

Å lære å gi og motta konstruktiv kritikk gjennom medstudentvurdering

Jill Walker

Seksjon for humanistisk informatikk, Det historisk-filosofiske fakultet

E-post: jill.walker@uib.no

”Det er likevel bare eksamenskommisjonen som leser det,” pleide vi å si da jeg var student. Visst hadde vi skrivekurs og kollokviegrupper, men i praksis så vi ikke mange andre studenters oppgaver. Da jeg for første gang var sensor på semesteroppgaver, var jeg forbauset over hvor mye jeg lærte om oppgaveskriving gjennom å lese meg gjennom og vurdere bunken med eksamensbesvarelser. Det var med denne intense egenlæringen i tankene at jeg i høst valgte å la studentene mine vurdere hverandres oppgaver som en del av læringsprosessen deres.

Når studenter vurderer hverandres oppgaver, kalles det ”peer assessment”, eller på norsk, medstudentvurdering. Dette er en uvanlig vurderingsform i Norge, men har internasjonalt vært i bruk i et par tiår. Kvalitetsreformen åpner for nye vurderingsformer, men Stortingsrapport 27 (2001) nevner ikke medstudent-vurdering eksplisitt. Derimot nevnes medstudentvurdering som en av de nye formene vurdering som er blitt aktuell i nyere pedagogikk. Olga Dysthe og Knut Steinar Engelsen (Dysthe and Engelsen: 2003) setter opp følgende oversikt over internasjonale tendenser i vurdering:

Frå		I retning av
Skriftleg slutteksamen	→	Alternative vurderingsformer, t.d. mapper
Lærer- og sensorvurdering	→	Sjølvvurdering, medstudentvurdering
Individuell eksamen	→	Gruppearbeid og prosjektarbeid
Implisitte kriterium	→	Eksplisitte kriterium
Konkurransse	→	Samarbeid
Lærarstyrt eksamen	→	Studentinnflytelse over vurdering
Produktvurdering	→	Prosessvurdering

Medstudentvurdering nevnes her, og det samme gjør samarbeid og student-innflytelse over vurdering. I en overgangsfase som den vi er i nå, er studentene fortsatt vant til vurderingsformene i den første kolonnen, og det er ikke gjort i en håndvending å få nye vurderingsformer til å fungere.

“Men hvordan skal vi vurdere hverandre når vi aldri har gjort det før?”, spurte studentene. “Vi kommer bare til å såre hverandre!” De hadde mange innvendinger, og vi diskuterte lenge for å komme fram til en modell som vi alle var enige i var rettferdig. Mange av studentene var i utgangspunktet svært skeptiske til medstudentvurderingen, men innstillingen deres endret seg radikalt fra før til etter den ble gjennomført. Den jevnt tvilende holdningen snudde seg til en mye mer positiv innstilling i etterkant av medstudentvurderingen.

I denne artikkelen vil jeg i hovedsak presentere opplegget vi brukte for å gjennomføre medstudentvurdering, og diskutere dette i forhold til tilbake-meldinger fra studentene om hvordan de opplevde prosessen. Siden medstudent-vurdering er så sjeldent i Norge, er det verdifullt å se hvordan norske studenter reagerer på en slik prosess, og å diskutere gode og dårlige erfaringer jeg og studentene har gjort i denne prosessen. Jeg innleder til dette med en beskrivelse av studentgruppen og opplegget jeg brukte samt en rask oversikt over tidligere forskning på medstudentvurdering.

Om studentgruppen, opplegget og materialet

Medstudentvurderingen ble gjennomført i en gruppe på tolv studenter hvor tre var laveregradsstudenter og ni var Mastergradsstudenter. Alle var tilknyttet Seksjon for humanistisk informatikk ved Universitetet i Bergen. Emnene HUIN204: Digitale medier og kultur og HUIN303: Digital medieestetikk har felles undervisning og er tilrettelagt for erfarne studenter. Emnene skal gi studentene en oversikt over praksis og teori innen digital estetikk og kultur, og fokus i høst var på nettkunst, interaktive fortellinger og dataspill, samt på historien til feltet. Studentene velger selv tema for en semesteroppgave som tidligere har utgjort hele vurderingsgrunnlaget, og som har vært karactersatt på vanlig måte av emne-ansvarlig og en ekstern sensor.

Erfaringer fra tidligere semestre hadde vist at selv om seminarene var vellykkede, med engasjerte studenter og mye læring, var det få studenter som leverte semesteroppgaven. For å oppmuntre til at flere studenter gjennomførte emnet, la vi høsten 2003 om vurderingsformen for emnet og krevde at studentene leverte et tellende utkast til semesteroppgaven sin som skulle karactersettes av medstudenter. Dette utkastet skulle karactersettes ved medstudentvurdering. Ved å kreve innlevering av et utkast håpet vi å sikre at studentene kom godt i gang med skrivingen tidlig i semesteret, og ved å bruke medstudentvurdering som vurderingsmetode ønsket vi å gjøre studentene mer bevisste på forventninger og prosess samtidig som vi søkte å fostre et åpent klima mellom studentene hvor gjensidig tilbakemelding og støtte studentene mellom var naturlig.

