

Studentaktivisering gjennom bruk av hverandrevurdering for førstesemesterstudenter i Canvas LMS: en forsøksstudie

Omid Mirmotahari, Yngvar Berg og Dag Langmyhr

Institutt for Informatikk, Universitetet i Oslo

Ester Fremstad og Crina Damsa

Institutt for Pedagogikk, Universitetet i Oslo

Sammendrag

I denne forsøksstudien har vi gjennomført hverandrevurdering på et førsteårs-emne i informatikk. Hverandrevurdering fremheves i forskningen som en metode for å aktivisere studentene, og å styrke deres læring og metakognisjon. Ikke minst kan det å vurdere andres faglige prestasjoner være en måte for begynnerstudenter å utvikle innsikt i de faglige kriteriene og arbeidsmåtene i faget. Det kan også bidra til deres forståelse av egen læringsprosess og hvordan de best kan jobbe med studiene. Vi brukte Canvasfunksjonen Rubrics og SpeedGrader til å administrere hverandrevurdering av 524 studentbesvarelser. Det vi ønsket å undersøke var hvordan studentene erfarte denne typen læringsaktivitet; hvorvidt studentenes vurderinger var i tråd med faglærers vurdering; på hvilken måte og i hvilken grad hverandrevurderingen bidro til studentenes læring; samt hvordan data generert gjennom denne aktiviteten gir informasjon til faglærer som kunne brukes til å forbedre og tilpasse undervisningen. Til tross for enkelte tekniske mangler og utfordringer ga forsøket mange positive resultater.

Keywords: Studentaktivisering, Transparent undervisning, Constructive alignment, Hverandrevurdering, Rubrikk vurdering, Tilbakemeldinger, Data-dreven beslutning, Canvas.

1 Introduksjon

Ansvar for egen læring, metakognisjon, transparens knyttet til læringsmål og vurdering, tilbakemeldinger, og ikke minst en undervisning der studentene aktiviseres og der det er samstemthet mellom arbeidsformer, mål og vurdering (eng «constructive alignment» [Biggs and Tang, 2007]) fremheves som relaterte og sentrale elementer for å sikre kvalitet i høyere utdanning [Kunnskapsdepartement, 2017, Evans, 2013, Bransford et al., 2000, Nicol and MacFarlane-Dick, 2006].

Forskningsresultatene til [McMahon and Thakore, 2006] viser at samstemt undervisning bidrar til økt transparens, bedre læring, mer valid og pålitelig vurdering, økt studentaktivisering, og økt koherens i læringsutbyttet. [Morris, 2008] viser at det i emner med samstemt undervisning oppnås høyere snittkarakter, samt at en sterk korrelasjon mellom

Denne artikkelen ble presentert på konferansen NIK-2018; se <http://www.nik.no/>.

studentenes mestringsfølelse og prestasjon på eksamen. [Volante et al., 2010] viser at involvering i vurderingsarbeidet fører til at deltakerne tar mer ansvar og styring for egen læring. [Dochy et al., 1999] omfattende litteraturstudie påpeker positive effekter på studentenes læringsubytte ved å bruke hverandrevurdering (eng. «peer review»). Studentene blir mer involvert i lærings- og vurderingsprosessen. [Gibbs, 2006] studier viser at det ikke er kvaliteten av tilbakemeldingen fra hverandrevurderingen som øker studentaktiviseringen, men snarere det at studentene vet at deres oppgave blir vurdert av medstudenter. Spesielt viktig er det for førsteårs studentene at de får en bedre og dypere innsikt i faget, samtidig som de gjennomgår en personlig utvikling relatert til generell profesjonell kompetanse [Dochy et al., 1999, Boud, 2000, Ngar-Fun Liu and Carless, 2006, Hounsell et al., 2008, Mirmotahari and Berg, 2017b]. Hverandrevurdering gir nyttig læringsutbytte for både den som mottar tilbakemelding og den som vurderer oppgaven. Studentene lærer hva som verdsettes i en besvarelse, og å se etter vanlige feil og mangler, noe som bidrar til å gi studentene et meta-perspektiv på egne kunnskaper, ferdigheter, og læring [Mirmotahari and Berg, 2017a].

Det administrative aspektet ved hverandrevurdering kan være tidskrevende og utfordrende for faglærer, spesielt for emner med over 100 studenter. Det finnes flere online programmer som håndterer den administrative delen av innlevering og utlevering av oppgaver. Vi har valgt å benytte oss av Canvas som læringsplattform (LMS) og bruker Canvas sin egen hverandrevurdering gjennom Rubrics og SpeedGrader for å administrere hverandrevurdering for de 562 påmeldte studentene i innføringsemnet IN1020 ved Universitetet i Oslo.

