

Anne Skorkjær Binderkrantz

Er kvinder dårligere undervisere end mænd – ifølge de studerende?

En række internationale studier har vist, at kvindelige undervisere får dårligere evalueringer end deres mandlige kolleger. Artiklen undersøger, om det også er tilfældet i en dansk kontekst. Artiklens analyser bygger på evalueringer foretaget på Aarhus BSS, hvor et fælles evalueringsskema bruges på tværs af uddannelser, der varierer ganske meget i forhold til kønssammensætningen af både studerende og undervisere. Artiklen inddrager næsten 30.000 evalueringer af mere end 1.000 forskellige undervisere. På den baggrund vises, at der snarere end en helt generel bias i evalueringer ser ud til at være en sammenhæng mellem sammensætningen af undervisergruppen, den studerendes køn og vurderingen af undervisere af forskelligt køn.

Evalueringer af undervisningen bruges overalt i uddannelsessystemet. På universiteterne – som er i centrum for denne artikel – findes næppe et undervisningsforløb, hvor de studerende ikke bliver bedt om at udfylde et evalueringsskema. Traditionelt har hovedformålet med evaluering været, at den enkelte underviser får de studerendes vurdering af undervisningsforløbet og forslag til forbedringer – og dermed et udgangspunkt for at forbedre fremtidig undervisning. Evalueringer bruges imidlertid også til andre formål, og der er især et stigende fokus på brugen af evalueringer af undervisningsforløb til generel kvalitetssikring af uddannelser. De enkelte universiteter har allerede i vidt omfang integreret evalueringresultater i deres kvalitetssikringssystemer, og senest hedder det i Regeringens udspil til fremtidens universiteter, at ”Universiteterne skal mere systematisk bruge studerendes evalueringer af kurser og undervisere til at styrke uddannelserne” (Regeringen, 2018). Endelig bruges evalueringer også i beslutninger om ansættelse og forfremmelse på universiteterne, hvor der mange steder stilles krav om, at der ved jobansøgninger vedlægges evalueringer af undervisningsforløb.

Meget tyder imidlertid på, at evalueringer ikke giver et retvisende billede af kvaliteten i undervisningen (Dalsgaard et al., 2017). Evalueringerne ser fx ud til at blive påvirket af de studerendes forventning til, hvilken karakter de vil få, ligesom undervisere, der stiller store krav til studerende, risikerer at få dårlige evalueringer. Forskellige studier har også påpeget, at der ikke er en sammenhæng mellem de studerendes læringsudbytte og evalueringerne (Braga, Paccagnella og Pellizzari, 2014; Uttl, White og Gonzales, 2017). Endelig viser

en række nyere undersøgelser, at kvindelige undervisere systematisk evalueres dårligere end mandlige undervisere (Boring, Ottoboni og Stark, 2016; Mengel, Sauermann og Zölitz, 2017). Er der en bias i de studerendes evalueringer, stiller det spørgsmålstegn ved anvendelsen af studenterevalueringer til kvalitetssikring af uddannelser og rekruttering af universitetsansatte. Denne artikel undersøger derfor, om en sådan forskel i evalueringen af mandlige og kvindelige undervisere kan findes på et dansk universitet.

En mulig bias i de studerendes evalueringer er også interessant i lyset af den generelle underrepræsentation af kvinder på de danske universiteter især i de højeste stillingskategorier. Blandt professorer på de danske universiteter er således kun 17 pct. kvinder, mens kvinderne udgør omkring 30 pct. af samtlige fastansatte i underviserstillinger på universiteterne (Nielsen, 2017a). Evalueringerne kan spille ind på rekrutteringen af kvinder på to måder. For det første indgår evalueringerne i beslutninger om ansættelse og forfremmelse. For det andet kan evalueringerne påvirke de enkelte underviseres beslutning om at blive i universitetsverdenen. En af årsagerne til de manglende kvindelige ansatte er således, at flere kvinder end mænd fravælger at søge videre ansættelse efter de første år på universitetet (Nielsen, 2017b). Hvis kvinder systematisk evalueres dårligere end deres mandlige kolleger, kan det derfor være en medvirkende årsag til den manglende rekruttering af kvinder i de højeste stillingskategorier – især i en situation, hvor det fra alle niveauer understreges, at netop gode evalueringer er en central indikator for kvalitet i undervisningen.

Der findes flere bud på, hvad der kan ligge bag en kønsforskel i evalueringer. En del forskning peger på, at forskellen skyldes stereotyper om mænd og kvinder, og at kvinder får dårligere evalueringer, fordi de ikke lever op til de studerendes forventning om, hvordan en god underviser er. En oplagt – men måske kontroversiel – modtese er, at kvinder ganske enkelt evalueres dårligere end mænd, fordi de er dårligere undervisere. Et tredje argument er, at der er variation i den undervisningsstil, som mænd og kvinder har størst tilbøjelighed til at bruge, og at de studerende generelt foretrækker de mandlige underviseres undervisningsstil. Endelig kan der peges på betydningen af den lokale kontekst – fx i form af meget mandsdominerede universitetsmiljøer. Artiklen diskuterer disse forskellige argumenter for en eventuel forskel i evalueringen af mænd og kvinder. Det vil dog være op til efterfølgende studier at komme tættere på, hvilke mekanismer der ligger bag en eventuel forskel mellem evaluering af mandlige og kvindelige undervisere.

Empirisk undersøger artiklen, om der er forskel på de studerendes evalueringer af mandlige og kvindelige undervisere på et dansk universitet. Konkret analyseres studerendes evalueringer af undervisningen på Aarhus BSS i 2016

og 2017. Det giver mulighed for at inddrage uddannelser med betydelig variation i kønssammensætningen især i studentergruppen, hvor andelen af kvinder blandt nyoptagne studerende spænder mellem mindre end en fjerdedel til mere end 80 pct. Samtidig er der også i hvert fald en vis variation i undervisergruppen. Selvom der er en klar overvægt af mænd blandt det fastansatte videnskabelige personale på de fleste uddannelser, er der også på Aarhus BSS uddannelser, hvor de studerende møder flest kvindelige undervisere. Undersøgelsen baseres på i alt 39.546 udfyldte evalueringsskemaer, der evaluerer undervisning foretaget af 1.176 forskellige undervisere.

