Uniped
Tidsskrift for universitets- og høgskolepedagogikk
VITENSKAPELIG PUBLIKASJON

Universitetsforlaget

Årgang 45, nr. 1-2022, s. 39–52
ISSN Online: 1893-8981
https://doi.org/10.18261/uniped.45.1.5

«Jeg tror jeg brukte den som taus kunnskap, om du skjønner meg»: En studie av sensurveiledning på bachelornivå

"I think I used it as tacit knowledge, if you know what I mean": A study on the use of examiner guidelines at bachelor's level.

Vidar Gynnild
Professor i universitetspedagogikk, Institutt for pedagogikk og livslang læring (IPL), NTNU
vidar.gynnild@ntnu.no

Sammendrag
Dette er en studie av sensurveiledning på bachelornivå i et stort utdanningsvitenskapelig emne. De sju sensorene ble intervjuet i etterkant av sensuren om sin bruk av veiledningsteksten. Artikkelen dokumenterer i hvilken grad sensurveiledningen ble benyttet, og om det var godt samsvar mellom veiledningsteksten og oppgavene. Veiledningen hadde ingen referanser til læringsutbyttebeskrivelsen, og de generiske karakterbeskrivelsene ble forklart på måter som bidro til uklarhet blant sensorene. Sensurveiledningen ble benyttet i ulik grad og på ulik måte, men felles for alle sensorene var et ønske om kollegialt samarbeid. Studien diskuterer også utfordringer som oppstår når sensurveiledning med et positivistisk utgangspunkt blir benyttet i et emne fra et tolkningsbasert kunnskapsdomene.

Nøkkelord
vurdering, sensur, pålitelig, gyldig, transparent, rettferdig

Abstract
This study deals with the use of examiner guidelines in a large educational science course. The seven examiners were interviewed after the examination about their use of the guidance text. The article documents the extent to which the examiner guidelines were used, and whether there was good correspondence between the guidance text and the assignments. The examiner guidelines had no reference to learning outcome descriptions, and the generic grade descriptions were explained in ways that contributed to ambiguity among the examiners. The examiner guidelines were used in different ways and to different degrees; however, common to all examiners was a desire for a greater degree of collegial cooperation along the way. The study also discusses challenges that arise when examiner guidelines with a positivist anchoring are used within an interpretation-based knowledge domain.

Keywords
assessment, grade decision, reliable, valid, transparent, fair

Innledning
Vurdering og sensur har betydning for studieprogresjon og yrkesmuligheter for den enkelte student, og som dokumentasjon for framtidige arbeidsgivere. Det er derfor naturlig at myndighetene setter nasjonale krav til vurderingsordninger og sensur. Krav om to sensorer har røtter helt tilbake til Københavns Universitet i 1788 (Lauvås & Jakobsen, 2002), og Thomsen (1975) hevder at ordningen kan spores tilbake til opplysningstidens ønske om rettssikkerhet

for borgerne. I nyere tid satte universitetsloven av 1995 § 50 krav om to sensorer, hvorav minst én ekstern, men ordningen kom under sterkt press blant annet fordi kun et fåtall andre nasjoner benyttet en tilsvarende ordning. Kravet om to sensorer ble derfor opphevet etter relativt kort tid ved lov av 28. juni 2002 nr. 62 (NOU 2020: 3). Denne endringen åpnet muligheten for å benytte kun én sensor, som kunne være faglæreren selv. Dette representerte en kraftig omlegging og privatisering av sensurarbeidet samtidig som kostnadene trolig ble noe redusert.

Endringene beskrives ved to sentrale kjennetegn, nemlig færre sensorer og en omdefinering av den eksterne sensorens rolle. Myndighetene satte etter dette krav om ekstern evaluering av vurderingsordningene, en ordning som fortsatt gjelder i universitets- og høyskoleloven § 3–9. Kandidatenes rettssikkerhet ble likevel svekket fordi ordningen med enesensur økte faren for feil, og myndighetene ønsket derfor tiltak for å ivareta studentenes rettssikkerhet ved sensur. Regjeringen ba etter dette om at institusjonene skal «stille krav om sensorveiledning til alle eksamener» (Meld. St. 16 (2016–2017) *Kultur for kvalitet i høyere utdanning*). Samme sted slås det fast at mer og bedre veiledning «vil dessuten sikre mer ensartet vurdering [. . .] så vel mellom de ulike sensorene i førstegangssensuren som mellom førstegangs- og klagesensuren» (s. 58). Enesensur forutsetter godt sensorsamsvar, noe som slett ikke alltid er tilfellet, og som er vel dokumentert i mange klagesaker.