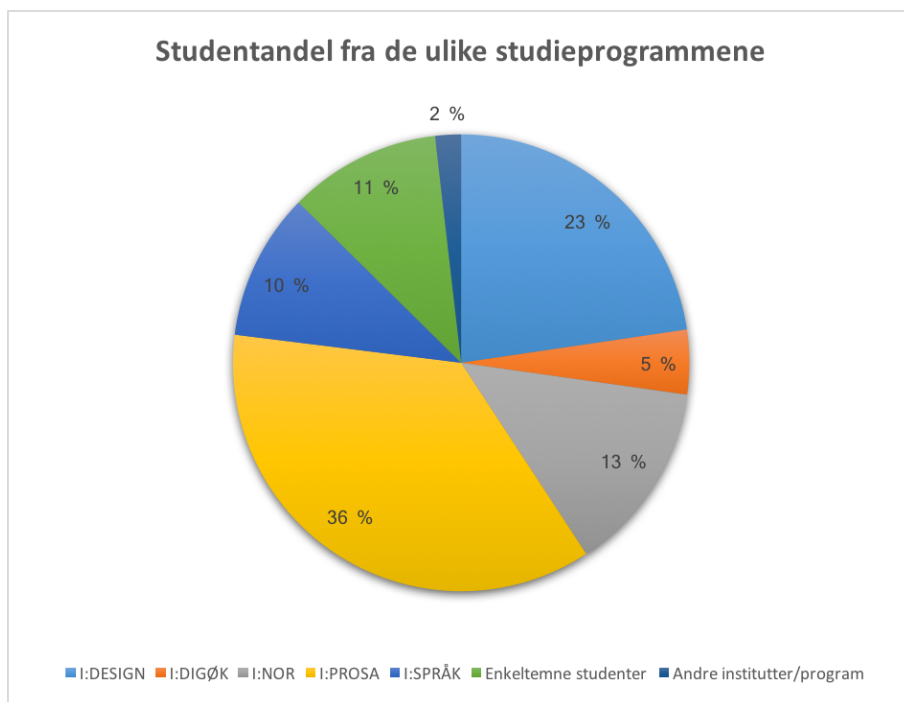
Motivasjonen for å gjennomføre denne forskningsstudien var ønsket om å belyse:

- (1) I hvor stor grad anerkjenner de sine medstudenters kunnskapsnivå?
- (2) Hva vektlegger studentene i sine vurderinger?
- (3) Hvordan opplever studentene bruk av Canvas?
- (4) Hvordan data innsamlet gjennom forsøket gir informasjon til faglærer om studentenes læring og faglige kompetansenivå.

I denne artikkelen gis først en beskrivelse av forskningsmetoden i avsnitt 2, mens resultatene blir presentert i avsnitt 3. Artikkelen avsluttes med en diskusjon av de ulike funnene og resultatene. Her besvarer vi forskningsspørsmålene med tilhørende forslag for videre arbeid.

2 Metode

Emnet IN1020 - Introduksjon til datateknologi er et obligatorisk førstesemesters emne som ble undervist for første gang høsten 2017. Ettersom emnet er obligatorisk for alle studenter tatt opp til de fem studieprogrammene ved instituttet er studentmassen en særdeles heterogen gruppe. Bakgrunnen til studentene er veldig varierende spesielt innen matematikk, fysikk og informatikk på grunn av opptakskravene til de representative studieprogrammene. Det var totalt 562 som var oppmeldt til emnet høsten 2017 og fordelingen mellom de ulike studieprogrammene vises i Figur 1. Emnet består av fire fokusområder innen feltene maskinvare, programmering (lavnivå), nettverk og sikkerhet. Hvert av disse fokusområdene har blitt undervist av egne faglærere. Undervisningen i emnet strekker seg over 14 uker med to timer forelesning, to timer gruppeundervisning og åpent tilbud om en time med populærvitenskapelig foredrag per uke. I forkant av eksamen



Figur 1: Fordelingen av de 562 påmeldte studentene og deres representative studieprogrammer. Det er veldig ujevnt fordeling av kjønn på de ulike programmene, men i snitt er kjønnsfordelingen på 70% menn og 30% kvinner.

skal alle studentene ha bestått 3 obligatoriske oppgaver. Sluttkarakteren, bestått/ikke-bestått, er kun basert på avsluttende skriftlig 4-timers eksamen. Eksamensoppgavene i 2017 bestod av 37 deloppgaver med ulik vekt, men 25% for hver fokusområde.

I dette forsøket tar vi for oss den første obligatoriske innleveringen. Denne innleveringen er to-delt, hvor første del er en kretsimplementasjon av en 4-bits multiplikator, mens andre del er hverandrevurdering. I første innleveringen skal de laste opp både kretsdesignet de har utført i LogiSim, samt en utfyllende rapport som beskriver fremgangsmetoden og løsningen de har brukt. Det ble innlevert 524 besvarelser innen fristen. Som del 2 (hverandrevurdering) fikk alle automatisk tildelt minst tre medstudentbesvarelse som de skulle vurdere.

Involverte parter i dette forsøket har vært faglærere, gruppelærere og studentene. Innsamlingen av data er gjort gjennom Canvas, spørreskjema, kursevaluering foretatt av fagutvalget og kvalitative intervjuer med studentene.

Oppsett i Canvas

Vi valgte å ta i bruk Canvas for administreringen av hverandrevurderingen. Det finnes mange gode instruksjonsartikler for måter å sette opp et hverandrevurderingssystem på Canvas sine egne nettsider og andre utdanningsinstitusjoner, men dette er ikke tema i denne artikkelen. Vi valgte å benytte oss av innebygde funksjoner som Rubrics og SpeedGrader. Rubrics er et rubrikkskjema som faglærer setter opp som en vurderingsmal. Studentene skal bruke dette skjema for å vurdere hverandres oppgaver. Rent administrativt var det satt opp slik at studentene skulle levere inn sin obligatoriske oppgave først innen en gitt frist. Deretter vil de som har levert innen fristen får utdelt tre medstudenters besvarelse etter fem dager. For hverandrevurderingen var det satt av en tidsfrist på