En bias i evalueringer – den eksisterende litteratur

Spørgsmålet om, hvorvidt mænd og kvinder evalueres forskelligt, har været på forskningsdagsordenen i USA i flere årtier men er først for nylig blevet undersøgt også af europæiske forskere. Den tidlige amerikanske forskning kom til modstridende konklusioner. Nogle studier fandt ganske små eller ingen forskelle i de studerendes evalueringer af mænd og kvinder, og andre studier fandt forskelle i enten kvindernes eller mændenes favør (Centra og Gaubatz, 2000: 18; Feldman, 1993; 1992; Wright and Jenkins-Guarnieri, 2011). I den nyere litteratur er man nået frem til mere konsistente resultater, der tyder på, at der er en kønsbias i studenterevalueringer. Især fire studier er interessante: to amerikanske, der bruger forskellige typer eksperimenter, og to europæiske, der analyserer faktiske evalueringer.

Det første studie fokuserede på, om en undervisers køn og alder påvirker de studerendes vurdering af en præsentation i forhold til, om underviseren fx opleves som entusiastisk, interesseret og selvsikker. Forskerne lod 352 psykologistuderende høre en 35 minutter lang præsentation med en stemme, der på forhånd var verificeret som kønsneutralt opfattet, imens de blev vist en slide med en computergenereret profil af en professor, der hverken kunne identificeres som mand eller kvinde. Efterfølgende fik de studerende et spørgeskema med forskellige oplysninger om, hvorvidt forelæseren var mand eller kvinde, henholdsvis ung eller gammel. Selvom de havde hørt den samme forelæsning, vurderede studerende, der troede, at forelæseren var mand, en række aspekter – entusiasme, interesse og stemme – bedre, end dem der troede, der var tale om en kvinde, og især den gruppe, der troede, der var tale om en yngre mand, var positive i deres vurdering (Arbuckle og Williams, 2003).¹

Et andet eksperimentelt studie fra USA bygger også på, at de studerende evaluerer uden at kende underviserens sande køn men tager udgangspunkt i faktiske undervisningsforløb og evalueringer. Studiet inddrager fire parallelle online undervisningsforløb med i alt 43 studerende. De studerende mødte ikke

på noget tidspunkt underviseren fysisk eller virtuelt. To af de fire forløb blev undervist af en kvinde og to af en mand, men de studerende på to af holdene fik oplyst, at underviseren havde det modsatte køn af det korrekte. Efter forløbene udfyldte de studerende evalueringer, og studiet viste, at de studerendes evaluering fulgte det køn, de troede underviseren havde. De studerende, der havde fået oplyst, at de havde en kvindelig underviser, gav således dårligere evalueringer – på en samlet evalueringsscore på baggrund af 12 spørgsmål var forskellen på 0,6 på en skala med fem point – end dem, der troede at deres underviser var mand (Macnell, Driscoll og Hunt, 2015).

Der findes – i hvert fald så vidt vides – indtil nu kun to europæiske undersøgelser af, hvorvidt underviserens køn påvirker de studerendes evaluering af et undervisningsforløb. Begge undersøgelser analyserer faktiske studenterevalueringer. Mengel, Sauermann og Zölitz (2017) inddrager næsten 20.000 evalueringer foretaget af studerende på en hollandsk handelshøjskole, mens Boring (2017) analyserer 23.000 evalueringer fra et fransk universitet. Begge studier fokuserer på undervisningsforløb, hvor de studerende tilfældigt blev fordelt på hold med henholdsvis mandlige og kvindelige undervisere. Samtidig inddrager studierne også de studerendes karakterer i de evaluerede fag for at kunne undersøge, om der er forskel på, hvor effektive mænd og kvinder er som undervisere – i hvert fald i forhold til at påvirke de studerendes karakterer.

De to studier konkluderer begge, at kvinder evalueres lavere end mænd. I den hollandske undersøgelse er der ingen sammenhæng mellem underviserens køn og de studerendes karakterer, men der er en sammenhæng med de studerendes evaluering af underviseren. Både mandlige og kvindelige studerende evaluerer de kvindelige undervisere dårligst, men forskellen er størst for de mandlige studerende, der i deres generelle vurdering af underviseren giver mandlige undervisere en 0,2 point højere score på en skala med fem point (Mengel, Sauermann og Zölitz, 2017: 19). Det franske studie finder nogenlunde tilsvarende resultater. Heller ikke her ser der ud til at være en sammenhæng mellem underviserens køn og den karakter, de studerende opnår til eksamen, og også her evaluerer mandlige studerende deres mandlige undervisere højere end de kvindelige. Konkret peger Boring på, at sandsynligheden for, at mandlige professorer vurderes som ”excellente” af deres mandlige studerende, er mellem 33 og 41 pct. højere end for de kvindelige professorer (Boring, 2017: 32).

Samlet set rejser disse studier tvivl om, i hvilket omfang studerendes evalueringer er en god indikator for undervisningens kvalitet eller snarere afspejler fx forskellige forventninger til mandlige eller kvindelige undervisere. Der er imidlertid også grund til at tro, at konteksten kan have betydning for, hvorvidt og i hvilket omfang der er en bias i studenterevalueringer. Selvom fx stereotype

opfattelser af mænd og kvinder er relativt ens i en nutidig vestlig kontekst, hænger stereotyper også sammen med sociale og strukturelle forhold, som kan variere mellem lande og på tværs af forskellige kulturer (Rudman og Glick, 2001: 744-745). En undersøgelse af evalueringer i en dansk sammenhæng kan derfor bidrage både til den danske debat om brugen af evalueringer og til den generelle forståelse af den mulige kønsbias i de studerendes svar på evalueringer.

Baggrunden for en mulig forskel – stereotyper eller undervisningsstil?

