

# Peer review i et obligatorisk bachelorkursus for civilingeniørstuderende i medikoteknik

Jens E. Wilhjem, professor, BME, DTU Elektro, Danmarks Tekniske Universitet

Sidsel-Marie Winther Prag, e-læringskonsulent, DTU LearningLab, Danmarks Tekniske Universitet

## Videnskabelig artikel, fagfællebedømt

*Et nyudviklet peer review-system, der integrerer studenterdata, administration og grafiske procesoversigter, blev afprøvet i et obligatorisk bachelorkursus. Overensstemmelse mellem de studerende og hjælpelærernes kvantitative bedømmelser blev undersøgt for en relativt udfordrende opgave med 15 delspørgsmål. 49 studerende afleverede besvarelser. Efterfølgende blev hver besvarelse bedømt af tre medstuderende, svarende til at 735 delbesvarelser blev bedømt. Ud af disse var der enighed mellem studenterbedømmerne og hjælpelærerne i 480 tilfælde (65 %). I under 3,5 % af tilfældene var der total uenighed. Der var tendens til, at jo mere rigtigt en besvarelse blev bedømt, eller jo mere konkret svaret var, des mere enighed. En interviewundersøgelse hos de studerende viste stor opbakning til fremgangsmåden og viste også, at de studerende fandt det meget nyttigt at se andre studerendes besvarelser såvel som rettevejledningen. Afslutningsvis gives en række anbefalinger til processen og systemet.*

## 1. Indledning

Der er flere faktorer, der stedse øger undervisningsbyrden på universiteterne. Markante faktorer er stigende optag af studerende og ønskerne om at holde de studerende i et aktivt læringsforløb, samtidig med at de får vigtig feedback på deres aktiviteter (Andersen, 2016; Boud & Molloy, 2013). Ved aflevering af, for eksempel, skriftlige opgavebesvarelser kan det ensformige rettearbejde blive af betydeligt omfang og dermed udgøre (d)en begrænsende faktor (Andersen, 2016; Sims, 1989).

Peer review er et af flere tiltag, man kan ty til for at begrænse dette rettearbejde, når de studerende skal udarbejde en opgavebesvarelse. Hvis underviseren kan få alle aspekter af processen til at forløbe tilfredsstillende, vil peer review potentielt hurtigt og effektivt kunne bidrage med en stor mængde relevant feedback. Yderligere vil peer review fordre af de studerende, at de påtager sig nye roller til formodet gavn for læringen og egen udvikling (Nicol et al., 2014; Sims, 1989).

Den administrative proces ved peer review kan imidlertid være omfattende (Boase-Jelinek et al., 2013; Sims, 1989), men arbejdsbyrden kan reduceres kraftigt ved at implementere en computerbaseret styring af processen, der opfylder følgende: fuldstændig integration af alle studenterdata, instruktioner og opsætning, al fil- og resultathåndtering, direkte adgang til resultater af processen inklusive overskuelige grafiske oversigter samt en høj grad af datasikkerhed (Boase-Jelinek et al., 2013; Cho et al., 2006a).

Ved normal opgaveaflevering – og efterfølgende rettelser af underviser eller hjælpelærer – tager de studerende to roller: *forfatter* af besvarelse samt *modtager* af rettelser. Da sidstnævnte kommer fra underviserne, vil de ofte blive modtaget som sandhed og ikke debatteret yderligere. I peer review er der yderligere roller, hvorfor de studerende bliver (Hamer et al., 2015):

- *forfatter* af en opgavebesvarelse,
- *bedømmer* af andre medstuderendes arbejde, hvorved de skal danne sig en mening om besvarelsen samt generere en form for feedback; her vil de ofte få ny indsigt i alternative måder at besvare på, samt
- *modtager* af kritik, som kommer fra medstuderende og derfor ofte må antages at blive modtaget med en lidt større skepsis og kræve flere undersøgelser af validiteten.

Peer review involverer derfor mere og anderledes arbejde for de studerende, og nogle af disse elementer ser ud til at resultere i en større læring. For eksempel virker det plausibelt, at selve det at give (skrive) seriøs feedback til medstuderende tilføjer en ny, værdifuld læringsproces (Althaus, 2001; Cho & Mac-Arthur, 2011; Cho et al., 2011; Li et al.; Nicol et al., 2014), og ligeledes kan det at blive præsenteret for andre studerendes gode og dårlige besvarelser potentielt gøre de studerende mere kritiske over for deres eget arbejde (Hamer et al., 2015).

I nærværende artikel rapporteres om anvendelse af et nyudviklet peer review-modul i CampusNet (Arcanic A/S); CampusNet (i DTU Inside) er en fælles platform for alle DTU-kurser. Peer review i CampusNet integrerer studenterdata, administration og grafiske procesoversigter.

Modulet blev afprøvet i et obligatorisk kursus i fysikken i medicinsk billeddannelse (Kursushjemmeside, 2017) på bacheloruddannelsen i Medicin og Teknologi, der er en samarbejdsuddannelse på DTU og Det Sundhedsvidenskabelige Fakultet, Københavns Universitet. Kurset ligger på 5. semester, forløber over 13 fire-timers eftermiddagsmoduler fra september til december og har cirka 60 studerende. I kurset skal de studerende, opdelt i teams af fire, afdække vævstyper i en boks med ukendt biologisk væv indstøbt i agar (Wilhelm et al., 2008). Dertil skannes med ultralyd, MRI, PET, CT samt røntgenfotoafregning, og disse data analyseres efterfølgende af de studerende. Disse analyser muliggør, at de studerende kan se præcis det samme vævssnit med alle teknikker, men da teknikkerne er forskellige, vil billederne også være det. For at kunne bruge denne information er de studerende nødt til at studere fysikken bag billeddannelsen, hvilket er det andet hovedformål med kurset. Der foregår naturligvis en kamp mellem, hvor meget tid de studerende bruger på disse to hovedformål. De to hovedformål er forskelligartede og svære, og fremgangsmåden fordrer, at de studerende skaber overblik over, hvad der skal gøres hvornår og hvorfor. Til at understøtte dette afleveres fire obligatoriske opgaver i løbet af kurset (opgaverne dækker begge hovedformål). En af opgaverne (nummer 2) blev derfor udvalgt til et forsøg med peer review. Dette havde den yderligere fordel, at de studerendes nødvendige brug af en rettevejledning hjalp dem til at holde fokus på dét, der var vigtigt (Boase-Jelinek et al., 2013).

Peer review blev implementeret som en dobbelt anonymiseret proces (se senere), som nogle anbefaler (Howard et al., 2010; Sims, 1989), selvom andre er fortalere for total åbenhed (Boase-Jelinek et al., 2013).