Kriterier	Vurderinger					
(1) Virker implementasjonen i LogiSim?	Ja		Delvis		Nei	
(2) Hvordan er implementasjonens struktur og ryddighet?	Veldig oversiktlig og høy grad av struktur og ryddighet.	God struktur.	Har struktur, men alt er i ett vindu (ingen subkretser).	Rotete, noe orden.	Helt fraværende for struktur eller ryddighet.	
(3) Oppfyller rapporten kravene satt i oppgaven?	Kravene er oppfylt og rapporten inneholder en konseptuell skisse av kretsen som helhet i tillegg til å beskrive subkretser i detalj. Organiseringen er gjort på en logisk og effektiv måte som kobler viktig informasjon til hverandre.	Kravene er oppfylt, men kunne ha vært bedre med en liten omorganisering av innholdet.	Kravene er oppfylt, men flyten er noe svak (ikke god sammenheng); burde vært gjort noen endringer for å gjøre den lettlest.	Enkelte krav er ikke oppfylt.	Har vesentlige mangler for grunnleggende informasjon.	
(4) Hvor god er kandidaten til å « snakke fag » og formidle sin løsning i de innledende 300 ords avsnittet som beskriver løsningen og andre alternative løsninger?	Mange fagrelevante ord er brukt. Teori og matematikk blir brukt for å beskrive anvendelsen på en måte som gjør det lett å forstå løsningen. Avsnittet er skrevet med god struktur og relevant informasjon er koblet sammen.	Fagrelevante ord er brukt. Teori og matematikk blir brukt for å beskrive anvendelsen, men er skrevet på en måte hvor flyten gjør det vanskelig å forstå.	Enkelte fagrelevante ord er brukt og beskrivelsen er lettlest, men ikke komplett. Relevante matematiske konsepter er nevnt, men kunne ha hatt mer detalj.	Teksten er vanskelig å følge og forstå. Beskrivelsen er ikke komplett.	Få/ingen fagrelevante ord er brukt og beskrivelsen av løsningen er vanskelig å forstå. Irrelevant/ingen matematikk er brukt for å gjøre logikken konkret.	
(5) Hvor bra sammenheng det mellom beskrivelsen av løsningen og selve løsningen?	Løsningen er som en visuell tegning av løsningsbeskrivelsen. Alt som er beskrevet, er blitt implementert på en ryddig måte.	Løsningen er relevant, men inneholder viktige elementer som ikke er beskrevet.			Løsningen har ingen likhetstrekk med beskrivelsen. Oppgaven er løst på en helt annen måte enn beskrevet.	
(F) Ville du ha godkjent denne oppgaven?	JA			Nei		
(G) Hvilke karakter vil du gi denne besvarelsen?	A	B	C	D	E	F

Tabell 1: Tabellen viser vurderingskriteriene som ble brukt i Canvas Rubrics.

en uke. Selve vurderingsrubrikken fikk studentene kun se når de skulle starte med hverandrevurderingen. Det betyr at de ikke har fått kjennskap til vurderingsrubrikken tidligere. Vi valgte å benytte oss av en rubrikk med ulike vektning og attributter, se Tabell 1. Rubrikken er inndelt i fire deler, hvor de to første delene, punkt (1)-(2) og (3)-(5), er vurderinger basert henholdsvis på den innleverte kretsimplementasjonen og rapporten. Del tre og fire, (F) og (G), er rent administrative punkter for studentenes vurdering av oppgaven til godkjent/ikke-godkjent og en karaktervurdering. Disse to siste punktene er ikke tellende med i poengsummen som hver student får. I analysen som følger vil fordeling mellom de ulike punktene fremkomme tydeligere. Etter gjennomføringen av hverandrevurderingen får studentene tilgang til vurderingene som er gjort av sin egen besvarelse.

3 Resultat

Studentenes eksamensresultatert er krysskoblet med resultatene fra den første obligatoriske oppgaven. Selv om eksamenvurderingen har vært bestått/ikke-bestått, har det vært mulig å hente ut hva studentene har oppnådd av poeng på eksamensdelen som er relevant for dette forsøket. Slik eksamen i 2017 var utformet var det totalt mulig å oppnå 25 poeng på denne delen i eksamen og for enkelhets skyld for å kunne sammenligne har vi benyttet oss av en karakterfordeling basert på 91-80-60-49-40% (av 25 poeng) for henholdsvis A-B-C-D-E og hvor F er under 40%.

Det var 36% av studentene som har fått samme karakter på eksamen som den har både gitt til sine medstudenter og selv har fått fra andre, mens det er 46% som har fått ± 1 karakter. Dette betyr at hele 82% av studentene har allerede ved første obligatoriske oppgave fått muligheten til å vite hvordan de ligger ann i emnet. (Ettersom dette er en tilfeldig generert utvalg og ikke styrt kan vi ikke direkte si at det er en korrelasjon med at en «A-student» gir de andre «A».) I forhold til tidsrammen på 7 kalendardager som studentene fikk for å gjennomføre hverandrevurdering sier 82% at det har vært passe med tid for den faglige vurderingen, dog mener de at det har gått mye tid til spesifikke tekniske Canvas problemer. Når det gjelder anonymiteten i forsøket mener majoriteten at det er et viktig grep, 17% av studentene sier at det er viktig og 42% mener det er veldig viktig. Et av punktene forskningslitteraturen tar opp er studentenes tillitt til den faglige kompetansen til sine medstudenter. I dette forsøket svarer majoriteten av studentene som forventet, 76% har ingen eller liten tillitt. Spesielt tydelig fremkommer dette fra studentsitatet 1.

