskap vi faktisk har tilgang til (om den enn er aldri så ufullkommen) er absolutt det verste. Til sist bør vi ha et kritisk og reflektert forhold til de mekanismer økt bruk av studentevalueringer kan ha både i forhold til kurskrav og faglig integritet i høyere utdanning. Her ligger imidlertid ikke bare ansvaret hos den enkelte faglærer, men også i særdeleshet hos faglig ledelse ved norske høgskoler og universiteter.

Litteratur

- Ambady, N., & Rosenthal, R. 1993. Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology*, 64: 431-441.
- Arbuckle, J., & Williams, B. D. 2003. Students' perceptions of expressiveness: Age and gender effects on teacher evaluations. *Sex Roles*, 49(9-10): 507-516.
- Basow, S. A. 1995. Student evaluations of college professors: When gender matters. *Journal of Educational Psychology*, 87(4).
- Campbell, H. E., Gerdes, K., & Steiner, S. 2005. What's looks got to do with it? Instructor appearance and student evaluations of teaching. *Journal of Policy Analysis and Management*, 24(3): 611-620.
- Crader, K. W., & Butler, J. K. 1996. Validity of students' teaching evaluation scores. The Wimberly-Faulkner-Moxley questionnaire. *Educational and Psychological Measurement*, 56(2): 304-314.
- Eiszler, C. F. 2002. College students' evaluation of teaching and grade inflation. *Research in Higher Education*, 43(4): 483-501.
- Greenwald, A. G., & Gillmore, G. M. 1997. Grading leniency is a removable contaminant of student ratings. *American Psychologist*, 52(11): 1209-1217
- Hand, L., & Rowe, M. 2001. Evaluation of student feedback. *Accounting Education*, 10(2): 147-160.
- Haskins, A. R., RoseStPrix, C., & Elbaum, L. 1997. Convert bias in evaluation of physical therapist students' clinical performance. *Physical Therapy*, 77(2): 155-163.
- Hiim, H., & Hippe, E. 1998. *Læring gjennom opplevelse, forståelse og handling: En studiebok i didaktikk*. Oslo: Universitetsforlaget.
- Inglehart, R. 1997. *Modernization and post-modernization: Cultural, economic, and political change in 43 societies*. Princeton, NJ: Princeton University Press.
- Isely, P., & Singh, H. 2005. Do higher grades lead to favorable student evaluations? *Journal of Economic Education*, 36(1): 29-42.
- Kelly, H. H. 1973. The processes of causal attribution. *American Psychologist*, 28: 107-128.
- Kidd, R. S., & Latif, D. A. 2004. Student evaluations: Are they valid measures of course effectiveness? *American Journal of Pharmaceutical Education*, 68(3): Art. No. 61.
- Krautmann, A. C., & Sander, W. 1999. Grades and student evaluation of teachers. *Economics of Education Review*, 18(1): 59-63.

- Kravitz, D. A., Stone-Romero, E. F., & Ryer, J. A. 1997. Student evaluations of grade appeal procedures: The importance of procedural justice. *Research in Higher Education*, 38(6): 699-726.
- Leone-Perkins, M., Schnuth, R., & Kantner, T. 1999. Preceptor-student interactions in an ambulatory clerkship: Gender differences in student evaluations of teaching. *Teaching and Learning in Medicine*, 11(3): 164-167.
- McPherson, J. M. 1982. Hypernetwork sampling: duality and differentiation among voluntary organizations. *Social Networks*, 3: 225-249.
- Moore, S., Eng, E., & Daniel, M. 2003. International NGOs and the role of network centrality in humanitarian aid operations: A case study of coordination during the 2000 Mozambique floods. *Disasters*, 27: 305.
- Murray, H. G., Rushton, J. P., & Paunonen, S. V. 1990. Teacher personality traits and student instructional ratings in six types of university courses. *Journal of Educational Psychology*, 82: 250-261.
- Radmacher, S. A., & Martin, D. J. 2001. Identifying significant predictors of student evaluations of faculty through hierarchical regression analysis. *The Journal of Psychology*, 135(3): 259-268.
- Renaud, R. D., & Murray, H. G. 1996. Aging, personality, and teaching effectiveness in academic psychologists. *Research in Higher Education*, 37(3): 323-340.
- Riniolo, T. C., Johnson, K. C., Sherman, T. R., & Misso, J. A. 2006. Hot or not: Do professors perceived as physically attractive receive higher student evaluations? *The Journal of General Psychology*, 133(1): 19-35.
- Seiler, M. J., Seiler, V. L., & Chiang, D. 1999. Professor, student, and course attributes that contribute to successful teaching evaluations. *Financial Practice and Education*, Fall/Winter
- Waters, M., Kemp, E., & Pucci, A. 1988. High and low faculty evaluations. *Teaching of Psychology*, 15: 203-204.
- Yount, W. R. 1996. *Created to Learn*. Nashville, Tennessee: Broadman & Holman Publishers.