Når det kommer til at forklare, at der er forskel i studerendes evaluering af mænd og kvinder, argumenteres ofte for, at de studerendes evalueringer reflekterer stereotyper om mænd og kvinder. Ifølge teorien om såkaldt ”rolleinkongruens” kolliderer de generelle forventninger til kvindelig adfærd med forventningerne til universitetsundervisere (Feldman, 1993: 179; Laube et al., 2007). Studerende har således både forestillinger om, hvordan mænd og kvinder bør være, og om, hvordan en professor bør være. Studier har vist, at der er større overensstemmelse mellem de træk, som opfattes som attraktive for mænd og professorer, end mellem attraktive træk for kvinder og professorer (Burns-Glover og Veith, 1995; Jaschik, 2015; Sprague og Massoni, 2005).

Når vi skal foretage en bedømmelse – som tilfældet er, når de studerende bliver bedt om at evaluere en underviser – gør vi det i forhold til et referencepunkt, og her er det centralt, at mænd og kvinder bliver vurderet ud fra forskellige referencepunkter (Arbuckle og Williams, 2003; Laube et al., 2007: 92-94; Sinclair og Kunda, 2000). Opfører et individ sig anderledes end forventet, opstår rolleinkongruens, og det kan påvirke opfattelsen af individet negativt. Konkret kan undervisningssituationen lægge op til, at en kvinde optræder assertivt og med autoritet, og da det er træk, som oftest forbindes med mænd, kan det føre til rolleinkongruens og negative reaktioner i form af dårlige(re) evalueringer (Martin, 2016: 315; Rudman og Glick, 2001).

Sammenhængen mellem underviserkøn og de studerendes evalueringer kan imidlertid også skyldes forskelle i den måde, mænd og kvinder er mest tilbøjelige til at undervise på – eller i hvor effektive mænd og kvinder i gennemsnit er som undervisere – selvom der naturligvis også vil være variation inden for hver gruppe (Feldman, 1993:180). Selvom de nævnte studier af Boring og Mengel, Sauermann og Zölitz i et vist omfang kontrollerer for, om mænd er bedre undervisere end kvinder, er det ikke afklaret, om sammenhængen kan skyldes forskelle i undervisningsstil. Det er fx påvist, at mænd oftere benytter sig af forelæsning, mens kvinder er mere tilbøjelige til at anvende klasses Diskussioner

(Centra og Gaubatz, 2000: 29; Laird, Garver og Niskode-Disset, 2011; Stes, Donche og Petegem, 2004). Det er også oplagt, at forskelle i, hvor kompetente mænd og kvinder oplever sig selv som undervisere, kan påvirke deres adfærd i klasselokalet (Bandura, 1997). I det omfang studerende foretrækker nogle undervisningsstile frem for andre, kan det derfor være en medvirkende årsag til, at mandlige og kvindelige undervisere vurderes forskelligt af de studerende. Det gælder uanset, om de forskellige undervisningsstile reelt også har en betydning for, hvor meget de studerende lærer.

Samtidig kan den lokale kontekst også have betydning. Det er et forhold, der står centralt i teorier om ”den kønnede organisation”, hvor fx et studie af et finsk statskundskabsinstitut med en overvægt af mandlige ansatte har peget på en række mekanismer, der skaber forskellige vilkår for mandlige og kvindelige ph.d.-studerende (Kantola, 2008). På nogle danske universitetsuddannelser er der en klar overvægt af undervisere af det ene køn, mens der på andre uddannelser er en mere ligelig balance, hvilket oplagt vil kunne påvirke interaktionsmønstrene på uddannelsen. De studerendes erfaring med henholdsvis mandlige og kvindelige undervisere kan også påvirke deres forventninger til underviserne og dermed deres evaluering. Det er derfor vigtigt at undersøge mulige forskelle mellem uddannelsesmiljøer med forskellig kønsmæssig balance.

Endelig er den studerendes eget køn også oplagt vigtigt. Nogle studier peger på, at studerende generelt foretrækker undervisere af samme køn som dem selv (Basow, 2000; Centra og Gaubatz, 2000), mens andre ikke har fundet sådanne forskelle eller ovenikøbet peger på en bias i den modsatte retning (Arbuckle og Williams, 2003; Macnell, Driscoll og Hunt, 2015). Uanset hvad er det klart, at både den enkelte studerendes køn og den overordnede sammensætning af studerende i forhold til køn potentielt kan have betydning for, hvordan underviserne bliver evalueret.

Forskningsdesign

Spørgsmålet om en eventuel forskel på evalueringer af mandlige og kvindelige undervisere kan undersøges på forskellig vis. En afgørende skelnen er, om undersøgelser bygger på faktiske studenterevalueringer eller mere eksperimentelle design. Fordelen ved den første type undersøgelser er, at de belyser, om der reelt er empiriske forskelle på de evalueringer, mænd og kvinder får, og dermed maksimerer den eksterne validitet i forhold til virkelige evalueringer af undervisningsforløb. De mere eksperimentelle design er derimod velegnede til at komme tættere på kausalmechanismerne bag en eventuel forskel i evalueringer.

gerne, fordi man fx kan sørge for, at det ikke er forskelle i undervisningsstil, der påvirker de studerendes evalueringer.

I denne artikel analyseres faktiske evalueringer af undervisningsforløb. Dermed giver artiklen et første indblik i, hvorvidt der findes en kønsbias i danske studerendes evalueringer af deres undervisere. Det vil til gengæld være vanskeligt at identificere, hvilke mekanismer der ligger bag en eventuel forskel i evalueringer af mænd og kvinder. Analysen baseres på evalueringer af undervisningsforløb ved Aarhus BSS. Fakultetet udbyder en række forskellige uddannelser inden for samfundsvidenskab og har samlet 7.627 bachelorstuderende og 4.989 kandidatstuderende. Uddannelserne på BSS spænder ganske vidt med klassiske samfundsvidenskabelige uddannelser som statskundskab, jura og psykologi og erhvervsrettede uddannelser som cand.merc.