«Unødvendig med kameratvurdering, i så fall må det stilles bedre krav til refleksjon og gjennomgang av eget arbeid versus studiekamerat.»
(Student #2476934)

Sitat:1

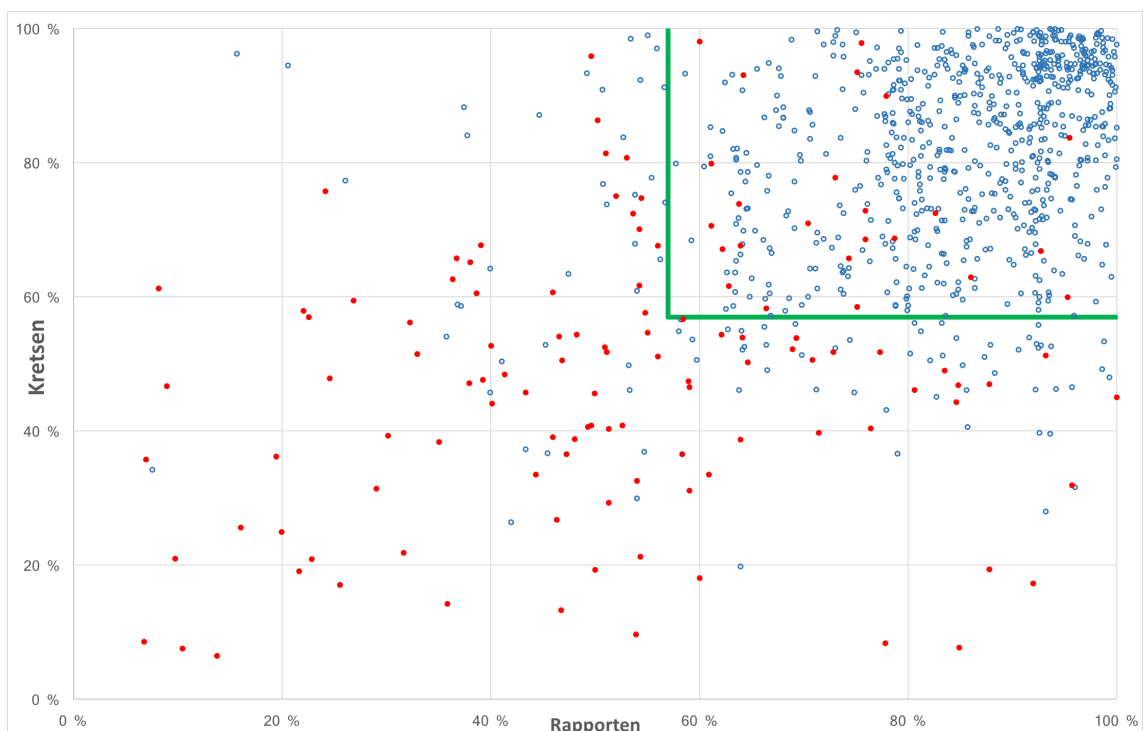
Den obligatoriske oppgaven har vært todelt, kretsimplementasjon og rapport, hvor begge delene skal telle med i vurderingen. Rubrikken er inndelt for å foreta vurdering for begge delene og angi om oppgaven er bestått eller ikke-bestått. For hver rad i rubrikken er det angitt en poengskala 1 – 5, bortsett fra kriteriet (1) med poeng 0-2 og kriteriet (5) med poeng 0-3-5. Det er mulig å oppnå totalt 22 poeng. Avkrysningen i rubrikk (F) og (G) er ikke med i poengberegningen. Det har ikke vært oppgitt fra faglærer hva poenggrensen for å få godkjent/ikke-godkjent skal være, ei heller hva karaktertersklene. Det er flere interessante analyser for hvordan studentene har vurdert hverandre og hva de vektlegger i forhold til godkjent/ikke-godkjent og karakter. I Figur 2 ser vi distribusjonen av vurderingene gjort med hensyn til relasjonen poengskår (omregnet til %) for henholdsvis kretsimplementasjonen og rapporten. De røde punktene i figur 1 representerer oppgaver som har fått ikke-godkjent, mens de blå representerer godkjent. Selvom studentene ikke har fått oppgitt grensen for godkjent og ikke-godkjent har faglærer definert en grense, denne er merket inn med grønn linje. Som figuren viser er majoriteten av de som har fått godkjent innenfor den øverste kvadrant til høyre, mens majoriteten til ikke-godkjent utenfor. Hva kjennetegner avvikene kan vi se nærmere av figurene, Figur 3 og Figur 4. Disse figurene viser hvilke del har vært mest vektet for å gi godkjent eller ikke-godkjent. Det viser seg en direkte korrelasjon om at det er om kretsimplementasjonen virker eller ikke som er den signifikante avgjørende faktoren for om oppgaven blir godkjent eller ikke. Figur 3 viser oversikten over de oppgavene som har fått godkjent, men som har en total poengsum under 13 poeng. Som radarplottet viser er det rapporten som er den signifikante faktoren, hvor utslaget hovedsakelig ligger på 0 og 1 poeng, mens

kretsimplesimentasjonen har lite utslag i samme område. For motsatt tilfelle ser vi samme trend i Figur 4, der radarplottet viser utslagene for besvarelser som har fått ikke-godkjent, men mer en 13 poeng. Selvom utslaget for rapporten også slår mest ut i den øverste delen av poengskalaen så er det likevel kretsimplesimentasjonens virkemåte som overstyrer vurderingen ifht godkjent og ikke-godkjent. Det er tydelig at de som har fått ikke-godkjent med mer enn 13 poeng har levert særdeles gode rapporter.