I nærværende studie undersøges følgende spørgsmål for peer review anvendt til kvantitativ bedømmelse med formativ evaluering for øje:

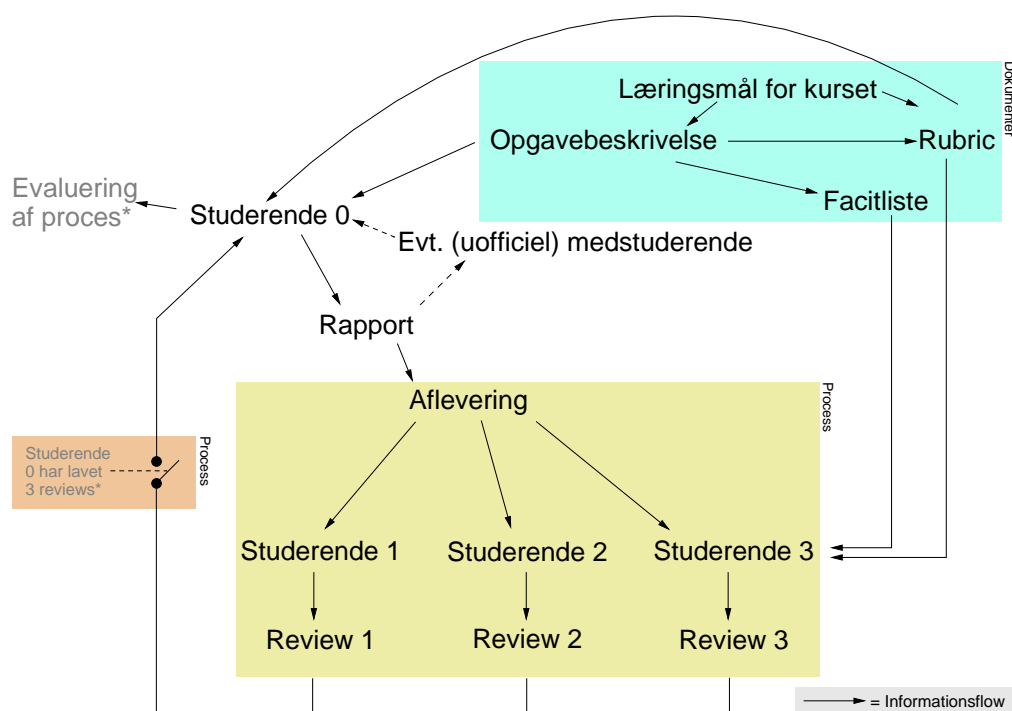
1. I hvor høj grad er de studerendes bedømmelser valide? (her vurderet med hjælpelærernes rettelser som "facit")
2. Hvor stor var de studerendes *a priori*-lyst til at deltage, selvom deltagelse ikke er obligatorisk for at bestå kurset?

3. Hvad syntes de studerende om peer review-processen umiddelbart efter?
4. Hvor meget kan underviserens tid mindskes, uden at det går ud over proces og de studerendes følelse af at blive retfærdigt behandlet?

I det følgende kapitel beskrives design og fremgangsmåde ved brug af peer review i CampusNet. Det efterfølgende kapitel beskriver de materialer og metoder, der blev anvendt til undersøgelsen af de fire ovenstående spørgsmål: den kvantitative sammenligning, som blev brugt til at undersøge spørgsmål 1, 2 og 4, og de semistrukturerede interviews, der blev brugt til at undersøge spørgsmål 2 og 3. Efter resultatafsnittet diskuteres svarene på de fire spørgsmål. Dette efterfølges af en konklusion og perspektiver for brugen og for fremtidig forskning samt praktiske anbefalinger.

## 2. Procesbeskrivelse for peer review i CampusNet

I det følgende beskrives den generelle proces, som er indtænkt i peer review-systemet i CampusNet. Den del, der bruges i nærværende studie, er en anelse mindre, hvilket fremgår af afsnit 3.



**Figur 1:** Oversigt over peer review-processen. Studerende 0 laver (eventuelt i samarbejde med medstuderende) en rapport (eller besvarelse) som afleveres til peer review-systemet. Tre tilfældige studerende (1-3) bedømmer denne rapport og melder - via peer review-systemet - tilbage til studerende 0. Dette ses dog kun, hvis sidstnævnte har lavet mindst 3 bedømmelser af *andre* studerendes opgaver. Endelig kan den studerende evaluere processen og kommentere de bedømmelser, der er modtaget. De med \* mærkede funktioner er ikke undersøgt her, men tænkes implementeret

Processen er vist skematisk i figur 1. Underviseren laver opgavebeskrivelsen og eventuelt en tilhørende såkaldt *rubric* (Panadero & Jonsson, 2013) under skyldig hensyntagen til læringsmålene for kurset. Rubric viser, hvad der er vigtigt og i hvilken grad. Om nødvendigt, udarbejdes også en rettevejledning, som i nogle tilfælde kan træde i stedet for en rubric. Den studerende laver sin

besvarelse (rapport) ud fra opgavebeskrivelsen (og rubric, når den findes). Besvarelsen bliver afleveret, og (her tre) tilfældigt valgte medstuderende laver hver deres bedømmelse.

Inden peer review begyndes, inddrages de studerende (Ulriksen, 2014) ved at rammesætte processen (klargøring af formålet og oprids af det forventede udbytte etc.), idet jo mere aktivt de studerende inddrages i processen, des højere sandsynlighed for succes (Evans, 2015).

### 3 Materialer og metoder

En beta-udgave af DTU CampusNets Peer Review-modul blev afprøvet i efteråret 2016 i kurset Indledende medicinsk billeddannelse (Wilhjem et al., 2008; Kursushjemmeside, 2017). Modulet blev først frigivet ved semesterstart, hvorfor deltagelse i dette forsøg var frivilligt. Beta-versionen havde visse begrænsninger i forhold til den endelige version, blandt andet kunne det kun lade sig gøre at give kvantitative bedømmelser. I nogle få tilfælde blev anonymiteten ikke opretholdt på grund af programmeringsfejl.

#### 3.1 Opgavebeskrivelse

Opgaven (Wilhjem, 2017), der skulle løses, fremstod som 15 spørgsmål indlejret i en *guided tutorial*, som de studerende blandt andet kunne arbejde individuelt med under to besøg på det hospital, hvor de medicinske skanninger foregik. De studerende skulle blandt andet lave en grafisk afbildning af, hvordan planar røntgenfotografering fungerer, samt producere en række visualiseringer af fysiske fænomener relateret hertil. Der var ofte mere end én rigtig besvarelse. Den kursusansvarlige var til stede for at besvare spørgsmål og guide.

#### 3.2 Procedure for den computerunderstøttede proces

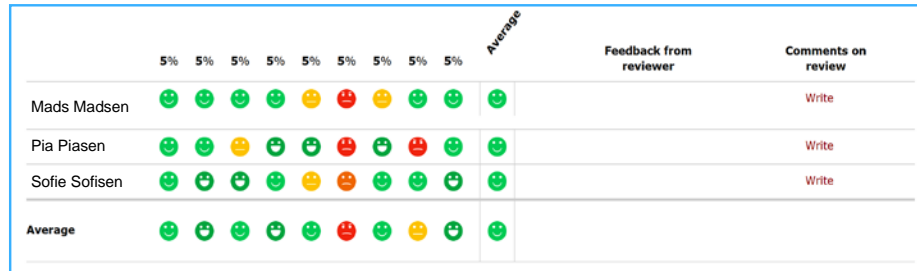
Besvarelsene blev af de studerende uploadet som en anonym pdf-fil på CampusNet i et dedikeret afleveringsmodul.

Efter aflevering blev de studerende tilfældigt fordelt med tre bedømmere på hver studenters besvarelse. De fik desuden udleveret en 6 sider lang rettevejledning (Wilhjem et al., 2016) (facitliste), som angav eksempler på korrekte svar for de 15 spørgsmål. Herefter kunne de studerende bedømme de medstuderendes svar. Idet opgaven var så svær, at en rettevejledning var nødvendig, blev der ikke anvendt en rubric; de studerende skulle blot bedømme hinanden på følgende skala: "Meget godt" (100 point), "Akkurat tilfredsstillende" (50 point) og "Ikke tilfredsstillende" (0 point) med udgangspunkt i rettevejledningen. Der blev opereret med tre niveauer for at gøre det så enkelt som muligt samt for at fremtvinge vurderingerne 100 point og 0 point. Se eksempel på grafisk oversigt i figur 2.