Medstudentvurderingen ble altså brukt på et utkast til semesteroppgavene som studentene skulle levere på slutten av semesteret, og karakteren som ble gitt telte som 40% av sluttkarakteren. Siden utkastene skulle vurderes, ga jeg studentene svært spesifikke krav til hva som skulle med. De skulle bl.a. ha en klart definert problemstilling, et *abstract* som beskrev det tenkte innholdet i semesteroppgaven, en disposisjon, en tidsplan for arbeidet mot sluttproduktet og en litteraturliste for litteratur hvor hver kilde var annotert med en setning som forklarte hvordan den var tenkt brukt i oppgaven. Studentene fikk også eksplisitte kriterier for karactersettingen utdelt i god tid før de leverte utkastene sine. Her er for eksempel kriteriene for å få karakteren A på utkastet sitt:

A: Fremragende. Fremragende prestasjon som klart utmerker seg. Viser stor grad av selvstendighet.

Tydelig og selvstendig eller kreativ problemstilling, litteraturlisten er korrekt satt opp og oppfylder kravene, minst en kilde utover det som er vist i timene, det er tydelig hvorfor akkurat disse kildene er valgt, gjerne en viss kreativitet i valget av kilder. Arbeidsplanen er i samsvar med ambisjonene og det virker sannsynlig at den kan gjennomføres. Svært få eller ingen slurvefeil, omtrent riktig antall ord for hvert element. **For analytiske oppgaver:** beskrivelsen av analyseobjektet er klart og godt skrevet, slik at leseren kan danne et godt bilde av det. Disposisjonen er godt organisert. Det er god sammenheng mellom og gjerne noe originalitet i valg av problemstilling, kilder og analyseobjekt. **For praktiske oppgaver:** Demoen og beskrivelsen er godt utarbeidet og gir et tydelig inntrykk av hva som tenkes laget.

For 300-nivå: Selvstendighet vektlegges alltid for å få en A: det betyr at man ikke legger opp til en plankeoppgave men har en genuint original idé. Dette er viktigere på 300-nivå enn på 200-nivå. Valg av en problemstilling og kilder som viser egen utforskning utover pensum og forelesninger forventes på dette nivået, mens en oppgave på 200-nivå kan få A for selvstendig forståelse og bruk av det som er gjennomgått. Selvsagt teller selvstendighet og kreativitet også her.

Før og etter medstudentvurderingen hadde vi flere samtaler i gruppen om prosessen, og jeg har brukt noen av uttalelsene fra disse samtaleene i denne artikkelen. Jeg diskuterer også noen av mine egne erfaringer og observasjoner fra medstudentvurderingen. I tillegg gjennomførte vi en skriftlig evaluering uken etter medstudentvurderingen. Skjemaet studentene fylte ut er gjengitt i **Figur 2**. Syv studenter deltok i denne evalueringen, og det er i hovedsak disse resultatene jeg bygger konklusjonene mine på.

Med kun tolv studenter, hvorav bare syv leverte en skriftlig evaluering, blir utvalget for lite til å generalisere, men det kan være verdifullt som eksempel og for å antyde typiske bekymringer studenter har for et slikt opplegg.

Forskning rundt medstudentvurdering

De siste par tiårene har medstudentvurdering, eller ”peer assessment”, vært utprøvd på mange universiteter og på mange forskjellige måter (Mowl and Pain: 1995; Sivan: 2000; Topping et al.: 2000; van den Berg, Admiraal and Pilot: 2003). Det er mange valg og muligheter. Skal studentene sette karakterer på hverandres arbeid eller kun gi kvalitativ tilbakemelding? Skal vurderingen skje anonymt eller åpent? Skal man la flere studenter samarbeide om karaktersetting? Skal faglærer godkjenne karakteren eller skal faglærers vurdering få en bestemt vektning i forhold til studentens vurderinger? Dersom faglærers vurdering skal telle, skal studentene få vite faglærers vurdering før eller etter

de selv har kommet til en konsensus om hvilken karakter som bør settes? Man kan også velge å droppe karakterene helt bare la studenter gi hverandre uforpliktende tilbakemeldinger på hverandres arbeider.