Figur 5 viser avviket for studentenes karaktergivning relativt til faglærers. Studentene har ikke fått oppgitt faglærerens karaktergrenser. De grønne søylene, for karakter A og C, representerer der faglærer har flere kandidater til den respektive karakteren, mens rød er der studentene har flest kandidater. Det totale avviket relativt til antall besvarelser er 11%. Den multimodale karakteristikk har en snittverdi på 17% som faktisk er signifikant i forhold til karakterskalaen og som igjen kan bety mer enn en hel karakterforskjell i snitt. Spesielt tydelig blir det i Figur 6 hvor plottet viser karakterdistribusjonen gitt poengsum som studentene har vurdert. Ettersom forsøket tar for seg en obligatorisk oppgave er det ikke helt uventet at majoriteten av studentene gjennomfører oppgaven med høy poengscore og dermed også tilsvarende A.

Erfaringer med funksjonaliteten i Canvas

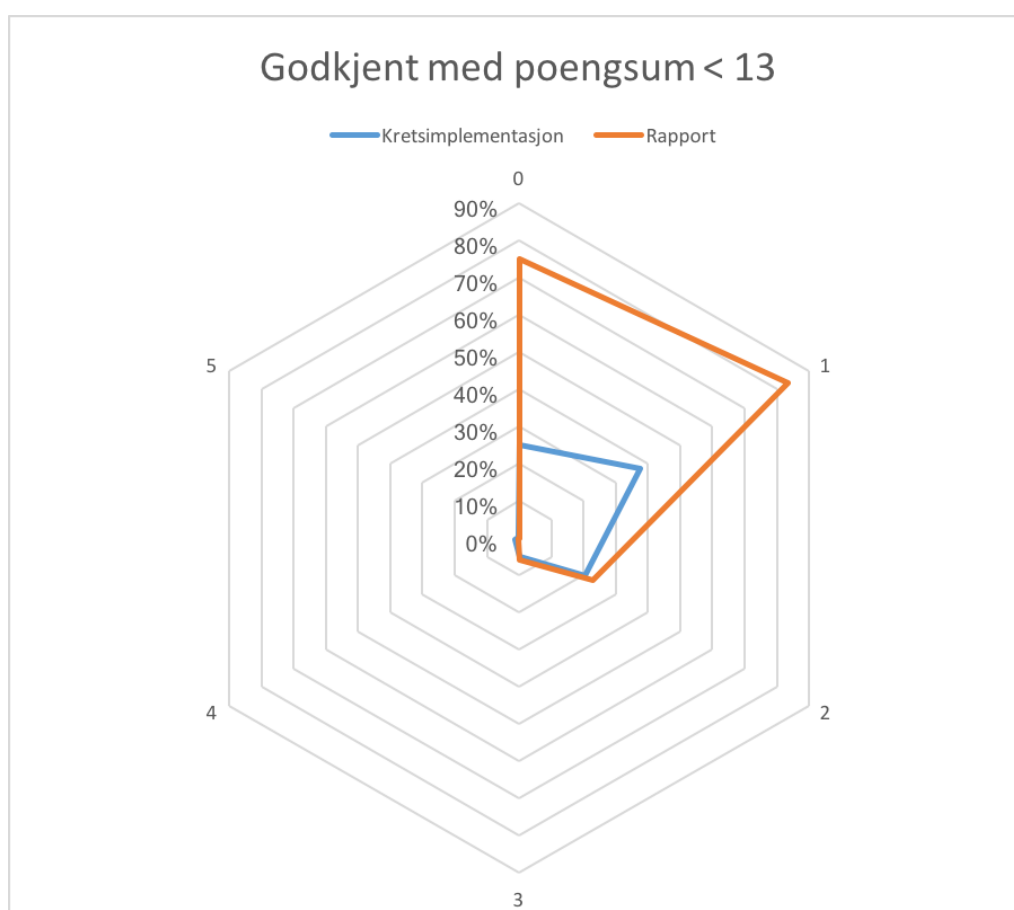
Hverandrevurderingen ble konfigurert som anonym i Canvas, men dessverre leverte alle studentene inn sine besvarelser med en forside med navn. Dette hindret oppnåelsen av full anonymitet selv om Canvas hadde skjult filnavn og avsender. Det vi oppnådde var derfor en delvis anonymisering, hvor alle vurderinger ble anonyme for mottakeren. En