#### 3.3 Indsamling af resultater

##### 3.3.1 Kvantitativ analyse af bedømmere

Med henblik på at udføre nærværende studie blev de studerendes bedømmelser manuelt trukket ud af systemet af Arcanic. Med udgangspunkt i rettevejledningen blev samtlige afleverede besvarelser også bedømt af hjælpelærerne, som var to erfarne kandidatstuderende på samme uddannelse, der tidligere havde taget kurset med godt resultat. De studerendes bedømmelser og hjælpelærernes bedømmelser kunne derefter sammenlignes direkte.



Figur 2: Dele af skærbilledet i peer review-systemet. Øvre panel: tre af de cirka 60 studerende er vist sammen med navnene på dem, som har bedømt, efterfulgt af en tre-delt ring, der viser hvor højt hver bedømmer har bedømt besvarelsen. Nedre panel: Detaljeret oversigt (for den midterste studerende i det øvre panel) over evaluering af besvarelsen af del-spørgsmål. Data fra undersøgelsen, men med ændrede navne og studienumre

### 3.3.2 Kvalitativ feedback fra studerende

Som supplement til den kvantitative undersøgelse af de studerendes bedømmelser af hinandens arbejde – og for at undersøge de studerendes oplevelse af det at benytte peer review i undervisningen – indsamlede underviseren systematisk mundtlig feedback fra de studerende to gange i løbet af kurset. Samtalerne fandt sted i pauser og under grupperegning (det vil sige i øvelsestiden), hvilket betød, at nogle studerende var grupperet tilfældigt, mens andre sad i deres selvvalgte teams. Samtalerne var organiseret som semistrukturerede interviews i grupper på baggrund af en foruddefineret spørgeguide (se Appendiks 11.1).

## 4 Resultater

Kurset havde 54 aktive studerende, som gik til eksamen. Indeholdt i de 54 var en enkelt studerende uden for programmet. Antallet af svar og bedømmelser for opgaven var som følger:

Rettidige afleveringer: 52 (= 95 %). Af disse 52 opnåede en enkelt studerende ikke at blive bedømt (i øvrigt til trods for, at han selv havde lavet tre bedømmelser, men det kunne jo kun den kursusansvarlige se), og to studerende blev hverken bedømt eller bedømte selv. Af de resterende 49 (som er dem, der analyseres i denne artikel), var der 19 studerende, der havde afleveret uden efterfølgende selv at have foretaget bedømmelser. De resterende 30 studerende lavede 1-5 bedømmelser; i gennemsnit blev der foretaget 2,9 bedømmelser per studerende.

### 4.1 Kvantitativ analyse af bedømmere

Der er, som nævnt tidligere, en vis vigtighed i, at selve review-processen forløber med så høj grad af faglig korrekthed som muligt. Ikke kun for at opnå et højt fagligt niveau, men også for at fastholde de studerende i en så seriøs tilgang som muligt. Endvidere er det meget ønskværdigt,

at en studerende, der har lavet noget forkert 1) får det at vide og 2) helst også får at vide, hvad der er rigtigt, eller kraftigt anspores til selv at finde ud af det. På den anden side skal dét, en studerende *har* lavet rigtigt, udløse en form for anerkendelse for at bidrage til, at den studerende føler sig fair behandlet og motiveret.

Det blev derfor undersøgt, i hvor høj grad studenterbedømmelserne var korrekte.

Figur 3 viser en grafisk oversigt over de studerendes performance, både med hensyn til de afleverede besvarelser og bedømmelsesprocessen. Samtlige tre plots har samme rækkefølge for de 49 studerende (den vandrette akse). De to første plot viser, hvordan de 49 studerende blev bedømt ("student som *bedømt*"), mens det sidste plot viser, hvordan de samme studerende har bedømt ("student som *bedømmer*").

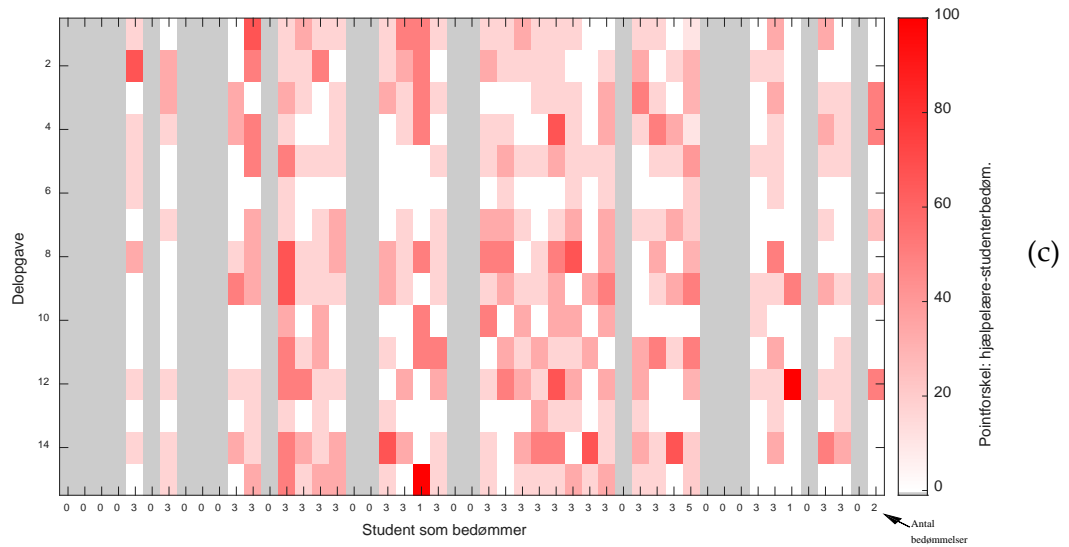
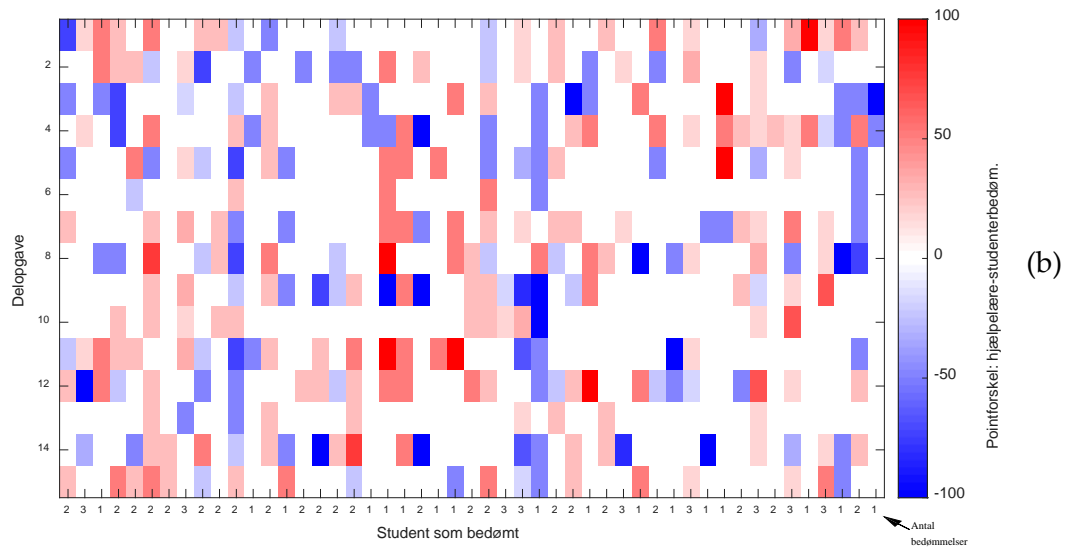
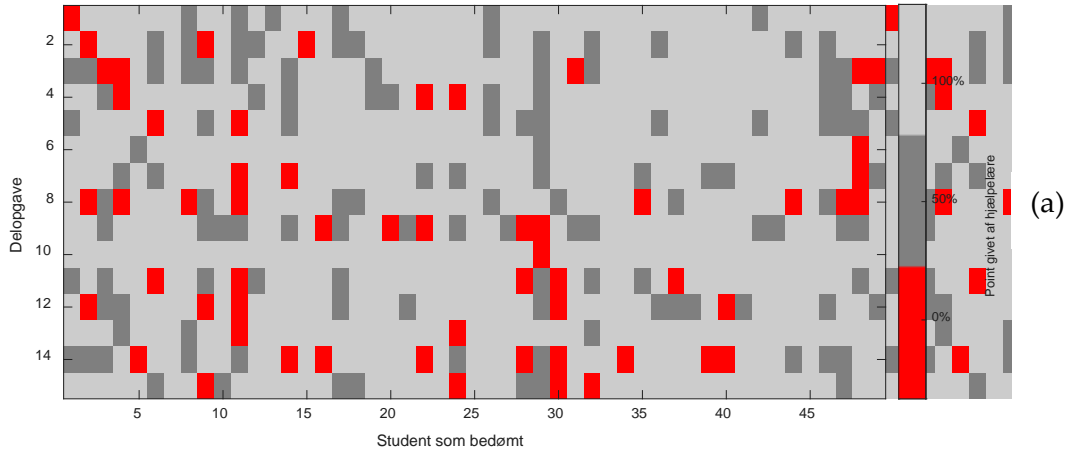
Figur 3a viser hjælpelærernes bedømmelser af de afleverede besvarelser. Der var 15 delspørgsmål, som hver i sær blev bedømt, og der var 49 studerende inkluderet i analysen. Der var således i alt 1549 = 735 delspørgsmålsbesvarelser. Langt hovedparten af disse blev bedømt "meget godt" (100 point) og "akkurat tilfredsstillende" (50 point). Dette kan også ses af den midterste og højre grønne søjle i histogrammet i figur 4 (beskrives senere). Antallet af uacceptable delspørgsmålsbesvarelser udgør resten, det vil sige de 57 røde rektangler i figur 3a svarende til den venstre grønne søjle i figur 4.