Medstudentvurderinger har spesielt vært brukt som en *formativ* vurderingstype. Her skiller man mellom summativ vurdering, hvor en karakter settes etter semesterets arbeid er over, og formativ vurdering, hvor vurderingen integreres i læringsprosessen. Den summative vurderingen gjøres etter læringsprosessen, og gir et uttrykk for hvorvidt studenten har lykket eller feilet, mens den formative vurderingen søker å hjelpe studenten å lykkes mens hun er i læringsprosessen (Topping et al.: 2000). Ved å gi en vurdering underveis i semesteret ønsket jeg å skifte fokus til en formativ vurderingsform hvor studentene kunne bygge videre på læringserfaringene fra underveisvurderingen. I tillegg håpet jeg at studentene ville dra lærdom av å vurdere og diskutere andres arbeider:

To promote learning, assessments must incorporate genuine feedback that learners can employ in redirecting their efforts. In other words, assessment information must *reveal* to learners an understanding of how their work compares to a standard, the consequences of remaining at their current level of skill or knowledge, as well as information about how to improve, if improvement is needed. (Huba and Freed: 2002)

Dette var tydeligvis vellykket: evalueringene fra studentene ga uttrykk for at tilbakemeldingen de fikk fra medstudentene var svært nyttig og grundigere enn vanlig, og at de var glade for å ha kommet så godt i gang med arbeidet i god tid før den endelige semesteroppgaven skulle leveres. Men det var også negative responser til medstudentvurderingen, som jeg vil diskutere seinere i denne artikkelen. Spørsmålet for fremtidige forsøk med medstudentvurdering blir hvordan man kan utnytte de positive effektene av en medstudentvurdering samtidig som man minsker de negative effektene.

Det er lite arbeid gjort på medstudentvurdering i Norge. Sykepleie-utdanningen ved Høgskolen i Oslo er et av de få stedene som har prøvd ut vurderingsformen, og da er det gjort som en del av en større reform hvor hovedvekten ligger på mappevurderinger og sammenheng i læringsprosessen gjennom hele studiet. En rapport som oppsummerer erfaringene hittil diskuterer i korte trekk hvordan medstudentvurdering har blitt brukt som tillegg til vurdering av faglærer (Johnson: 2003). Der noteres det at studentene ikke synes å se stor nytte i medstudentvurderingen før et stykke ut i studiet. Sisteårsstudentene, som har opplevd jevnlig

medstudentvurdering under hele studiet, er mye mer fornøyde med denne vurderingsformen enn hva førsteårsstudentene er. Dette tyder på at undersøkelser må gjøres over tid ettersom studentene venner seg til vurderingsformen.

Metode for konstruktiv kritikk

Studentene ga uttrykk for at de hadde liten trening med å vurdere eget eller andres arbeid. De var bekymret for det sosiale ansvaret ved å muligens gi medstudenter dårlige karakterer, men de var også i tvil om de var kvalifiserte til å vurdere andres arbeid. Dette kom til uttrykk i de første diskusjonene vi hadde om medstudentvurderingen:

”Men hvordan skal vi vurdere hverandre når vi aldri har gjort det før?”
(Student i muntlig diskusjon før medstudentvurderingen)

Det ble også tatt opp i de skriftlige evalueringene i etterkant av medstudentvurderingen:

Skriv noe du synes ikke var bra:

”At vi må vurdere medstudenter, på tross av at vi ikke har tyngde nok til det. Vi vet f.eks. ikke hva som er en gjennomsnittlig oppgave.” (Student 5)

For å gi studentene noe erfaring med å gi hverandre tilbakemeldinger, gjennomførte jeg en styrt tilbakemeldingsrunde på et første idéutkast til semesteroppgaven. Da studentene seinere ga hverandre karakterer, brukte de den samme tilbakemeldingsstrukturen.

Strukturen jeg valgte var basert på Edward de Bonos tenkehatter (de Bono: 1999). De Bono har arbeidet med det han kaller parallell tenkning i flere tiår, og hans teknikker egner seg spesielt godt for kollaborativt tankearbeid og for å oppnå gode resultater når grupper skal arbeide sammen.

De Bono mener at deltagere i diskusjoner ofte blir fastlåste i bestemte roller. Dersom du gir uttrykk for en mening i starten av diskusjonen, forventes det at du skal forsvare denne meningen i fortsettelse, og du ”vinner” diskusjonen ved å argumentere ut fra dette ene perspektivet. Diskusjonen blir konfliktorientert heller enn konstruktiv. Dersom man i stedet for å la hver enkelt argumentere ut fra ett perspektiv kan la alle tenke *parallelt*, mener de Bono at man oppnår bedre resultater raskere. Parallell tenkning vil si at alle i en gruppe tenker og argumenter ut fra det samme perspektivet på

samme tid. Så bytter alle til et annet perspektiv. Dette frigjør individer fra å bli fastlåst i bestemte posisjoner eller typer.

Tenkehattene er en konkretisering av denne grunntanken. De Bono har gitt seks forskjellige farger til seks forskjellige perspektiver til et tema. Idéen er at man tar på seg forskjellige hatter etter tur for å tydeliggjøre hvilket perspektiv man snakker ut fra. Det er selvfølgelig ikke nødvendig å bruke virkelige hatter, men det å dele innfallsvinklene opp så eksplisitt kan være verdifullt for å tillate tenkemåter man kanskje ellers ville ha sensurert. Når man innleder en negativ bemerkning med ordene ”Nå har jeg den svarte hatten på meg” så slipper man å identifisere hele seg med de negative ordene. Man kan lettere tillate seg å også se andre sider ved idéen.