Et illustrerende eksempel er en student som fikk tre ulike karakterer på samme eksamen. Ved ordinær sensur ble besvarelsen vurdert til F (ikke bestått), og studenten valgte da å klage. Ved ny sensur ble besvarelsen overraskende vurdert til A (beste karakter). Til tross for dette ble det ingen A på vitnemålet på grunn av universitets- og høyskolelovens krav om ny vurdering ved avvik på to eller flere karakterer fra opprinnelig sensur. Sluttkarakteren ble C, med følgende kommentar fra studenten: «Jeg bare synes det er vanskelig å vite hvordan jeg skal forholde meg til karaktersystemet når jeg får så forskjellige tilbakemeldinger fra forskjellige folk innenfor det samme fagfeltet», gjengitt i et intervju med nettstedet www.trd.by, som først omtalte saken (21.09.21).

Det framstår som et høyst beklagelig faktum at den endelige karakteren, i ukjent omfang, kan være mer avhengig av sensorens kvalifikasjoner og dømmekraft enn besvarelsens kvalitet. Alvoret i situasjonen motiverte artikkelforfatteren til en utforskende studie av sensur der konteksten fungerer som en plattform for analysen. I motsetning til evalueringsstudier tar denne artikkelen ikke stilling til om praksis er god eller dårlig, ei heller er forfatteren opptatt av normative perspektiver ved sensur.

Begrepene «sensorveiledning» og «sensurveiledning» benyttes om hverandre i dagligtalen. Selv benytter jeg begrepet «sensurveiledning» fordi jeg mener dette gir den riktigste referansen.

Bakgrunn og problemstillinger

Lavt sensorsamsvar er et velkjent fenomen, og i en norsk artikkel diskuterer forfatterne om dette kan endres (Bjølseth et al., 2011). Det blir vist til eksempler på omfattende karaktersprik og at begrunnelser ved klagesensur ofte peker i ulike retninger. Tilsvarende utfordringer blir rapportert i en artikkel om reliabilitet ved vurdering av essaybesvarelser (Asmyhr, 2011). Skriftlige kriterier ble ikke benyttet i særlig grad, mens bruken av subjektive kriterier var utbredt. I en eksperimentell studie fant forfatterne at sensorer påvirket hverandre dersom den opprinnelige karakteren først ble gjort kjent for den andre sensoren (Eriksen & Rasch, 2009). Psykologiske mekanismer ved press og sosial tilpasning kan trolig bidra til

å forklare disse observasjonene. Samtidig kan naturligvis en ekstern sensor også bidra til å oppdage og rette opp tilfeldige feil i vurderingen (Eriksen & Rasch, 2009).

Medieoppslag og forskning tegner alt i alt et dystert bakteppe for denne undersøkelsen (Bloxham et al., 2016; Lauvås, 2018). Blind klagesensur ved Universitetet i Oslo førte for eksempel til større avvik og større spredning i karakterene sammenliknet med ordinær sensur (Gynnild, 2019), og en internasjonal studie rapporterer om bekymringsfulle utviklingstrekk ved britiske universiteter (Bloxham, 2009). Selv om situasjonen kan være annerledes andre steder, gis det tankegods og erfaringer som kan være gjenkjennelig også her til lands, slik dette sitatet illustrerer:

> [. . .] in developing rigorous moderation procedures, we have created a huge burden for markers which adds little to accuracy and reliability but creates additional work for staff, constrains assessment choices and slows down feedback to students. [. . .] The general lack of discourse on marking in higher education (HE) allows assumptions on reliable standards to continue unchallenged perhaps because it is too uncomfortable to discuss these matters [. . .] (Bloxham, 2009, s. 209).

En britisk studie konkluderer med at tid og energi ved universitetene mest har vært rettet mot praktiske tiltak som scoringsskjema og vurderingsrubrikker med mindre vekt på oppfølgende studier som undersøker effekten av tiltakene. Internasjonal forskning viser imidlertid at tiltakene oftest er basert mer på antakelser enn forskningsbasert kunnskap, og at verdien er svært begrenset:

> These findings indicate that criteria are likely to have limited power in achieving consistent judgement. Shared language is insufficient to ensure shared interpretation of common criteria. In addition, we cannot be confident that only published criteria will be drawn upon for judgement or that they will be weighted similarly by markers (Bloxham et al., 2016, s. 478).

I nyere forskning har flere undersøkt tolkningsrommet for kriteriebasert vurdering og bruken av eksplisitt versus taus kunnskap ved sensur (Bloxham et al., 2011b; Bloxham et al., 2016; Hendry et al., 2012; Sadler, 2005, 2009b). Her til lands har det vært tradisjon for å gjøre vurderingskriteriene mest mulig transparente, for eksempel i sensurveiledning. Et sosiokulturelt utgangspunkt vektlegger imidlertid et mer dynamisk syn på standarder der konsensus mellom fagkyndige er et bærende prinsipp (Ajjawi et al., 2019). Et tredje perspektiv som har vært trukket fram, er aktiv deltakelse fra studentenes side som bidrag til læring og egenvurdering. Ulike perspektiver avspeiler ulike kunnskapssyn for ulike formål og med ulike metoder som løsning.