Figur 3b viser forskellen mellem, hvordan hjælpelærerne har bedømt, og hvordan besvarelsen blev bedømt af de studerende. Antallet af bedømmelser, der ligger til grund (mellem 1 og 3), fremgår af talrækken nederst, der har erstattet talrække "1-49" fra figur 3a af pladmæssige årsager.

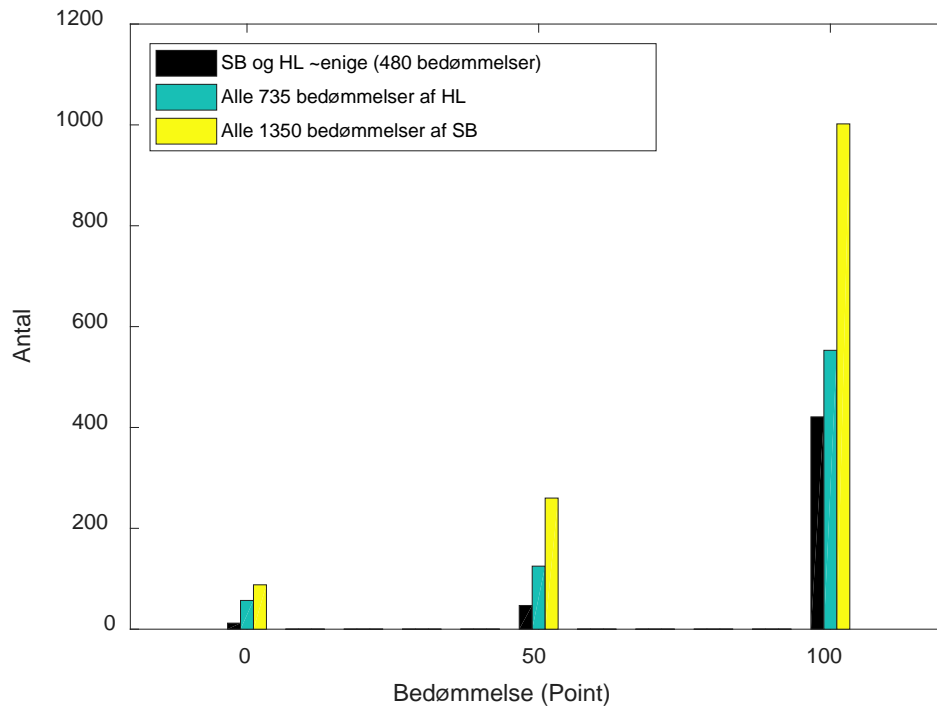
Hvid svarer til enighed. Rød svarer til, at hjælpelærerne har bedømt mere positivt end de midlede studenterbedømmelser (hvis der er mere end en studenterbedømmelse, tages gennemsnittet af dem). Det betyder, at de helt røde felter er situationer, hvor hjælpelærer har givet topkarakter, mens studenterbedømmer (alle) har givet bundkarakter. Dette skete kun 7 gange (under 1 % af delvurderingerne), hvilket også afspejles af de to højre søjler i histogrammet på figur 5b (beskrives senere).

Blå svarer til, at de studerende har bedømt mere positivt end hjælpelærerne. I sin yderste konsekvens betyder det, at de studerende har givet 100 point for noget, som var helt forkert. Det skete i 14 tilfælde (lige under 2 % af delbedømmelserne).

Figur 3c viser, hvordan de studerende klarer sig som bedømmere. De 19 studerende, som ikke nåede at lave en bedømmelse, fremstår med grå farve. Antallet af bedømmelser, hver enkelt studerende har lavet, fremgår af talrækken nederst. Dette plot sammenligner bedømmelserne, som hjælpelærerne og studenterbedømmerne afgav, men da plottet viser gennemsnittet af op til cirka tre bedømmelser, er der regnet med absolutværdier, således at alle værdier er positive i dette plot.



Figur 3: (a) De studerendes performance i selve afleveringen evalueret af hjælpelærerne. (b) De studerendes enighed med hjælpelærerne ved bedømmelse af besvarelsene. (c) De studerendes performance som bedømmere (enighed med hjælpelærerne). Antallet af individuelle bedømmelser, der ligger til grund er vist i linjen nederst for (b) og (c).