Gruppediskusjon etter Edward de Bonos hattemetode

Først presenterer den som ber om tilbakemelding sin idéskisse.

Så tar gruppen en runde for hver hattefarge. Snakk etter tur uten avbrudd, så får den som har presentert sin idé igjen ordet, så tar dere eventuell diskusjon på dette perspektivet. Gå så videre til neste hattefarge og neste perspektiv.

Hvit hatt: sakens fakta

Fortell hva du har oppfattet at taleren ønsker å gjøre. ”Slik jeg forstår deg, planlegger du å... og du vil finne ut...”

Rød hatt: intuisjon

Hva er **magefølelsen** din? Har du noen intuisjoner som du ikke helt vet hvorfor du mener? Du er kanskje ikke sikker på om de er riktige?

Sort: djevlels advokat

Vær skeptisk. Si noe som kan bli **problematisk** i dette prosjektet? Hva kan gå galt?

Gul: optimist

Hva **liker** du? Hvorfor liker du det? Hva kan dette bidra med?

Grønn: muligheter

Hva **mer** kan gjøres, hva kunne gjøres annerledes, hva kunne dette føre til?

Hvit igjen:

Kjenner du til **fakta eller informasjon** som kan være til hjelp for personen? Kanskje du kommer på andre referanser han eller hun bør bruke? Kan du bidra med noe?

Blå: oppsummering

Den som presenterte prosjektet sitt oppsummerer hva hun eller han vil ta videre fra diskusjonen.

Figur 1: Disse instruksjonene ble delt ut til studentene både på treningsrunden og på selve medstudentvurderingen.

Jeg delte ut et ark til hver student med instruksjoner for hvordan man skulle gå gjennom de forskjellige perspektivene (se figur 1). Jeg ga hver farge eller tenkemåte en kort og konkret beskrivelse tilpasset oppgaven foran oss. Jeg delte studentene inn i grupper på fire, og hver gruppe fikk et sett med fargede hatter og gikk i gang med tilbakemeldingsprosessen.

Etter tilbakemeldingene diskuterte vi muntlig hvordan prosessen hadde fungert. Alle var enige om at det hadde vært svært produktivt, også de som i utgangspunktet hadde vært skeptiske til det. En student, en ung kvinne, ga uttrykk for at tenkehatten frigjorde henne fra hennes vante rolle som støttende og positiv: "Når vi kom til den svarte hatten måtte vi jo finne noe negativt å si - det pleier jeg aldri å gjøre! Det var deilig!" Andre studenter ga også uttrykk for at de hadde brukt andre sider av seg enn vanlig. En student opplevde det tvert imot som "overpedagogisk, og sa: "Jeg tenker sånn likevel. Tror ikke vi er så konfliktorienterte. Overflødig."

Studentene brukte tenkehattstrategien flere ganger seinere i semesteret, uten at jeg tok initiativ til dette. For eksempel kunne det hende at en student presiserte at "Nå snakker jeg med en sånn rød intuisjonshatt, altså", mens en annen gjorde en negativ kommentar mindre personlig ved å utheve at "dette sier jeg med sort hatt".

Sosialt ubehag

Jeg hadde ikke tatt høyde for at studentene ville oppleve medstudentvurderingen som sosialt ubehagelig. Det var først i diskusjonen like før vurderingen skulle foregå at jeg forsto at dette var et stort og viktig spørsmål for dem. Da hadde alle studentene lest to medstudenter sine utkast, og de hadde notert kommentarer og anbefalt karakter, og vi var samlet for å diskutere oss fram til de endelige karakterene alle skulle få. Før vi begynte på vurderingsdiskusjonene, brukte vi lang tid på å diskutere studentenes innvendinger. Noen gikk, som tidligere nevnt, på at studentene ikke stolte på sin egen evne til å vurdere objektivt og rettferdig, men det som ble tydelig var at studentene først og fremst fryktet at det å skulle vurdere sine medstudenter kunne skape konflikter og misstemninger i gruppen. Dette kommer tydelig fram i den skriftlige evalueringen studentene gjorde i etterkant av medstudentvurderingen:

Skriv noe du synes ikke var bra:

- ” - Ansvar for andres karakter.
- Dårlig samvittighet overfor medstudenter.
- Sårbart for psykososiale stressorer. Gruppen kan lide, konflikter kommer til overflaten.” (Student 4)

”Litt skremmende å være herre over andres fremtid.” (Student 5)

”Noen kunne kanskje føle det som vanskelig hvis de trodde medstudentene ikke ble fornøyd med vurderingen.” (Student 6)

”Vanskelig situasjon å være ansvarlig for 40% av en karakter hvis oppgaven du vurderer ikke holder mål.” (Student 7)

”Vi lå alle på et såpass høyt nivå at det ikke egentlig var noe problem å vurdere hverandre. Jeg tror problemene ville blitt større hvis en/noen lå an til å få en mye dårligere karakter. Jeg ville ikke likt å være den som skulle ”dra noen ned”. (Student 6)

Det er interessant at studentene i liten grad bekymrer seg for hvordan det ville ha vært å motta en dårlig vurdering fra en medstudent. I stedet er de opptatt av hvor vanskelig det ville være å gi en dårlig karakter. Noen bekymrer seg for at *andre* kan oppleve det som ubehagelig å motta en dårlig karakter fra en medstudent.