På tværs af alle uddannelser fordeler de studerende sig nogenlunde ligeligt mellem kønnene med 49 pct. kvindelige studerende og 51 pct. mandlige studerende (Aarhus Universitet, 2017). Der er imidlertid en betydelig variation i kønsfordelingen på tværs af fakultets uddannelser, og ser man på optagne studerende, varierer andelen af kvindelige studerende således mellem 22 pct. på bacheloruddannelsen i økonomi og 81 pct. på psykologi (Aarhus Universitet, 2016). På undervisersiden er der knap så meget variation, idet der på næsten alle institutter på Aarhus BSS er en overvægt af mandlige videnskabelige medarbejdere. Køns sammensætningen varierer dog trods alt noget. Institut for Statskundskab har med 25 pct. kvinder den laveste andel af kvindelige videnskabelige medarbejdere, mens Institut for Psykologi har en lille overvægt af kvinder, om end dette især skyldes et større antal kvinder end mænd i gruppen af ikke-fastansatte medarbejdere (Aarhus BSS, 2016a).

Siden forårssemestret 2016 er undervisning på Aarhus BSS blevet evalueret ved hjælp af et fælles evalueringsskema, der indeholder spørgsmål om fx den studerendes samlede udbytte af undervisningen, undervisningens tilrettelæggelse og de studerendes eget engagement i undervisningen. Undervisning på Aarhus BSS kan have forskellige formater, og alt efter undervisningsformatet anvendes forskellige versioner af evalueringsskemaet. På bacheloruddannelsen undervises ofte ved en kombination af forelæsninger og holdundervisning, og her evalueres både den underviser, den studerende har haft som forelæser, og holdunderviseren. Sidstnævnte kan ofte være studenterinstruktører eller ph.d.-studerende. På kandidatuddannelserne anvendes både forelæsninger og seminarundervisning.

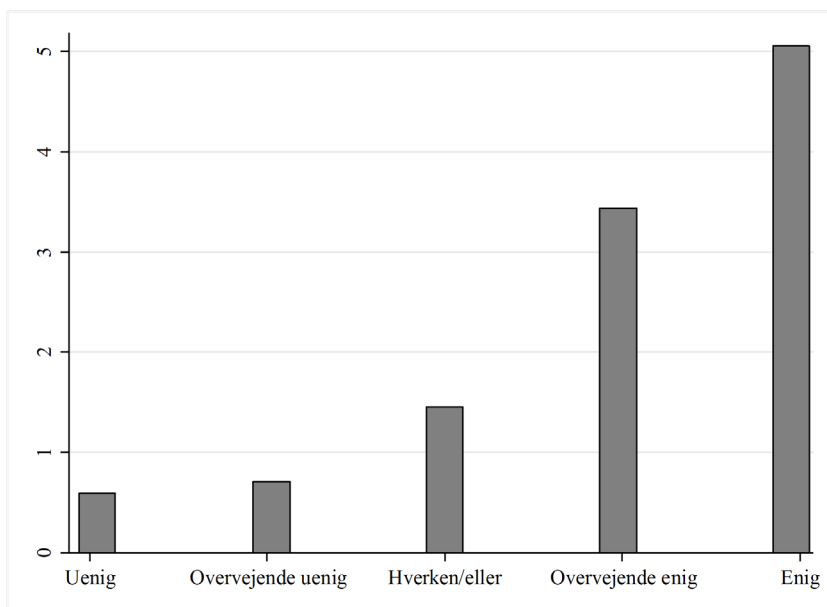
Datagrundlaget for artiklen består af evalueringer fra 2016 og 2017 på tværs af alle uddannelser på Aarhus BSS. I analysen indgår 39.546 evalueringer foretaget af 8.452 forskellige studerende. På undervisersiden er der tale om 1.176

forskellige undervisere, hvoraf 62 pct. er mænd og 38 pct. kvinder. Der er altså tale om et stort datamateriale men med den væsentlige begrænsning, at antallet af evaluerede undervisere – især kvindelige undervisere – trods alt er begrænset især i forhold til analyser, der bevæger sig ned på de enkelte uddannelser eller inkorporerer interaktionseffekter.

Artiklens analyser koncentrerer sig om det spørgsmål i evalueringsskemaet, der går mest direkte på den enkelte undervisers formidlingsevne. Konkret spørges de studerende, om ”Underviseren var god til at formidle stoffet”, og har mulighed for at svare med kategorierne ”enig”, ”overvejende enig”, ”hverken eller”, ”overvejende uenig” og ”uenig”. Dette spørgsmål stilles både til forelæsere og holdundervisere og giver derfor mulighed for at inddrage begge undervisningsformater. I undervisningsforløb med flere undervisere varierer det, om spørgsmålet stilles til underviserne samlet eller gentages for hver underviser. I analyserne anvendes kun undervisningsforløb, hvor evalueringssvarene gælder en enkelt underviser.

Som det fremgår af figur 1, er der et ganske skævt mønster i svarfordelingen på spørgsmålet om undervisernes formidling. Over halvdelen af de studerende

Figur 1: Fordeling af svar på spørgsmål om ”formidling”, pct.



Note: Figuren viser fordelingen af svar på spørgsmålet, ”Underviseren var god til at formidle stoffet”. N = 39.546.

vælger kategorien ”enig”, når de skal vurdere underviserens formidling og yderligere 1/3 vælger ”overvejende enig”. I de videre analyser opdeles svarene derfor i tre hovedkategorier: ”enig”, ”overvejende enig” samt en samlet kategori for de tre nederste værdier. Analysen heraf foretages som en generalized ordered logistic regression.² For at tage højde for at de samme undervisere indgår en række gange i datamaterialet, anvendes klyngerobuste standardfejl justeret for undervisere.

Evalueringer af undervisning: Er der en kønsbias?

Indledningsvis er det relevant at se på sammensætningen af de studerende og undervisere, der indgår i analyserne – og på hvordan dette varierer på tværs af fag. Blandt de studerende er der stort set en ligelig repræsentation af mænd og kvinder med 48 pct. mandlige og 52 pct. kvindelige studerende. Blandt underviserne er der noget større skævhed med samlet set 62 pct. mænd og 38 pct. kvinder. I tabel 1 ses fordelingen mellem de største uddannelsesretninger på Aarhus BSS.