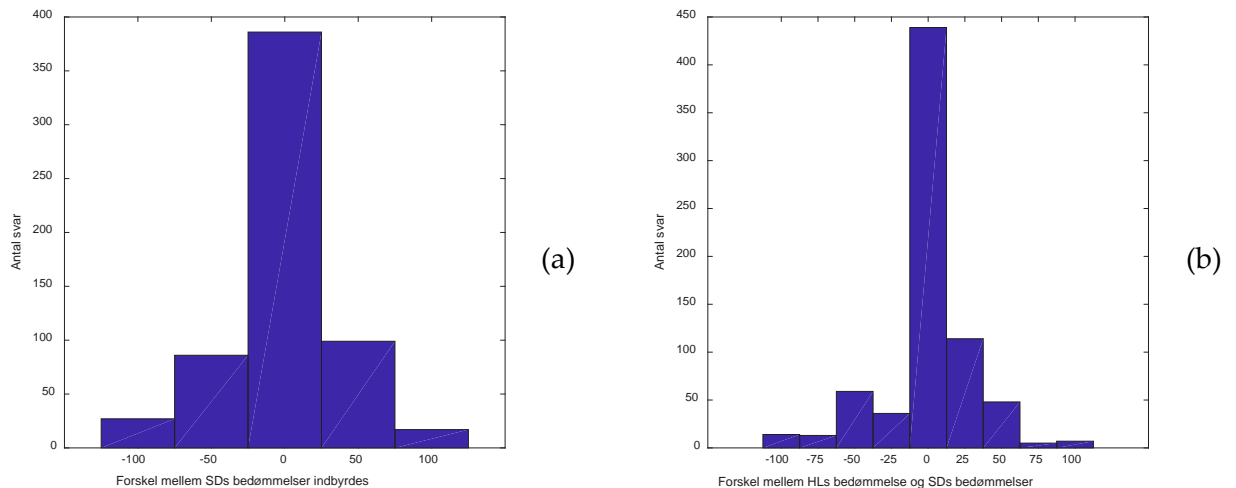
**Figur 4:** Histogrammer for de tre situationer som vist i boksen: Sort: fordelingen af antal bedømmelser hvor studenter- bedømmerne og hjælpelærerne er enige eller næsten enige (forskul  $\leq 17$  point). Grøn: fordelingen af alle 735 bedømmelser foretaget af hjælpelærerne. Gul: fordelingen af alle 1350 bedømmelser foretaget af studenterbedømmerne.

Figur 4 viser fordelingen af bedømmelserne i tre forskellige situationer. De tre grønne søjler angiver fordelingen af samtlige 735 delspørgsmålsbedømmelser foretaget af hjælpelærerne. Her er givet 100 point i 75 % af tilfældene. De tre sorte søjler repræsenterer de 480 tilfælde (delspørgsmål), hvor studenterbedømmerne og hjælpelærerne var enige eller næsten enige. Her er det relative antal af bedømmelsen "100 point" højere, nemlig 88 %. Så studenterbedømmerne og hjælpelærerne var bedre til at være enige, når en opgave blev bedømt til 100 point. En tilsvarende analyse for "0 point" viste, at de ikke er lige så gode til at være enige, når der blev svaret forkert på opgaven. De tre gule søjler i figur 4 repræsenterer samtlige 1350 delspørgsmålsbedømmelser foretaget af studenterbedømmerne. De svarer i store træk til fordelingen for hjælpelærer, omend studenterbedømmerne var mere positive end hjælpelærerne.

Figur 5a viser fordelingen af forskellene i de studerendes bedømmelser, *bedømmerne imellem*, for samtlige bedømmelser. I de fleste tilfælde var der enighed svarende til søjlen i midten, der udgør 63 %, men i 37 % var der en grad af uenighed. I 7 % af tilfældene var der total uenighed.

Figur 5b viser fordelingen af forskelle mellem hjælpelærernes vurdering og de studerendes vurdering. I dette tilfælde blev gennemsnittet af de studerendes bedømmelser udregnet (for op til tre studenterbedømmeres bedømmelser). Her er der 60 % helt enige vurderinger, 80 % helt enige eller næsten enige vurderinger, mens omkring 3 % var totalt uenige (begge sider af histogrammet).





Figur 5: (a) Inter-bedømmer overensstemmelse. (b) Fordelingen af forskelle mellem hjælpelærernes vurdering og de studerendes vurderinger. Søjler ved negative værdier på den vandrette akse angiver, at de studerende er mere positive end hjælpelærerne. HL = hjælpelærerne og SD = studenterbedømmerne.

## 4.2 Kvalitativ feedback fra studerende

De studerende blev interviewet individuelt eller i grupper a op til seks studerende. I alt blev 20 studerende (lige kønsfordeling) interviewet fordelt på seks grupper/individer. Alle svar blev skrevet ned i et organiseret format, og de vigtigste iagttagelser og kommentarer præsenteres nedenfor.

### 4.2.1 Studerendes vurdering af formål, proces og udbytte

Generelt udtrykte de studerende i interviewene tilfredshed med både rammesætningen af peer review-opgaven, den vejledning, der blev givet, selv det at udføre peer review samt selv blive bedømt af medstuderende. På spørgsmålet: "Hvad oplevede du var formålet med peer review-modulet?" svarede de studerende blandt andet, at det var rigtig godt at se andres besvarelser, samt at det blev klart, hvad der var vigtigt via rettevejledningen.

En mindre fremherskende oplevelse blandt de studerende var, at formålet med at benytte peer review var at spare penge ved at spare på brugen af hjælpelærere. To studerende nævnte dog, at det er meningsfuldt at benytte hjælpelærerressourcerne til hjælp i projektarbejdet frem for til at rette opgaver.

Kun én studerende gav udtryk for ikke at have følt sig godt nok klædt på til at skulle bedømme sine medstuderende, og én svarede, at det var "nogenlunde", på trods af at rettevejledningen gav eksempler på facit. Alle de øvrige adspurgte udtrykte, at de følte sig godt klædt på til at skulle bedømme medstuderende. Lidt færre følte sig trygge ved selv at blive bedømt, og den manglende anonymitet nævnes som grund til dette. Det nævnes desuden i fem interviews, at det burde være muligt at give kvalitativ feedback (kommentarer) i forbindelse med review-processen, hvilket systemet ikke understøttede på det tidspunkt.

De studerende blev spurgt, om det var en god idé at gentage peer review-processen i forbindelse med kursets næste opgave, og her svarede hovedparten "ja". Én af de adspurgte studerende udtrykte, at peer review var en dårlig idé, og at vedkommende hellere ville lære fra en ekspert.

Det blev dog bemærket, at det var "bedre end ingenting" at blive bedømt af sine medstuderende.

Én gruppe fremhævede, at selve fordelingen af smileys (point) i bedømmelsesarbejdet blev "gjort med positive briller og med meget løs hånd".

To studerende nævnte, at "man gør sig ekstra umage, når det er ens medstuderende, som retter, i modsætning til, hvis det er underviseren".

#### 4.2.2 Studerendes vurdering af tidsforbrug

Hovedparten af de studerende brugte efter egen vurdering ca. 10-15 minutter pr. review. Kun én nævner at have brugt ca. 30 minutter pr. review, og to studerende har brugt omkring to minutter pr. review.

#### 4.2.3 Anonymitet

I fem interviews blev det fremhævet, at peer review bør foretages anonymt, forstået således at studenterbedømmerne ikke kan se identiteten på den, der bedømmes. En studerende nævnte som begrundelse sin blufærdighed over at afsløre for medstuderende, at vedkommende staver dårligt. Andre mente, at rettelsen blev mindre fordomsfuld, hvis besvareren er anonym.

En enkelt studerende bemærkede, at "de medstuderende i kurset er ens venner", så denne studerende følte sig fuldstændig tryk ved et åbent peer review, men fremhævede, at det nok ville være anderledes på et kursus med mange ukendte deltagere.

#### 4.2.4 Variation i udstyr

De studerende bruger en bred vifte af operativsystemer og browsere, hvilket stiller store krav til selve peer review-systemet. I nærværende kursus brugte cirka halvdelen Microsoft (MS) Windows 10 og Chrome. Apple Mac iOS udgjorde 20 % for Safari og 16% for Chrome. Resten fordelte sig ligeligt med 4 % på hver af MS Windows 7 og Chrome, MS Windows 9 og Chrome samt MS Windows 10 og Firefox.