Noe av grunnen til dette er selvfølgelig at studentene har fylt ut evalueringsskjemaet etter at de vet hvilke karakterer de mottok, og karakterene var jevnt over svært gode: ingen fikk dårligere karakter enn B.

Manglende objektivitet når man skal sette karakterer på mennesker man kjenner fra studiehverdagen var også noe studentene så som en fare ved medstudentvurderingen:

”Personlige meninger om medstudenter kan påvirke hvilken karakter de ender opp med. Dette kan ha en negativ og ”urettferdig” virkning. (..) Noen vil kanskje synes at det er en urettferdig evaluering. Kritikken fra medstudenter kan være ”vondere” enn fra foreleser.” (Student 1)

”Vanskelig å gi dårlig karakter, usikker på eget faglig grunnlag.” (Student 7)

”Vanskelig å være objektiv.” (Student 3)

En student oppgir å ikke stole på sine medstudenters vurderingsevne:

”Jeg var utrolig skeptisk til at andre studenter skulle vurdere min innsats.” (Student 7)

Den samme studenten rapporterer at studentene ikke ønsket å gi dårlige karakterer:

”Jeg føler ikke at karakterskalaen ble brukt etter intensjonen. Som en annen student sa: ’Jeg gir ikke dårligere karakter enn B.’ For å få en reell vurdering av sin egen innsats burde nok det være flere elever (større gruppe gjør at man lettere uttaler seg kritisk).” (Student 7)

I denne gruppen gjorde alle studentene et svært godt arbeid, og de fortjente alle å få gode karakterer. Jeg håper at studentene ville ha vært ærlige dersom de hadde vurdert arbeider til dårligere karakter enn B, men det er umulig å vite hva som ville ha skjedd dersom det hadde vært virkelig dårlige oppgaver i gruppen.

Noe som overrasket meg var oppdagelsen av hvor sterkt studentene ønsket at jeg skulle beholde min tradisjonelle rolle som autoritet. Jeg tror at det er en sammenheng mellom studentenes mistro til sin egen evne til å vurdere og deres redsel for at det å vurdere sine medstudenter skulle skape sosiale konflikter. Dersom læreren tar hele ansvaret for vurdering, kan studentene lettere være en gruppe. Deres gruppetilhørighet er delvis definert ved at de *ikke* er lærere og at de *ikke* har ansvar for gruppen som helhet eller for vurdering av gruppen. Hvordan skal man som student kunne klage på en karakter, for eksempel, dersom det er ens likemenn som har bestemt karakteren?

Når studenter tvinges til å ta ansvar for gruppen, frykter de at de skal miste rollen som likeverdige medlemmer av gruppen. De har ingen modell for en gruppe som er selvledet, i hvert fall ikke innenfor utdanningsinstitusjonen. I sluttevalueringen for hele emnet skrev en av studentene, kun halvt ironisk, ”Ikke legg vår egen skjebne i våre egne hender.” Det er mye enklere å frasi seg ansvaret.

På den andre siden kan man godt si at det er urealistisk å forvente at studenter skal takle åpen medstudentvurdering når vi selv som profesjonelle forskere stort sett kun må tåle anonym *peer review*. Ofte er *peer review* til og med organisert som en *double blind review*, slik at verken den som vurderer eller den som vurderes kjenner den andres identitet.

Det ville sannsynligvis ha gitt mindre sosialt ubehag å hatt ”blind” eller anonym medstudentvurdering. Et vellykket forsøk har vært gjort hvor geografistudenter anonymt vurderte hverandres essayer med kvalitativ skriftlig tilbakemelding i tillegg til karaktersetting. Her telte to studentvurderinger tilsammen 50% av karakteren, mens en lærervurdering telte 50%. Studentene diskuterte så de skriftlige vurderingene fra studenter og lærer i smågrupper, fortsatt uten at de som hadde gitt vurderingen ble

avslørt (Mowl and Pain: 1995). Dette krever kanskje en større studentgruppe, men ville antageligvis ha lettet på mye av det sosiale ubehaget studentene følte.

Utdanningssystemet er i dag under endring fra barneskolen og oppover. I dagens barneskole jobber for eksempel andreklassinger med jevnlig selvevaluering. Når disse barne kommer til universitetet vil kanskje deres forventninger til hvordan en studentgruppe skal fungere være annerledes.