Som forventet er der stor variation mellem uddannelsesretningerne. På psykologi er der fx en overvægt af kvinder både blandt studerende og undervisere, mens der på økonomi omvendt er flest mænd i begge grupper. På andre uddannelser er der en kontrast mellem sammensætningen af studentergruppen og undervisergruppen – mest markant på jura, hvor omkring 2/3 af de studerende er kvinder, mens kun 1/3 af de undervisere, de møder på deres fag, er kvinder.

Tabel 1: Fordeling af studerende og undervisere på uddannelser, pct. mænd/kvinder

	Fordeling af studerende			Fordeling af undervisere		
	Mænd	Kvinder	N	Mænd	Kvinder	N
Erhvervsøkonomi	58	42	3.075	62	38	366
Virksomhedskommunikation	25	75	506	41	59	86
Jura	35	65	783	57	33	167
Psykologi	17	83	873	39	61	127
Statskundskab	49	51	1.031	73	27	91
Økonomi	68	32	651	76	24	102
Øvrige uddannelser	50	50	1.533	68	32	237
Samlet	48	52	8.452	62	38	1.176

Note: Tabellen er baseret på besvarede evalueringer. Der er i alt afgivet 39.546 evalueringer, som i tabellen er fordelt på unikke studerende og undervisere.

Materialet giver dermed et godt udgangspunkt for at undersøge, om der er forskellige eller ens mønstre i evalueringerne på tværs af disse forskellige uddannelser. Især er det interessant at undersøge, om uddannelser med en overvægt af kvinder i undervisergruppen adskiller sig fra uddannelser med en overvægt af mænd – fx fordi det påvirker de studerendes opfattelse af, hvordan en typisk universitetsunderviser er.

Analysen af sammenhængen mellem underviserens køn og de studerendes evalueringer foretages i første omgang på baggrund af det samlede datamateriale. I tabel 2 præsenteres resultaterne af en overordnet analyse af spørgsmålet om underviserens formidling. Analysen inddrager både den studerendes køn, underviserens køn og interaktionen mellem disse. Herudover kontrolleres for de største uddannelsesretninger, der indgår i materialet. Som det fremgår af tabellen, er der tale om en interaktionseffekt mellem underviserens køn og den studerendes køn. De højeste evalueringer ses, når mandlige studerende evaluerer mandlige undervisere, mens kvindelige studerende evaluerer kvindelige undervisere højest. En samlet test af interaktionen mellem underviser- og studenterkøn har en P-værdi på 0,000. Denne skal vurderes i lyset af, at standardfejlene er klyngejusteret for underviser, og må derfor betragtes som en klar indikation af, at der er tale om en signifikant effekt.

Tabel 2: Resultater af regressionsanalyse af evalueringer (generalized ordered logistic regression)

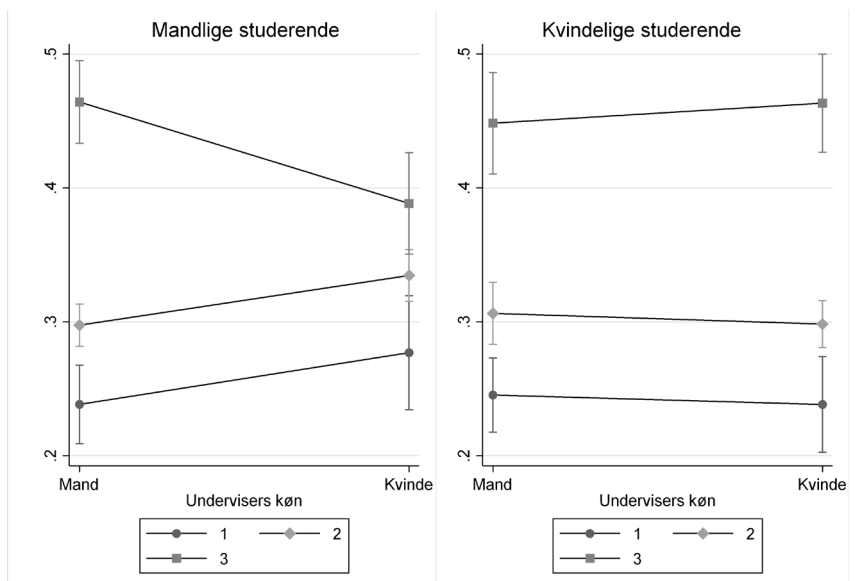
		Udfald 1			Udfald 2		
		Koeffi- cient	Robuste standard- fejl	P-værdi	Koeffi- cient	Robuste standard- fejl	P-værdi
Mandlig studerende	Mandlig underviser	Reference			Reference		
	Kvindelig underviser	-0,20	0,14	0,137	-0,31	0,11	0,003
Kvindelig studerende	Mandlig underviser	-0,13	0,06	0,025	-0,14	0,05	0,005
	Kvindelig underviser	-0,06	0,13	0,636	-0,06	0,10	0,586

Note: Der anvendes generalized ordered logistic regression. N = 39.546 (standardfejl er klyngejusteret for 1.037 undervisere). Afhængig variabel er spørgsmålet: ”Underviseren var god til at formidle stoffet” og udfaldene er: 3 = enig, 2 = overvejende enig, 1 = hverken/eller, overvejende uenig samt uenig. En chi² test af den samlede interaktion mellem underviser og studerendes giver chi² = 28,7 og P-værdi = 0,000. Nagelkerke R² = 0,02. Der indgår dummyvariable for de seks uddannelsesretninger i analyserne.

I figur 2 illustreres resultaterne grafisk. Her vises de forudsagte sandsynligheder for de tre forskellige evalueringsudfald på baggrund af henholdsvis en model, der kun medtager kvindelige studerende, og en model med kun mandlige studerende. Det kan blandt andet ses, at når mandlige studerende evaluerer mandlige undervisere, vil de i omkring 47 pct. af tilfældene give den højest mulige score sammenlignet med 40 pct. af tilfældene for kvindelige undervisere. Det modsvares af, at de i 27 pct. af evalueringerne af kvinder vælger en af de tre laveste kategorier, mens de vælger disse svarmuligheder i 23 pct. af tilfældene for mandlige undervisere. For de kvindelige studerende er mønstret modsat men mindre markant, idet der er en forskel på ca. 2 procentpoint i de kvindelige underviseres favør. Desuden viser figurernes konfidensintervaller, at resultaterne for især de kvindelige studerende er behæftet med stor usikkerhed.