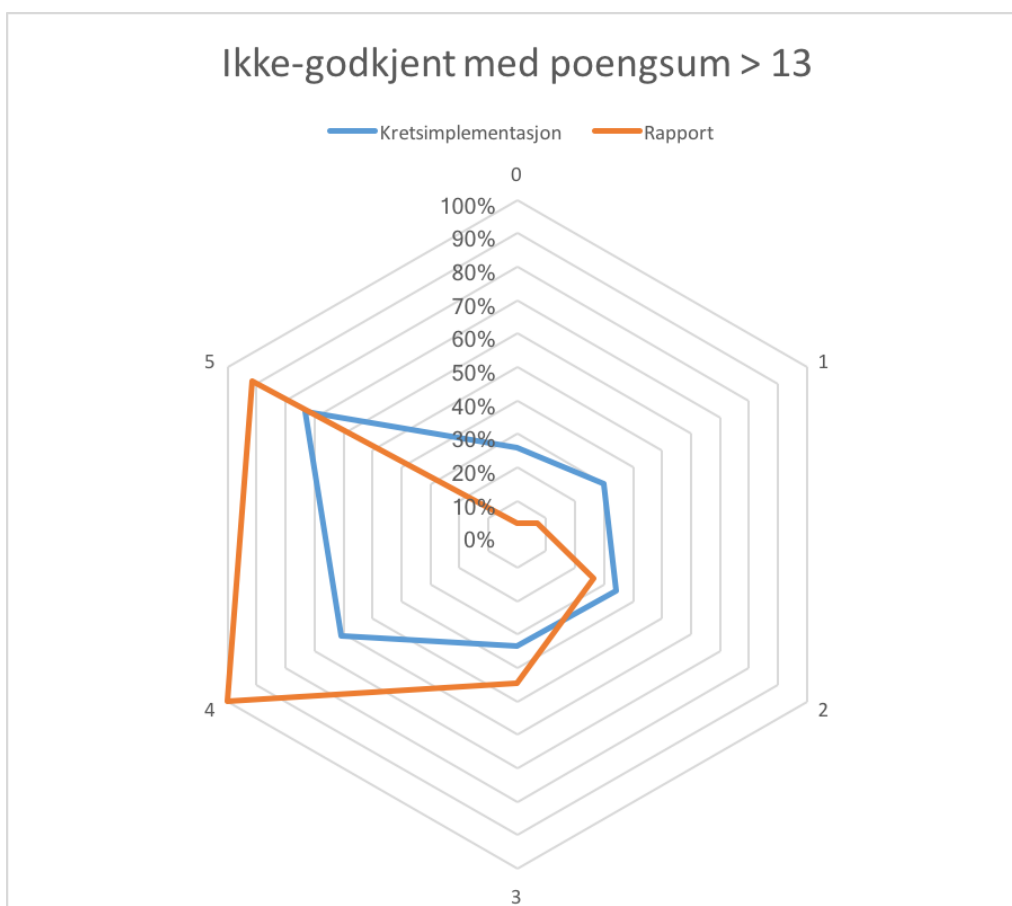
Figur 2: De blå punktene representerer godkjente oppgaver, mens de rød representerer ikke-godkjente oppgaver. Øverste kvadrant til høyre, avgrenset med grønne linjer, representerer mengde av oppgaver som faller inn under faglærerens krav om godkjent, tilsvarende 13 poeng.

gjennomføring av hverandrevurdering med et så stort antall deltakere bidrar til mange praktiske utfordringer som godkjente utsettelse av tidsfrister, stabilitet i Canvas når mange bruker den samtidig, systemkrasj på lokalmaskin osv. Dette bidrar til frustrasjon både for studenter og organisator for prosjektet. Det er veldig uheldig at førsteårs studenter får en slik opplevelse av canvas som gjen kan påvirke deres mestringsopplevelse og skygger for læringsutbyttet.

Studentene rapporterer mye frustrasjon i brukeropplevelsen med Canvas. De påpeker særlig to tilfeller, den ene i forhold til lagre-knappen i SpeedGrader som ofte og uforutsigende ble innaktiv. Dette førte til at studentene ble usikre på om deres arbeid var blitt videreformidlet eller ikke. Den andre var hovedsakelig rettet mot den tekniske funksjonaliteten til Canvas. Mange studenter ønsket å gjøre vurderingen over tid, men det er ikke mulig å mellomlagre en vurdering og vurderinger er heller ikke mulig å redigere i ettertid. Et element som indirekte har påvirket studentene i Canvas er muligheten for å sette opp at oppgaven skal overprøves av en faglærer, denne funksjonen valgte vi å ikke bruke på grunn av arbeidskapasitet. Således ble det derfor heller ingen sperre eller filter



Figur 3: Relasjonen mellom poeng-givning for kretsimplesmentasjonen og rapporten målt i signifikans. Utslaget representerer hvor mange som har fått de representative poengene (0-5) for henholdsvis kretsimplesmentasjonen og rapporten. Utvalget omfatter besvarelser som studentene har vurdert til godkjent, men som har fått en samlet poengsum som er lavere enn faglærers grense for bestått, som er 13 poeng. Slik grafen viser lar studentene kretsimplesmentasjonen være mer gjeldende og styrende for om medstudenter får godkjent eller ikke-godkjent.



Figur 4: *Relasjonen mellom poeng-givning for kretsimplementasjonen og rapporten målt i signifikans. Utslaget representerer hvor mange som har fått de representative poengene (0-5) for henholdsvis kretsimplementasjonen og rapporten. Utvalget gjelder hvilke studentene er blitt vurdert som ikke-godkjent men samlet poengsum mer enn faglærerens poeng-grense, som er 13 poeng. Slik grafen viser er ikke rapporten signifikant nok til å få godkjent eller ikke-godkjent.*

for hva studentene fikk tilsendt. Det vi savner er en mulighet for studentene til å kunne gi hverandre en tilbakemelding på vurderingen de har fått, alternativt kunne «flagge» vurderinger som de finner upassende. Vi opplevde kun én vurdering (av 1572) som var meget upassende.