Skal man fjerne karaktersettingen?

Studentene foreslår ikke selv at vurderingen bør foregå anonymt. Hovedendringen de foreslår er at man beholder den kvalitative tilbakemeldingen men enten fjerner karaktersettingen eller lar den telle mindre. Det sosiale ubehaget de føler ved å sette karakterer er sterkt. Det er godt mulig at studentene ikke ville oppgitt å være så fornøyde med medstudentvurderingen i ettertid dersom noen av dem hadde måtte tildele andre dårlige karakterer – vi hadde en liten gruppe med dyktige studenter som gjorde jobben sin, og det kan man ikke forvente i alle undervisningssituasjoner.

Det ville være enkelt å la studentene slippe å tildele karakterer. I praksis ble det i dette forsøket slik at vi delte på byrden. Dette var ikke intensjonen i utgangspunktet, men studentene ble først trygge på at de ville gjennomføre opplegget da jeg foreslo at jeg kunne godkjenne karakteren de satte, og slik ha det siste ordet. Jeg hadde selvfølgelig lest og kommentert utkastene samtidig som studentene gjorde det, og jeg hadde også gjort meg opp en mening om hvilke karakterer de forskjellige oppgavene fortjente, selv om jeg i utgangspunktet hadde tenkt å holde min mening hemmelig med mindre det var svært store forskjeller i min og studentenes vurderinger. Ved å gå med på å ha de siste ordet om karakterene tok jeg også ansvaret for karaktersettingen, et ansvar som studentene følte som svært sosialt ubehagelig. De to studentene som skulle vurdere den tredjes oppgave ble altså enige seg i mellom om hvilken karakter de mente var riktig, så snakket de med meg. I alle tilfeller unntatt ett, som jeg kommer tilbake til, hadde jeg selv vurdert oppgaven til samme karakter som studentene. Dersom det hadde vært forskjeller, ville vi kanskje ha kommet opp i problemer, men de meget klare kriteriene som var satt opp gjorde jobben mye lettere.

Jeg tviler på at tilbakemeldingen studentene ga hverandre ville vært like god uten alvoret som lå i situasjonen nettopp fordi tellende karakterer skulle settes. Selv om en student melder at ”Jeg tror ikke alle var like godt forberedt” (student 6), ga alle

uttrykk for at tilbakemeldingen de fikk på utkastet var eksepsjonelt god. Tilbakemeldingen fra meg var begrenset: jeg ga hver student en skriftlig kvalitativ vurdering på noen linjer etter at studentene hadde diskutert oppgaven seg i mellom. I de fleste tilfeller hadde studenten da allerede hørt de samme synspunktene fra de andre studentene, og nikket gjenkjennende til mine kommentarer. Jeg tror det å oppleve at ens medstudenter har så god innsikt og evne til å gi tilbakemelding, er svært verdifullt, spesielt for studenter på Masternivå som skal være i ferd med å utvikle selvstendige og kollaborative læringsstrategier som ikke er avhengige av å ledes av autoriteter. I arbeidslivet vil de måtte tåle å gi og motta tilbakemeldinger til og fra sine medarbeidere. Forhåpentligvis ga denne erfaringen studentene mer tro på egne vurderingsevner.

Det var ett tilfelle hvor studentene ikke ble enige. Studentene "Arne" og "Beate" skulle vurdere "Carina" sitt arbeid, og mens "Arne" mente at "Carina" fortjente en B, var "Beate" skråsikker på at hun burde stryke. "Beate" hadde selv mottatt karakteren B på sitt arbeid, som hun mente fortjente en A. Jeg hadde vurdert "Carinas" arbeid til B, og "Beate" var sjokkert over at hennes arbeid og "Carinas" arbeid kunne vurderes til samme karakter. Hennes arbeid var mye bedre enn "Carinas", mente hun bestemt.

Det som skjedde først var at gruppen havnet i en konflikt de ikke kunne løse, og ba meg om å megle. Jeg hørte på deres argumenter, og fremla så min egen vurdering, både en beskrivende kvalitativ vurdering og begrunnelsen jeg ga for at jeg syntes arbeidet fortjente en B. Det jeg synes var spesielt interessant i det som fulgte, var at "Beate" ikke godtok min begrunnelse, men når hun først hadde hørt den kunne diskusjonen i gruppen fortsette. De kom til slutt fram til en konsensus hvor "Beate" var enig i at arbeidet fortjente en B, og hvor det (enda viktigere) virket som om hun forsto grunnene til dette. "Beate" trengte også en lang diskusjon med sine medstudenter for å forstå hvorfor hennes eget arbeid ble vurdert til B og ikke til A som hun selv først mente var riktig.

Det er ikke sikkert at "Beates" evne til å vurdere kvaliteten av eget og andres arbeid er forbedret av denne prosessen, men det ga henne i hvert fall en sjelden sjanse til ikke bare å se hvorvidt ens egne vurderinger samsvarer med andres, men også til å granske *hvor* avvikene ligger og hva de skyldes. Uten karaktersettingen ville ikke disse avvikene kommet like godt til syne.