Resultatet af denne overordnede analyse peger på, at der for danske studerende – i hvert fald på Aarhus BSS – kun delvist kan findes de samme mønstre i evalueringen af mandlige og kvindelige undervisere som i de store franske og

Figur 2: Forudsagte sandsynligheder for forskellige evalueringsudfald for mandlige og kvindelige studerende



Note: N for mandlige studerende = 16.830/kvindelige studerende = 22.716. Figuren viser 95 pct. konfidensintervaller. Afhængig variabel er spørgsmålet: "Underviseren var god til at formidle stoffet" og udfaldene er: 3 = enig, 2 = overvejende enig, 1 = hverken/eller, overvejende uenig samt uenig.

hollandske studier. Der er ikke tale om en helt generel sammenhæng mellem underviserkøn og evalueringresultater³ men derimod en mere kompleks sammenhæng mellem de studerendes eget køn og deres vurdering af undervisere af forskelligt køn. Samtidig finder vi for de mandlige studerende et sammenligneligt resultat, idet de – i lighed med de franske og hollandske studerende – er mere tilbøjelige til at give den højeste score til mandlige undervisere end til kvindelige. Effektens størrelse er dog mindre end i fx det franske studie (Boring, 2017: 35). Hertil kommer, at de kvindelige danske studerende udviser en modsatrettet – om end svagere og klart usikker – tendens, idet de er mere tilbøjelige til at vurdere deres kvindelige undervisere i den højeste kategori.

Forskelle mellem uddannelsesmiljøer?

Næste skridt er at undersøge, om der er forskelle i evalueringsmønstrene mellem de uddannelser, der indgår i studiet. Det giver mulighed for at analysere samme spørgsmål i uddannelsesmiljøer med ret forskellige sammensætninger af både studenter og undervisere. Som beskrevet ovenfor fordeler de seks hoveduddannelser dels med: 1) nogle med en overvægt af kvinder i begge grupper (psykologi og virksomhedskommunikation), 2) nogle med flest mænd i begge grupper (økonomi og erhvervsøkonomi) og 3) nogle hvor undervisergruppen er domineret af mænd, men hvor der blandt de studerende enten er flest kvinder eller nogenlunde ligevægt (statskundskab og jura). I analysen nedenfor er uddannelserne delt i to hovedgrupper baseret på sammensætningen af undervisergruppen. I den ene gruppe er uddannelser med flest mandlige undervisere – mellem 62 og 76 pct. – og i den anden gruppe uddannelser med en overvægt af kvindelige undervisere – mellem 59 og 61 pct. Tabel 3 viser hovedresultaterne af separate analyser for henholdsvis mandlige og kvindelige studerende på disse to hovedkategorier af uddannelser. Resultaterne illustreres i figur 3.

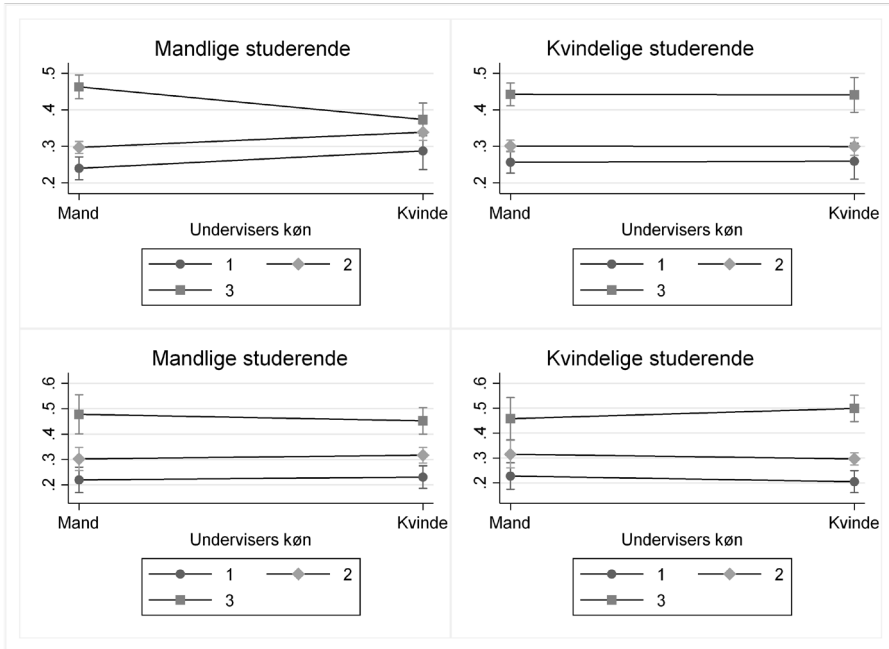
Inddragelsen af uddannelse i analyserne peger på endnu en dimension af sammenhængen mellem underviseres køn og de studerendes evalueringer af undervisningen. Det viser sig nemlig, at der kun i én af de analyserede situationer er en forskel i evalueringen af mandlige og kvindelige undervisere. Det gælder for den situation, hvor der er tale om evalueringer foretaget af mandlige studerende, der læser på uddannelser med en overvægt af mandlige undervisere. Her genfindes resultatet fra analysen af de samlede evalueringer, idet mænd vurderes i den højeste kategori i 46 pct. af tilfældene sammenlignet med 37 pct. af kvinderne. Omvendt får 29 pct. af kvinderne en score i en af de tre laveste kategorier af deres mandlige studerende sammenlignet med 24 pct. af de mandlige undervisere. I de tre øvrige situationer er der tale om ganske svage og ikke-signifikante sammenhænge.