«Det var vanskelig å finne fram, og når man skulle gi tilbakemelding til andre medstudenter var det problemer med tilbakemeldingsskjemaet, så man måtte vente til det plutselig funkete igjen. Tilbakemeldinger ble ikke lagret, knapper forsvant, layout og popup av tabeller fungerte dårlig, samt at det var kronglete å i det hele tatt finne fram til det man ville.»
 (Student #2486411)

Sitat:2

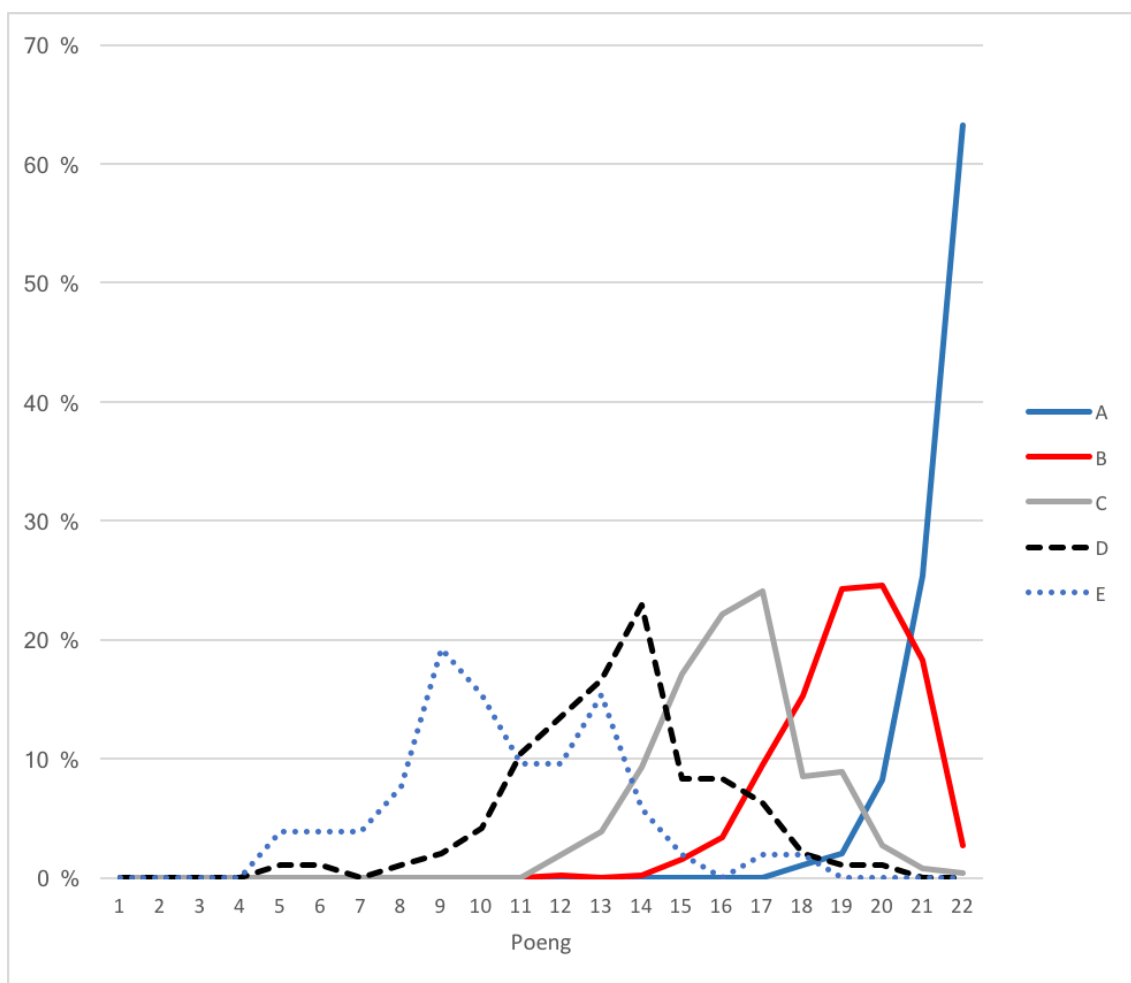


Figur 5: Avvikene mellom studentenes og faglærerens karaktergivning. Den relative verdien er positiv, grønn, der faglærer har flere kandidater og tilsvarende rød, hvor studentene har flere. Det totale avviket relativt til antall besvarelser er 11%.

4 Diskusjon

En interessant problemstilling som har dukket opp som følge av at studentene gjør hverandrevurdering er tillit til egen evne til å foreta rettferdig vurdering og å gi tilbakemelding til medstudenter. Dette er også påpekt av [Mulder et al., 2013]. Likeledes er spørsmålet om tillit til at egen besvarelse blir vurdert rettferdig av medstudenter sentral, slik også [Cheng and Warren, 1997] påpeker. Vi ser at studentene har begrenset tillit til hverandres kunnskapsnivå og dermed i utgangspunktet er noe skeptiske til vurdering gitt av medstudenter og ikke av faglærer. Dette er ikke unaturlig, i og med at studentene kan ha en negativ oppfatning av sitt eget kompetansenivå ved hverandreevalueringen. Dermed vektlegger de i mindre grad medstudenters evaluering. Vi ser av karakterfordelingen at studentene er gjennomgående mer kritisk enn faglærer og at de ikke evner i like stor grad å bedømme svært gode besvarelser (A). Det er ikke merkverdig at studentene mangler erfaring og selvtillit til å karaktersette besvarelser som er svært gode. Vi ser av karakterfordelingen at studentene bedømmer en større andel av oppgavene til B. Dette kan være et utslag av at mange gjennomsnittsstudenter bedømmer besvarelsene som litt bedre enn de selv oppfatter som sitt eget kunnskapsnivå. Vi ser også at andelen av svake karakterer er større blant studentenes karaktergivning. Dette kan bety at de legger merke til konkrete svake besvarelser og i mindre grad evner å se positive elementer i slike besvarelser. Selv om karaktergivingen av faglærer og studenter ikke viser betydelig avvik er det ingen tvil om at det er stor forskjell på erfaring i forhold til bedømming av besvarelser. Nyanseringer med hensyn på kvalitet i karaktergivingen kan ikke forventes av studenter i like stor grad som av faglærer.