Studentreaksjoner i ettertid

Uken etter vi hadde gjennomført medstudentvurderingen ba jeg studentene fylle ut et evalueringsskjema (se **Figur 2**). Det mest slående resultatet er at studentene i den numeriske tilbakemeldingen er svært positive *etter* å ha gjennomført vurderingen, mens de oppgir å ha vært svært negative *før* gjennomføringen. På en skala fra 1 til 10, der 1 er ”svært skeptisk” og 10 er ”kjempeentusiastisk” havner studentene i snitt på 3 før medstudentvurderingen men på 7.07 etterpå. Det er en radikal forandring.

Evaluering av medstudentvurderingen i HUIN204/303	
Hvilket nivå er du student på?	HUIN204 HUIN303
Er det greit om det du skriver her tas med i en artikkel Jill skal skrive om medstudentvurdering? (Anonymt)	Ja Nei
Sett ring rundt tallet som best angir din innstilling til medstudentvurderingen – 1 er dypt skeptisk, 10 er kjempeentusiastisk.	
Hva synes du om medstudentvurderingen da du først fikk høre om den?	1 2 3 4 5 6 7 8 9 10
Hva synes du om medstudentvurderingen rett før den skulle gjennomføres?	1 2 3 4 5 6 7 8 9 10
Hva synes du om den nå i ettertid?	1 2 3 4 5 6 7 8 9 10
Synes du det er noe som bør gjentas?	Ja Nei Kanskje
Skriv noe du synes var bra med medstudentvurderingen:	
Skriv noe du synes ikke var bra:	
Hva ville du gjort annerledes om du var læreren og du skulle lede en medstudentvurdering i dette emnet?	
Er det noe annet du vil si? Om du krysset avat det "kanskje" bør gjentas, forklar litt mer her.	
Seksjon for humanistisk informatikk, 30. november 2003.	

Figur 2: Studentene fylte ut dette evalueringsskjemaet uken etter medstudentvurderingen ble gjennomført.

Likevel er det kun to av de syv studentene som ga tilbakemelding som mener at medstudentvurderingen bør gjentas. Fire sier riktignok at den ”kanskje” bør gjentas, og de fleste av disse skriver i kommentarene at den bør gjentas, men da enten uten karactersetting eller med en karakter som teller mindre enn 40%, som de opplever som svært mye.

De negative kommentarene studentene kommer med har jeg allerede diskutert. Hovedpoengene er at de ikke liker å ha ansvaret for andres karakterer, og at de syns karakteren som ble gitt burde telle mindre.

Det var også mange positive reaksjoner til medstudentvurderingen. Spesielt satte studentene pris på den gode tilbakemeldingen de fikk, og de likte også at de fikk mulighet til å være med på å diskutere vurderingen.

Skriv noe du syns var bra med medstudentvurderingen:

”Diskusjon og feedback. Det ble brukt mer ’man hours’ på gjennomgang enn det som er vanlig med ekstern sensor. Mulighet til å forsvare/forklare designvalg etc.” (Student 3)

” - Konstruktiv tilbakemelding.
- Samarbeidet, flere synspunkter på arbeidet.” (Student 4)

”Forskjellig tilbakemelding og at vi fikk trening i å sette oss inn i rollen som sensor.” (Student 5)

”Fint å vurdere andre samt at vi ser hvor vi selv står i forhold. Også positivt at vi får andre vurderingsformer enn skoleeksamen.” (Student 6)

”Lese andres oppgaver, få innblikk i andres tema. Øve seg på å gi og få kritikk.” (Student 1)

”Alt i alt, en nyttig erfaring.” (Student 4)

Det kommer tydelig fram av disse kommentarene at studentene ønsker mer tilbakemelding enn de får i tradisjonelle vurderingsformer, og at de setter pris på å få tilbakemeldinger fra flere perspektiver. En så detaljert og grundig tilbake-melding er det nesten umulig å få til uten å bruke en form for medstudent-vurdering. Flere av studentene nevner at det er nyttig å få se eksempler på andres arbeid og at man lettere dermed kan vurdere sitt eget arbeid. Dette var mitt opprinnelige mål med å gjennomføre medstudentvurderingen: jeg ville gjerne at studentene selv skulle oppleve noe av den læringen jeg opplevde i sensorrollen.