### 5 Diskussion af resultater

#### 5.1 Kvantitativ analyse af validiteten af studenterbedømmelserne

Ved vurdering af omhyggeligheden i rettelserne gør mindst tre forhold sig gældende: 1) De studerende har bedømt, hvad andre har skrevet, lige efter de selv har – eller skulle have – lært det samme materiale. 2) Én gruppe studerende fremhævede i et af interviewene, at selve tildeelingen af point i rettelserne blev gjort med rund hånd, hvilket antageligt delvist skal ses i lyset af, at de studerende kendte hinanden godt, og at der kun var tre bedømmelsesniveauer. 3) Slutteligt, når en hjælpelærer sidder og bedømmer over 350 delbesvarelser, kan der også ske fejl her. Det bør derfor ikke undre, at hjælpelærere og studenterbedømmere ikke var fuldkommen enige (histogrammerne i figur 5).

Resultaterne er dog i god overensstemmelse med litteraturen: Hamer *et al.* (2009 og 2015) finder for delvist sammenlignelige studerende også, at studenterbedømmerne var mere positive end hjælpelærerne, og der var også en tendens til, at datapunkterne klumper sig sammen omkring  $x = y$  ved høje bedømmelser. Quarrie (2007) har – for en nok mindre sammenlignelig situation – forbløffende god overensstemmelse og bedre end i nærværende undersøgelse. Slutteligt, i

et andet sammenligneligt studie, viser Boase-Jelinek *et al.* (2013) et histogram med samme karakteristika som figur 5.

Som vist i figur 3b er det svært at undgå situationer, hvor studerende, der har lavet noget rigtigt, fejlagtigt får at vide, at det er forkert. Og omvendt. Risikoen for, at det sker, må antages at være størst, hvis der kun er én bedømmer (Cho *et al.*, 2006b). Når der er tre bedømmere, må risikoen for, at alle tre tager fejl, antages at være mindre; det viser sig også i nærværende undersøgelse, idet der i mindre end 1 % af tilfældene blev givet en samlet bedømmelse på 0 point, når hjælpelærerne mente, at der burde være givet 100 point. Det skete lidt oftere, at en studerende blev bedømt for positivt af studenterbedømmerne. Disse fejlratere kan næppe bringes ned til nul. I det færdige system – og i driftssituationen, hvor der ikke foretages bedømmelser af hjælpelærerne – tænkes forkerte bedømmelser fanget på denne måde:

- I de situationer, hvor studenterbedømmerne er indbyrdes uenige, kan disse fanges af underviseren ud fra de grafiske oversigter i figur 2.
- I de situationer, hvor studenterbedømmerne er enige om, at en besvarelse er helt forkert, når den faktisk er helt rigtig, skal de studerende kunne melde tilbage, hvis de føler sig urimeligt behandlet, hvorefter underviseren kan tage affære og give den studerende oprejsning.
- I de situationer, hvor studenterbedømmerne er enige om, at en besvarelse er helt rigtig, selvom den er helt forkert, bør man sikre sig, at de studerende også selv tjekker deres besvarelse med facit (hvis de har foretaget korrekte bedømmelser af medstuderende, er de i hvert tilfælde bekendt med facit; hvis de ikke har foretaget korrekt bedømmelse, bør underviseren skride ind. Korrekt bedømmelse kan her kun estimeres ud fra, om den pågældende studerende er enig med sine medbedømmere).

Endelig kan der tænkes (meget få) situationer relateret til de to sidstnævnte punkter, hvor den studerende ikke er klar over, at han/hun er forkert bedømt. Her er der næppe meget andet at gøre end at håbe på, at den studerendes kritiske sans alligevel vil lede denne på rette spor.

Slutteligt skal det nævnes, at typen af opgaven også ser ud til at have betydning for resultaterne: Det ses af figur 3b, at der er relativt få forskelle mellem hjælpelærernes og studenterbedømmernes bedømmelser for delspørgsmål 6, 10 og 13, som var de mest konkrete spørgsmål (for eksempel et tal med enhed, som direkte fremgår af facit).

## 5.2 Den indbyggede vilje til at deltage i peer review

Ud fra de 30 studerende som vist i figur 3c, der frivilligt bedømte, er den relative deltagelse  $30/49 = 61\%$ . Frivilligheden taget i betragtning er det en pæn tilslutning til ideen om at bedømme hinanden. Hvis man dertil lægger den langt overvejende positive holdning, de studerende giver udtryk for i interviewene, vurderes det, at langt de fleste af dem, som ikke bedømte, vil kunne finde tid til dette, når det bliver obligatorisk at deltage. Lignende resultater når Nicol *et al.* (2014) frem til i et studie, der involverer tekstuel feedback.

## 5.3 Faktorer, der påvirker studerendes holdning til peer review

De studerendes motivation og positive holdning til peer review var opmuntrende, og de virkede indforståede med, at med begrænsede ressourcer er peer review en god ide. Der var i lighed med fund i andre studier (Andersen, 2016; Harland *et al.*, 2016) blot enkelte studerende, som

enten ikke mente, at de andre studerende var fagligt klædt på til at bedømme eller ikke selv følte sig kvalificeret til at bedømme. Den positive holdning skal også ses i lyset af, at ikke alt virkede ved peer review-systemet i første omgang, men underviserens entusiastiske rammesætning har muligvis trukket i modsatte retning.

De studerende på uddannelsen i Medicin og Teknologi følger samme hold og bliver derfor fortrolige med hinanden i løbet af deres studietid. Dette kan have indflydelse på deres positive holdning til at indgå i peer review.

På det hold, der deltog i forsøget, var der ikke den store modstand mod, at navne på bedømmer og bedømt var kendt; det vil muligvis være anderledes på et kursus, hvor de studerende ikke kender hinanden. Med hensyn til de studerendes indsats ved opgavebesvarelsen er det muligt, at de studerende, hvis alle navne er kendt, vil føle, at de skal stramme sig an for ikke at miste anseelse. Hvis processen yderligere forløber i en positiv ånd, kan den efterfølgende kontakt mellem de studerende være yderst inspirerende (Boase-Jelinek et al., 2013), men det er nok vigtigt at have de studerende med i beslutningen om anonymitet.

#### **5.4 Estimat af reduktion af tidsforbrug**

Et forsigtigt skøn: Hvis det tager cirka 30 minutter at rette hver af de 49 besvarelser ved den traditionelle fremgangsmåde (uden peer review), og det tager cirka 5 timer at udrede sager og rehabiliter studerende, der er blevet uretfærdigt bedømt, udgør besparelsen cirka 20 timer svarende til en reduktion på cirka 80 %.

### **6 Diskussion af metode**

#### **6.1 Styrker**

Først og fremmest blev peer review afprøvet på et relativt svært, obligatorisk bachelorkursus med et effektivt, systematisk og computeriseret system, der opfyldte alle kravene stillet i indledningen.

Opgaven, der var genstand for peer review, var realistisk og vel indarbejdet i kurset. Det var en type, hvor der var mere end ét korrekt svar (som ofte skulle indeholde en grafisk fremstilling), sværhedsgraden var relativt høj, og den studerende skulle ikke kun bruge det, han/hun har lært, men også sin sunde fornuft. Det vurderes derfor, at undersøgelsen, har en god grad af realisme og derfor ikke er kunstigt positiv over for brugen af peer review.

#### **6.2 Svagheder**

Deltagelse var frivillig, så selvom alle havde afleveret, var der en mindre gruppe, som ikke deltog i bedømmelsesprocessen.

De studerendes egen vurdering af tid brugt på bedømmelse af opgaverne skal tages med forbehold, idet de ikke på forhånd blev bedt om at registrere tidsforbruget.

Udregning af underviserens tid er et estimat. Det er muligt, at inddragelse af de resterende studerende – som muligvis ikke er helt lige så begejstrede for at bedømme – ville have givet et mere broget billede, hvor forholdsmæssigt flere studerende blev forkert bedømt.