Table 3: Separate analyser for uddannelser med en overvægt af mandlige henholdsvis kvindelige undervisere (generalized ordered logistic regression)

Uddannelser	Studerende	Koefficient for underviser = kvinde	Robuste standardfejl	P-værdi	Antal studerende	Antal klynger
Overvægt af mandlige undervisere	Mandlig studerende	-0,24/-0,37	0,15/0,12	0,111/0,002	14.942	821
	Kvindelig studerende	-0,01/-0,01	0,15/0,12	0,928/0,962	14.007	830
Overvægt af kvindelige undervisere	Mandlig studerende	-0,06/-0,10	0,20/0,19	0,749/0,589	1.888	191
	Kvindelig studerende	0,13/0,17	0,21/0,21	0,521/0,419	8.709	211

Note: Tabellen viser resultaterne af fire separate analyser af de forskellige grupper. Afhængig variabel er spørgsmålet: "Underviseren var god til at formidle stoffet" og udfaldene er: 3 = enig, 2 = overvejende enig, 1 = hverken/eller, overvejende uenig samt uenig. Der anvendes generalised ordered logistic regression og tabellen angiver estimaterne for to udfald på den afhængige variabel, Nagelkerke $R^2 = 0,01, 0,00, 0,00, 0,00$ for de fire modeller.

Figur 3: Forudsagte sandsynligheder for forskellige evalueringsudfald for uddannelser med en overvægt af mandlige (øverst) henholdsvis kvindelige (nederst) undervisere



Note: Figuren viser forudsagte sandsynligheder på baggrund af de fire analyser, der er opsummeret i tabel 3. Figuren viser desuden 95 pct. konfidensintervaller. Afhængig variabel er spørgsmålet: "Underviseren var god til at formidle stoffet", og udfaldene er: 3 = enig, 2 = overvejende enig, 1 = hverken/eller, overvejende uenig samt uenig.

Diskussion og konklusion

Evalueringer fylder ganske meget i uddannelsessystemet. De bruges ikke kun til at arbejde med forbedringer af undervisningen men spiller også en rolle for beslutninger om, hvilke medarbejdere der skal ansættes eller forfremmes. Hertil kommer, at de i stigende grad bliver brugt i generelle kvalitetssikrings-systemer som en central indikator på, hvorvidt enkelte undervisningsforløb såvel som hele uddannelser er af høj kvalitet. På den baggrund er det bekymrende, at en række internationale studier har vist, at kvindelige undervisere får dårligere evalueringer end deres mandlige kolleger. Det er derfor afgørende at undersøge, om en sådan forskel også findes i en dansk universitetskontekst, og denne artikel har taget det første skridt i retning af at belyse dette.

Artiklens analyser bygger på evalueringer foretaget på Aarhus BSS, hvor et fælles evalueringsskema bruges på tværs af uddannelser, der varierer ganske meget i forhold til kønssammensætningen af både studerende og undervisere. Artiklen har således inddraget næsten 30.000 evalueringer af mere end 1.000 forskellige undervisere. På den baggrund er det vist, at der snarere end en helt generel bias i evalueringer ser ud til at være en sammenhæng mellem sammensætningen af undervisergruppen, den studerendes køn og vurderingen af undervisere af forskelligt køn. På uddannelser med en overvægt af mandlige undervisere vurderer mandlige studerende således gennemsnitligt deres mandlige undervisere højere end de kvindelige. Det samme mønster findes ikke for deres kvindelige medstuderende eller på de uddannelser, hvor der er en overvægt af kvinder blandt underviserne.

Selv for de mandlige studerende på uddannelser med en overvægt af mandlige undervisere er der tale om en mindre forskel end i nogle af de nylige udenlandske undersøgelser. Samtidig er det dog et markant resultat, at kvindelige undervisere på mandedominerede uddannelser har en større sandsynlighed for at få en evaluering i en af de laveste tre kategorier end deres mandlige kolleger. I det kvalitetssystem, der fx anvendes på Aarhus BSS, er det en ret markant forskel, idet fag skal vurderes med en samlet score for udbytte af faget på 4,0 for at blive vurderet som velfungerende (Aarhus BSS, 2016b). Ser man blot på mandlige studerende på uddannelser med flest mandlige studerende, giver de helt overordnet deres kvindelige undervisere en score på 3,88 og deres mandlige undervisere en score på 4,04 i vurderingen af deres formidling. I et system, hvor variationen i evalueringresultater ligger i et relativt snævert område, og hvor netop forskelle inden for dette snævre område tillægges betydelig vægt, kan resultaterne her dermed blive ret afgørende.

Analysen fortæller imidlertid ikke meget om, hvorfor dette mønster findes. En nærliggende tanke er, at der er forskel på, hvilke typer undervisere studerende bedst lærer fra. Det kan således tænkes, at mange mandlige undervisere har en undervisningsstil, som virker særligt godt for mandlige studerende og tilsvarende for kvindelige undervisere og studerende. Er det tilfældet, afspejler resultaterne, hvem de studerende lærer mest fra og er dermed for den enkelte studerende en reel indikation af det læringsudbytte, den studerende har fået. Denne forklaring stemmer imidlertid ikke ret godt overens med forskellen mellem uddannelser afhængig af undervisersammensætningen. Her virker det mere oplagt, at forestillinger om, hvordan en ”rigtig” universitetsunderviser er, præges af, hvilke undervisere de studerende møder på uddannelsen. Samtidig kan det også pege på eksistensen af nogle af de mekanismer, som studier af den ”kønnede organisation” har identificeret (Kantola, 2008). I det omfang

disse er til stede, kan det bidrage til at fastholde en skævhed i rekruttering og fastholdelse af mænd og kvinder.

Et oplagt næste skridt vil være at lave studier, der kommer tættere på årsags-sammenhængene bag resultaterne – fx i form af sammenligninger af forskellige underviseres undervisningsstil og inddragelse af flere uddannelser. Indtil da bør resultaterne mane til forsigtighed i anvendelsen af evalueringsresultater. De mønstre, der er beskrevet her, sår således tvivl om, hvorvidt evalueringerne giver et retvisende billede af kvaliteten i undervisning. Derudover er det oplagt at interessere sig for andre spørgsmål i forbindelse med evalueringssystemet. Mens interessen her har været for kønsmæssige forskelle i evalueringer, er det fx oplagt, at der lige såvel kan være forskel i vurderingen af undervisere af forskellig etnicitet, alder eller baggrund. Hertil kommer, at disse forskelssættende kategorier kan forstærke hinanden som understreget i intersektionalitetsperspektivet (se introduktion til dette temanummer (Fiig, 2019)). Tilsvarende er der også grund til at se nærmere på de sammenhænge mellem underviserens krav til studerende og de studerendes karakterforventninger, som er fundet i udenlandske studier. Kan tilsvarende mønstre findes i Danmark, udgør det endnu en grund til at være mere kritisk i brugen af undervisningsevalueringer.