Det er også tydelig at studentene vektlegger konkrete forhold, spesielt om



Figur 6: Plottet viser karakterfordelingen gitt hvordan studentene har vurdert hverandres besvarelser med hensyn på total karakter. Distribusjonen mellom de ulike karakterene er med snittpoeng 19 for B, 16 for C, 13 for D og 11 for E. Som for eksempel ser vi at i intervallet 11-15 poeng er majoriteten D-kandidater.

kretsimplementasjonen virker eller ikke, mer enn faglærer. Dette får da større betydning for godkjent/ikke godkjent bedømming. Slike konkrete, unyanserte, forhold bidrar til en forenkling for studentene når det gjelder å bedømme et sluttresultat. Det fremkommer også at studentene verdsetter rapporten lite. Sannsynligvis har de også lagt lite arbeid i sin egen rapport ved innlevering. For fremtidige undersøkelser vil det være svært interessant å se nærmere på hvordan de som har levert en god rapport har vurdert medstudenters rapporter. Som en videreføring av dette forsøket kan man implementere maskinlæringsalgoritmer til den tekstlige rapporten og bruke det til å finne ut av om det lar seg gjøre å finne hvilke faktorer som utgjør en god og en mindre god rapport. Gitt en så stor datamengde som ved årets innsamling vil det være tilstrekkelig for et første ordens utprøving av maskinlæring.

Når det gjelder selve prosessen og erfaringen med å delta i hverandreevaluering er det liten tvil om at dette bidrar til at studentene blir mer bevisst på hvordan besvarelser evalueres og hva som er viktig å være klar over for å oppnå en bedre evaluering av egne besvarelser. Bevissthet om evalueringsmetodikk er viktig for studentene i eksamenssituasjoner og til å styrke tilliten til karaktergivning.

Referanser

- [Biggs and Tang, 2007] Biggs, J. and Tang, C. (2007). *Teaching for Quality Learning at University Third Edition Teaching for Quality Learning at University*, volume 3th edition (1th edition 1999). Open University Press.
- [Boud, 2000] Boud, D. (2000). Sustainable Assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2):151–167.
- [Bransford et al., 2000] Bransford, J. D., Brown, A. L., and Cocking, R. R. (2000). *How People Learn*.
- [Cheng and Warren, 1997] Cheng, W. and Warren, M. (1997). Having second thoughts: Student perceptions before and after a peer assessment exercise. *Studies in Higher Education*, 22(2):233–239.
- [Dochy et al., 1999] Dochy, F., Segers, M., and Sluijsmans, D. (1999). The Use of Self-, Peer and Co-assessment in Higher Education: a review. *Studies in Higher Education*, 24(3):331–350.
- [Evans, 2013] Evans, C. (2013). Making Sense of Assessment Feedback in Higher Education. *Review of Educational Research*, 83(1):70–120.
- [Gibbs, 2006] Gibbs, G. (2006). How assessment frames student learning. In *Innovative Assessment in Higher Education*, chapter 2.
- [Hounsell et al., 2008] Hounsell, D., McCune, V., Hounsell, J., and Litjens, J. (2008). The quality of guidance and feedback to students. *Higher Education Research & Development*, 27(1):55–67.
- [Kunnskapsdepartement, 2017] Kunnskapsdepartement, D. K. (2017). Kultur for kvalitet i høyere utdanning. Technical report.
- [McMahon and Thakore, 2006] McMahon, T. and Thakore, H. (2006). Achieving constructive alignment: Putting outcomes first. *Aukšjojo mokslo kokybė*, (03):10–19.
- [Mirmotahari and Berg, 2017a] Mirmotahari, O. and Berg, Y. (2017a). Erfaringer fra strukturert peer review ved bruk av et egetutviklet sensureringsprogram. *Norsk Informatikkonferanse (NIK-UDIT)*, page 16.
- [Mirmotahari and Berg, 2017b] Mirmotahari, O. and Berg, Y. (mars 2017b). Individuell «automagisk» tilbakemelding på skriftlig eksamen. In *Artikkelsamling MNT-konferansen*, pages 109–114.
- [Morris, 2008] Morris, M. M. (2008). Evaluating university teaching and learning in an outcome-based model : replanting Bloom. *PhD thesis, School of Mathematics and Applied Statistics, University of Wollongong.*, page 370.
- [Mulder et al., 2013] Mulder, R., Baik, C., Naylor, R., and Pearce, J. (2013). How does student peer review influence perceptions, engagement and academic outcomes? A case study. *Assessment & Evaluation in Higher Education*, 39(6):1–21.
- [Ngar-Fun Liu and Carless, 2006] Ngar-Fun Liu and Carless, D. (2006). Peer feedback: the learning element of peer assessment. *Teaching in Higher Education*, 11(3):279–290.
- [Nicol and MacFarlane-Dick, 2006] Nicol, D. and MacFarlane-Dick, D. (2006). Formative assessment and selfregulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2):199–218.
- [Volante et al., 2010] Volante, L., Beckett, D., Reid, J., and Drake, S. (2010). Teachers' views on conducting formative assessment within contemporary classrooms. *Annual Meeting of the American Educational Research Association*, pages 1–23.