Hjælpelærernes rettelser af besvarelserne er ikke nødvendigvis altid korrekte, og de to hjælpelærere bedømte ikke nødvendigvis helt ens; de bedømte cirka halvdelen af opgaverne hver.

Underviser foretog interviewene, så en mindre risiko for bias var til stede.

Slutteligt indeholdt dette studie ikke afslutningsvis feedback fra de studerende om hele bedømmelsesprocessen. Det må formodes, at en sådan feedback kan give yderligere værdifuld viden om procesforløbet.

## 7 Konklusion

Resultaterne tillader besvarelse af de fire spørgsmål rejst i indledningen:

Validiteten af de studerendes bedømmelser er undersøgt, og kun i nogle få procent af tilfældene er der total uenighed mellem hjælpelærerne og studenterbedømmerne; underviseren behøver således ikke rette besvarelserne. Fejlbedømmelser kan næppe reduceres til nul, men med et passende tilbagemeldingssystem kan studerende gøre opmærksom på, at de føler sig forkert bedømt. Det kan endvidere konkluderes ud fra deltagerantal og svar i interviews, at det virker, som om der er en meget stor lyst til at deltage i peer review. Tilsvarende giver de studerende udtryk for en meget positiv holdning til peer review-processen umiddelbart efter, at den har fundet sted. Endelig er tidsbesvarelsen i forbindelse med udarbejdelse af feedback til de studerende estimeret til omkring 80 % i det aktuelle tilfælde for underviseren.

## 8 Perspektiver

### 8.1 Overvejelser om brug

Ifølge dansk lovgivning kan bedømmelser lavet af medstuderende ikke indgå i den endelige summative bedømmelse af den studerende, altså den endelige kursuskarakter (hvilket vores studerende og andre (Patton, 2012) påskønner). På den anden side vil frivillighed med hensyn til at aflevere og bedømme meget ofte ende med, at kun en del af de studerende deltager. Derfor tænkes deltagelse i peer review at indgå binært: Man skal have sin deltagelse godkendt for at kunne gå til eksamen, men det påvirker ikke den endelige karakter i kurset.

I fremtidige anvendelser af peer review kunne et af følgende kvantitative mål i den ordinære kursusafvikling overvejes:

1. Den studerende har afleveret en besvarelse til bedømmelse
2. Den studerende har *yderligere* bedømt for eksempel tre andre besvarelser
3. Den studerende har *yderligere* forfattet et kvalificeret feedback til underviseren om de modtagne bedømmelser

Disse mål tænkes både at have et pædagogisk sigte (det er lærerigt for de studerende at lave) og et standpunktsmæssigt sigte (den studerende får en ide om, hvor den står fagligt, og den kursusansvarlige kan med en moderat indsats orientere sig i forløbet for en stor gruppe studerende og dermed på et overordnet niveau vurdere de studerendes læring). En anden mulighed er, at den studerende bedømmes på den feedback, vedkommende giver sin(e) medstuderende. Boase-Jelinek et al. (2013) rapporterer gode erfaringer med denne fremgangsmåde, som dog kræver, at arbejdsmængden for underviser balanceres med fordelene ved at have flere studenterbedømmelser af samme opgave.

En rekursiv fremgangsmåde kunne også anvendes, hvor de studerende skal genaflevere besvarelsen efter første review (Andersen, 2016; Nicol et al, 2014; Sims, 1989), men det kræver et langstrakt kursusforløb for at kunne nå det hele, og det betyder også, at der vil blive brugt rela-

tivt lang tid på en sådan opgave. I en sådan fremgangsmåde er der endvidere risiko for, at de studerende "låner" (for) meget af hinanden (når de læser andres besvarelser), men det kan dog afsløres automatisk via Urkund (Urkund, 2017), omend det altid er svært for kursusansvarlige at trække grænsen mellem inspiration og direkte plagiat (Sims, 1989).

På grund af kompleksiteten og de indgående *threshold concepts* (Land et al., 2005) i opgaven burde de studerende kunne tilføje kommentarer til den kvalitative bedømmelse af hinandens arbejde, hvilket som nævnt også kom til udtryk i de studerendes feedback til underviser. Man kan argumentere for, at de studerende, som er på samme (novice-) niveau i forhold til det, der skal læres, vil have bedre forudsætninger for at forstå, *hvad* der er vanskeligt i opgaven, og *hvordan* det er vanskeligt, og dermed i deres kommentarer vil kunne formulere sig i et andet – og for medstuderende mere forståeligt – sprog end underviseren, som har ekspertviden (Land et al., 2005, Meyer & Land, 2003).

## 9 Anerkendelser

Arvid Klein Gregersen, Arcanic A/S, for udtræk af data for de studerendes vurdering i peer review-processen. Undervisningsassistenter Line Marcussen og Anders Emil Høgh Jakobsen for rettelse af aflevering A2. De studerende, der deltog i dette studie. De anonyme fagfællebedømmere og Birgitte Lund Christiansen, LearningLab DTU, for konstruktive kommentarer samt Nina Kjærgaard, DTU Elektro, for korrektur.

**Jens E. Wilhjelm** er professor i medicinsk billeddannelse og medikoteknisk uddannelse på DTU. Han har en PhD fra Worcester Polytechnic Institute, Massachusetts, USA, i medicinsk ultralyd. Han er kandidatstudeleder for den medikotekniske uddannelse Medicin og Teknologi og medstifter af samme. Inden for undervisning arbejder Jens blandt andet med peer review og animeret interaktiv grafik for visualisering af interaktioner i biologisk væv i medicinsk billeddannelse. Han er desuden underviser i medicinsk billeddannelse og medicinsk instrumentering.

**Sidsel-Marie Winther Prag** er cand.it. i Design, Kommunikation og Medier fra IT-Universitetet i København med speciale i konceptudvikling, læring og it. Hun har arbejdet som projektleder, konceptudvikler og rådgiver inden for feltet læring og it i både den private og offentlige sektor siden 2007. Siden 2012 har Sidsel-Marie været ansat som e-læringskonsulent i LearningLab DTU på Danmarks Tekniske Universitet (DTU). Som e-læringskonsulent fungerer Sidsel-Marie som underviser, rådgiver og sparringspartner inden for universitetspædagogik og brug af læringsteknologi med fokus på den løbende kvalitetsudvikling af uddannelser, undervisning og læring på DTU.

## 10 Referencer

- Althaus, R. & Darnall, K. (2001): *Enhancing critical reading and writing through peer reviews: an exploration of assisted performance*. Teaching Sociology, 29, pp. 23-35.
- Andersen, L. B. (2016): Styrket feedback gennem studerendes selvevaluering. *Dansk Universitetspædagogisk Tidsskrift*, 21, pp. 9-20.
- Boud, D. & Molloy E. (2013): Rethinking models of feedback for learning: the challenge of design, *Assessment & Evaluation in Higher Education*, 38 (6), 698-712, DOI: 10.1080/02602938.2012.691462.