Noter

1. Med effektstørrelser på op til 1 point på en 6-pointskala.
2. Denne model vælges, fordi antagelserne for at anvende almindelig ordered logistic regression ikke er overholdt.
3. En analyse uden inddragelse af interaktionen mellem underviserens og den studerendes køn viser kun en meget svag og klart ikke-signifikant forskel mellem mænd og kvinder i mændenes favør.

Litteratur

- Arbuckle, Julianne og Benne D. Williams (2003). Students' perceptions of expressiveness: Age and gender effects on teacher evaluations. *Sex Roles* 49 (9/10): 507-516.
- Bandura, Albert (1997). *The Exercise of Control*. New York: W.H. Freeman and Company.
- Basow, Susan A. (2000). Best and worst professors: Gender patterns in students' choices. *Sex Roles* 43 (5/6): 407.
- Boring, Anne (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics* 145: 27-41.
- Boring, Anne, Kellie Ottoboni og Philip B. Stark (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research*.

- Braga, Michela, Marco Paccagnella og Michele Pellizzari (2014). Evaluating students' evaluations of professors. *Economics of Education Review* 41: 71-88.
- Burns-Glover, Alyson L. og Dale Veith (1995). Revisiting gender and teaching evaluations: Sex still makes a difference. *Journal of Social Behavior and Personality* 10 (4): 69.
- Centra, John A. og Noreen B. Gaubatz (2000). Is there gender bias in student evaluations of teaching? *The Journal of Higher Education* 71 (1): 17-33.
- Dalsgaard, Peter, Mikkel Bille, Rasmus Bjørk, Steffen Dalsgaard, Henrik Dimke, Karen Gram-Skjoldager og Nikolaj Zinner (2017). <https://www.altinget.dk/forskning/artikel/undervisere-her-er-de-fem-stoerste-problemer-med-studenterevalueringer> (12. januar 2018).
- Feldman, Kenneth A. (1992). College students' views of male and female college teachers: Part I: Evidence from the Social Laboratory and Experiments. *Research in Higher Education* 33 (3): 317-375.
- Feldman, Kenneth A. (1993). College students' views of male and female college teachers: Part II: Evidence from students' evaluations of their classroom teachers. *Research in Higher Education* 34 (2): 151-211.
- Fiig, Christina (2019). Skarpere med end uden? Kønsforskning og politologi. *Politica* 51 (1): 5-18.
- Jaschik, Scott (2015). Rate my word choice. *Inside Higher Education*.
- Kantola, Johanna (2008). "Why do all the women disappear?" Gendering processes in a political science department. *Gender, Work and Organization* 15 (2): 202-225.
- Laird, Thomas F.N., Amy K. Garver og Amanda S. Niskode-Disset (2011). Gender gaps in collegiate teaching style: Variations by course characteristics. *Research in Higher Education* 52: 261-277.
- Laube, Heather, Kelley Massoni, Joey Sprague og Abby L. Ferber (2007). The impact of gender on the evaluation of teaching: What we know and what we can do. *NWSA Journal* 19 (3): 87-104.
- Macnell, Lillian, Adam Driscoll og Andrea N. Hunt (2015). What's in a name? Exposing gender bias in student ratings of teaching. *Innovative Higher Education* 40 (4): 291-303.
- Martin, Lisa L. (2016). Gender, teaching evaluations, and professional success in political science. *PS: Political Science & Politics* 49 (2): 313-319.
- Mengel, Frederikke, Jan Sauermann og Ulf Zölitz (2017). Gender bias in teaching evaluations. *Journal of the European Economic Association*.
- Nielsen, Mathias Wullum (2017a). Scandinavian approaches to gender equality in academia: A comparative study. *Scandinavian Journal of Educational Research* 61 (3): 295-318.

- Nielsen, Mathias Wullum (2017b). Reasons for leaving the academy: A case study on the “opt out” phenomenon among younger female researchers. *Gender, Work and Organization* 24 (2): 134-155.
- Regeringen (2018). *Fleksible universitetsuddannelser til fremtiden*.
- Rudman, Laurie A. og Peter Glick (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues* 57 (4).
- Sinclair, Lisa og Ziva Kunda (2000). Motivated stereotyping of women: She’s fine if she praised me but incompetent if she criticized me. *Pers Soc Psychol Bull* 26 (11): 1329-1342.
- Sprague, Joey og Kelley Massoni (2005). Student evaluations and gendered expectations: What we can’t count can hurt us. *Sex Roles* 53 (11/12): 779-793.
- Stes, Ann, Vincent Donche og Peter V. Petegem (2004). Understanding differences in teaching approaches in higher education: An evidence-based perspective. *Reflective Education* 9 (1): 21-35.
- Uttl, Bob, Carmela A. White og Daniela Wong Gonzalez. (2017). Meta-analysis of faculty’s teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation* 54: 22-42.
- Wright, Stephan L. og Michael A. Jenkins-Guarnieri (2011). Student evaluations of teaching: combining the meta-analyses and demonstrating further evidence for effective use. *Assessment & Evaluation in Higher Education* 37(6): 683-699.
- Aarhus BSS (2016a). *Kønsbalance jan-juni 2016*. Aarhus: Aarhus BSS
- Aarhus BSS (2016b). *Procedure til evaluering af kursusforløb på Aarhus BSS*. Aarhus: Aarhus BSS
- Aarhus Universitet (2016). *Optagelsesstatistik 2016*. Aarhus: Aarhus Universitet (Uddannelsesstrategisk sekretariat).
- Aarhus Universitet (2017). *Aggregerede studieledertal efteråret 2017*. Aarhus: Aarhus Universitet