Sammendrag av studentenes evaluering av medstudentvurderinge

student	1	2	3	4	5	6	7	snitt
før:	3	1	7	2	2	5	1	3
rett før:	5	1	3	2	1	7	2	3
etter:	8	5	5	8,5	5	8	10	7,07
gjenta	kanskje	kanskje	kanskje	ja	nei	ja	kanskje	
positivt								
God tilbakemelding	x	x	x	x			x	5
Trening i å gi og motta kritikk	x				x	x		3
Flere synspunkter på arbeidet				x	x			2
kunne forsvare seg			x					1
bedre gjennomgang enn vanlig			x					1
Bra å få se kriterier for karakter	x							1
Tvinges til å arbeide tidligere				x				1
Tvinges til å arbeide tidligere				x				1
Ser hvor man selv står						x		1
Positivt med annet enn skoleeksamen								
Negativt								
Burde telle mindre		x	x			x	x	4
Likte ikke ansvar for andres karakterer				x	x	x	x	4
Urettferdig, trynefaktor	x		x					3
Sosialt ubehagelig	x			x		x		3
Burde ikke telle på slutt karakter	x	x						2
Faglig ukvalifisert til å vurdere					x	x		2
For oppfatt av tekniske detaljer		x						1
Prinsippene må avklares tidligere				x				1

Fra evalueringen tilsutt i semesteret: Tenkehattene svært populære, må beholdes! En student skrev likevel i ettertid og sa bl.a. at "Den [tenkehattene] forsto jeg liksom ikke helt, og synes jeg er litt vel gammel for å bli utsatt for direkte korrekt pedagogikk."

Figur 3: Sammendrag av studentenes evaluering av medstudentvurderingen

Konklusjoner

Det er slående at til tross for de mange innvendinger og bekymringer var studentene overveiende positive til medstudentvurderingen *etter* å ha gjennomgått den. Før gjennomføringen ga studentene medstudentvurderingen i snitt 3 av 10, hvor 1 uttrykte dyp skepsis og 10 stor entusiasme. Etter gjennomføringen hadde snittet steget helt til 7,07 av ti.

Innvendingene studentene kom med er likevel alvorlige nok til at en lærer bør vurdere hvordan en eventuell medstudentvurdering skal gjennomføres nøye. Man skal være oppmerksom på at selv når man krever at studenter skal ta ansvar for egen læring, ønsker de sannsynligvis ikke å ta ansvar for egen vurdering. Man skal være klar over det sosiale ubehaget de opplever ved å skulle vurdere andres arbeid.

Likevel er fordelene ved en medstudentvurdering store. Tilbakemeldingen blir mer variert og av større omfang, og studentene kan få større tro på eller større innsikt i egen vurderingsevne. I tillegg gir medstudentvurderingen viktig kompetanse i samarbeid og i å gi og motta kritikk, og dette er ting som er viktige i videre læring også etter universitetsutdannelsen.

En enkel løsning ville være å fjerne karakterene, men jeg har ikke tro på at studentene vil legge like mye energi i tilbakemeldingene sine dersom dette alvorlig forsvinner. Kanskje det vil holde å senke vekten av karakteren, slik som studentene selv foreslår.

Det er mye man må tenke gjennom når man skal la studenter vurdere hverandre, ikke minst at det er en svært uvant og ofte ubehagelig sosial situasjon for dem å være i. Jeg mener likevel at det er en viktig lærerfaring for studentene å oppleve. I jobbsammenheng og i livet forøvrig er det helt nødvendig å kunne gi og motta kritikk på konstruktiv vis, og det bør vi gi studentene våre anledning til å øve seg på.

Litteraturliste

- de Bono, Edward. *Six Thinking Hats*. Back Bay Books, 1999.
- Dysthe, Olga, and Knut Steinar Engelsen. "Mapper Som Lærings- Og Vurderingsform." *LUNA*; (2003).
<http://luna.itu.no/Fokusomraader/digitale_mapper/mappeboka/mb04>
- Huba, Mary E., and Jann E. Freed. *Learner-Centered Assessment on College Campuses: Shifting the Focus from Teaching to Learning*. Boston: Allyn and Bacon, 2002.
- Johnson, Anne Sunniva Espolin. *Innføring Av Mappe Som Arbeids- Og Vurderingsform*. Oslo: Høgskolen i Oslo, Avdeling for sykepleierutdanning, 2003.
<<http://www.hio.no/content/view/full/1040/>>.
- Kirke-, utdannings- og forskningsdepartementet. *Gjør Din Plikt - Krev Din Rett: Kvalitetsreform Av Høyere Utdanning.*, 2001.
- Mowl, G., and Rachel Pain. "Using Self and Peer Assessment to Improve Students' Essay Writing: A Case Study from Geography." *Innovations in Education and Training International* 32.4 (1995): 324-45.
- Sivan, Atara. "The Implementation of Peer Assessment: An Action Research Approach." *Assessment in Education* 7.2 (2000): 193-213.
- Topping, K. J., et al. "Formative Peer Assessment of Academic Writing between Postgraduate Students." *Assessment & Evaluation in Higher Education* 25.2 (2000): 149-69.
- van den Berg, Ineke, Wilfried Admiraal, and Albert Pilot. "Peer Assessment in University Teaching. An Exploration of Useful Designs." *European Conference on Educational Research*. University of Hamburg, 2003.
<<http://www.leeds.ac.uk/educol/documents/00003178.htm>>