- Boase-Jelinek, D.; Parker, J.; Herrington, J. (2013): Student reflection and learning through peer reviews. *Issues in Educational Research*, 23 (2), pp. 119-131.
- Cho, K. & MacArthur, C. (2011): Learning by Reviewing. *Journal of Educational Psychology*, 103 (1), pp. 73-84.
- Cho, Y. H. & Kwangsu Cho, K. (2011): Peer reviewers learn from giving comments. *Instr. Sci.* 39, pp. 629–643. DOI 10.1007/s11251-010-9146-1.
- Cho, K., Schunn, C. D., Charney, D. (2006): Commenting on writing - Typology and perceived helpfulness of comments from novice peer reviewers and subject matter experts. *Written Communication*. 23 (3), pp. 260-294.
- Cho, K., Schunn, C. D. & Wilson, R. W. (2006): Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology*, 98 (4), 891–901. doi: 10.1037/0022-0663.98.4.891.
- Evans, C. (2015): Students' Perspectives on the Role of Peer Feedback in Supporting Learning. *Journal of Cognitive Education and Psychology*, 14 (1), pp. 110-125.
- Hamer, J., Purchase, H., Denny, P. & Luxton-Reilly, A. (2009): *Quality of Peer Assessment in CS1*. Fifth International Computing Education Research Workshop, Icer, pp. 27-35.
- Hamer, J., Purchase, H., Luxton-Reilly, A., Denny, P. (2015): A comparison of peer and tutor feedback. *Assessment and Evaluation in Higher Education*. *Open table of contents*, 40 (1), pp. 151-164.
- Harland, T., Wald N. & Randhawa H. (2016): Student peer review: enhancing formative feedback with a rebuttal, *Assessment & Evaluation in Higher Education*, DOI: 10.1080/02602938.2016.1194368.
- Howard, C. D., Barrett, A. F., Frick, T. W. (2010): Anonymity to promote peer feedback: pre-service teachers' comments in asynchronous computer-mediated communication. *J. Educational Computing Research*, 43 (1), 89-112.
- Land R., Cousin G., Meyer J. H. F. & Davies P. (2005): *Threshold concepts and troublesome knowledge (3): implications for course design and evaluation*. In Rust, C (ed) *Improving Student Learning Diversity and Inclusivity*. Oxford: Oxford Centre for Staff and Learning Development.
- Li, L., Liu, X. & Steckelberg A. L. (2010): Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology*, 41 (3), pp. 525–536. doi:10.1111/j.1467-8535.2009.00968.x.
- Meyer, J. & Land, R.(2003): *Threshold Concepts and Troublesome Knowledge: Linkages to Ways of Thinking and Practising within the Disciplines*. Occasional Report 4. ETL Project, Universities of Edinburgh, Coventry and Durham. www.ed.ac.uk/etl.
- David Nicol, Avril Thomson & Caroline Breslin (2014) Rethinking feedback practices in higher education: a peer review perspective, *Assessment & Evaluation in Higher Education*, 39 (1), 102-122, DOI: 10.1080/02602938.2013.795518.
- Patton, C. (2012) 'Some kind of weird, evil experiment': student perceptions of peer assessment, *Assessment & Evaluation in Higher Education*, 37 (6), 719-731, DOI: 10.1080/02602938.2011.563281.

- Panadero, E., Jonsson, A. (2013): *The use of scoring rubrics for formative assessment purposes revisited: A review*. Educational Research Review 9, pp. 129–144.
- Quarrie S.P. (2007): Student Peer Review as a Tool for Efficiently Achieving Subject-Specific and Generic Learning Outcomes: Examples in Botany at the Faculty of Agriculture, University of Belgrade. *Higher Education in Europe*, 32 (2/3).
- Sims G. K. (1989): Student peer review in the classroom: A teaching and grading tool. *J. Agron. Educ.* 18, 105-108.
- Ulriksen, L. (2014): *God undervisning på de videregående uddannelser*. Frederiksberg: Frydenlund.
- Urkund (2017). [www.urkund.com](http://www.urkund.com).
- Wilhjelm, J. E., Pihl, M., Lonsdale, M. N. & Jensen, M. (2008): An active learning approach to the physics of medical imaging. *Medical Engineering and physics*. 30 (5), pp. 607-614.
- Wilhjelm, J. E. (2017): Fra [bme.elektro.dtu.dk/31540/pdf/pixels\\_resolution.pdf](http://bme.elektro.dtu.dk/31540/pdf/pixels_resolution.pdf).
- Wilhjelm, J. E., Marcussen L. & Jakobsen A. (2016): *Official answers to Assignment 2*. Fra [bme.elektro.dtu.dk/31540/pdf/31540\\_official\\_answers\\_a2.pdf](http://bme.elektro.dtu.dk/31540/pdf/31540_official_answers_a2.pdf).
- Kursushjemmeside (2017). Fra [31540.kurser.dtu.dk/course/31540](http://31540.kurser.dtu.dk/course/31540) og [bme.elektro.dtu.dk/31540](http://bme.elektro.dtu.dk/31540).



## 11 Appendix

### 11.1 Spørgeguide

- Hvordan var den overordnede proces?
  - Hvad syntes du om selve rammesætningen/vejledningen?
  - Hvad oplevede du var formålet med peer review?
  - Følte du dig godt nok klædt på til at skulle reviewe dine medstuderende?
  - Var du tryk ved selv at blive reviewet?
  - Hvilken information ville du evt. ønske at du havde haft?
- Hvordan var selve review-processen:
  - Teknisk
  - Fagligt
  - Psykisk
- For dem, der lavede mindre end 3 reviews: Hvorfor?
- Hvordan fungerede facitlisten og rettearket (rubric):
  - Hvad fungerede godt?
  - Hvad fungerede dårligt?
- Savnede du at kunne give skriftlige kommentarer i et kommentarfelt?
- Besvarelsen var i en pdf-fil. Burde der kunne kommenteres i denne pdf-fil?
- Hvor lang tid brugte du på selve reviewsne?
- Kommentarer til anonymitet?
- Hvilket operativsystem og hvilken browser brugte du?

### 11.2 Anbefalinger

Ud fra dette studie, forfatternes erfaringer og litteraturen er der en række forudsætninger for et positivt resultat og effektiv brug af ressourcer (tid), når peer review implementeres:

For selve processen:

- Rammesætning af peer review-processen, herunder tydeliggjorte forventninger til de studerende, drøftelse af retteark/rubric med de studerende og måske ligefrem forudgående træning i bedømmelse af andres arbejde, er vigtige aspekter for en positiv holdning blandt de studerende og for deres udbytte (Andersen, 2016).
- Feedback skal være informativ, præcis, skrevet i en positiv ånd og skulle gerne være motiverende (Boase-Jelinek et al., 2013). Det betyder også, at kritik skal begrundes, så modtageren kan læse, *hvad* der er galt.
- Den kursusansvarlige skal nøje overveje, hvorvidt processen skal være anonym, jævnfør diskussionsafsnittet (Boase-Jelinek et al., 2013; Howard et al., 2010; Sims, 1989).

- Hvis anonym: De studerende bør informeres om, at de ikke – selv i et velfungerende system – kan forvente, at alt altid er 100 % anonymt (for eksempel grundet egne fejl), og at de altid skal kunne stå på mål for de bedømmelser, de giver.

Selve det server-baserede peer review-system med WEB-interface bør:

- Tjekke, hvor muligt, at en studerende ikke utilsigtet kommer til at røbe sin identitet (via filnavnet, metadata eller teksten) i en anonym proces.
- Udføre plagiattjek på alt afleveret af de studerende.
- Kunne liste alle, der har afleveret den oprindelige besvarelse (og hvem der *ikke* har), samt grafiske oversigter over bedømmelsesprocessen.
- Tillade, at man styrer timingen i processen, så man i nogle tilfælde kan vente på, at et tilstrækkeligt antal studerende har afleveret den oprindelige besvarelse.
- Tillade tekst-input (enten generelt eller for hvert element i rubricen), så den studerende kan give skriftlige kommentarer.
- Tillade, at den studerende giver feedback til kursusansvarlige på bedømmelsen (for eksempel så en studerende kan gøre opmærksom på, at hun/han mener at være blevet forkert bedømt).