Universitetet i Tromsø satte krav om sensurveiledning ved alle eksamener allerede fra 2009 (Prop. 59 L (2013–2014), s. 42). Samme år konkluderte imidlertid en NIFU-rapport med et «hovedskille i kvaliteten på sensuren mellom enesensur og det å sensurere i fellesskap, dvs. to eller flere personer» (Meld. St. 16 (2016–2017) *Kultur for kvalitet i høyere utdanning*, s. 58). Universitets- og høgskolerådets styre tilrådde etter dette to sensorer ved muntlig prøve og eksamener som etter sin natur ikke lar seg etterprøve, og at det bør være en ekstern sensor ved vurdering av hovedoppgaven. I 2014 presiserte departementet at det vil «være helt avgjørende å ha gode sensurveiledninger for å minimere sjansene for at sensorene i første og andre instans legger til grunn ulike kriteriesett for sine vurderinger» (Prop. 59 L (2013–2014), s. 58). I 2018 ble kravet om sensurveiledning obligatorisk.

Formålet med sensurveiledningen var at «like oppgaver vurderes likt, ved at sensorene legger de samme retningslinjene til grunn for vurderingen og gjør vurderingen opp mot det fastsatte læringsutbyttet for emnet og studieprogrammet» (NOU 2020: 3, s. 201). Det ble

likevel uttrykt bekymring for om sensorene ville omgå sensurveiledningen, for eksempel ved å vekte kriterier annerledes enn fastsatt, men konklusjonen var at tiltaket «bidrar til bedre kvalitetssikring av det fastsatte læringsutbyttet og trolig bedre samsensur» NOU 2020: 3, s. 201).

Formålet med denne undersøkelsen er å utforske bruken av sensurveiledning i en gitt kontekst der følgende problemstillinger danner utgangspunkt for datainnsamling og analyse:

- I hvilken grad, eventuelt på hvilken måte, blir sensurveiledningen benyttet av sensorene?
- I hvilken grad blir veiledningen opplevd som tilstrekkelig under sensurarbeidet?
- I hvilken grad er det samsvar (kongruens) mellom oppgavetekster og sensurveiledning?

Data og metode

Artikkelen bygger på en kasusstudie av et emne på bachelornivå med ca. 400 studenter høsten 2020. Dette emnet valgte jeg fordi jeg ønsket å gjennomføre en dybdestudie med flere sensorer i håp om å avdekke forskjeller og likheter under bruken av sensurveiledningen. Grunnet koronapandemien ble hjemmeeksamen over tre dager benyttet for første gang, og alle hjelpemidler var tillatt. Eksamen besto av to essayoppgaver til valg med støtte i to tekster, og besvarelsene kunne maksimalt være 3500 ord. Hver av de sju sensorene, seks kvinner og én mann, sensurerte 50–60 besvarelser hver med en tidsfrist på 14 dager. Sensorenes erfaringsbakgrunn varierte fra 0 til 40 år, og alle ble organisert i par for frivillig konsultasjon underveis, men arbeidet ble oftest gjennomført alene.

Universitetet setter krav om skriftlig sensurveiledning, men det det fantes ingen felles, institusjonell praksis for utforming av dette dokumentet, og den valgte løsningen bygde på etablert praksis. Foruten administrativ informasjon som emnekode/navn og eksamenstype inneholdt dokumentet oppgavetekstene, oversikt over pensumlitteratur, eksamenskrav og en kort beskrivelse av minstekrav for å bestå eksamen. Samlet omfang var på fire sider. Dokumentet inneholdt generiske karakterbeskrivelser, men ingen læringsutbyttebeskrivelser. Å lage sensurveiledningen var en privatisert oppgave, og instituttene benyttet ulik praksis for å utforme sensurveiledninger.

Datagrunnlaget består av sensurveiledningen og semistrukturerte intervju få dager etter at sensuren falt i januar 2021. Selv hadde jeg ingen bindinger eller relasjoner til noen av sensorene. Intervjuene ble gjennomført i løpet av én klokketime ved bruk av den digitale plattformen «Zoom», og deretter transkriberte jeg dem. Det semistrukturerte formatet muliggjorde oppfølgende spørsmål for å etablere innsikt i sensorenes synspunkter og begrunnelser under bruken av sensurveiledningen. Studien er gjennomført innen et tolkningsbasert paradigme der jeg er opptatt av mekanismer og sammenhenger, uten ambisjon om absolutte sannheter, men med ønske om å bidra til en diskurs om sensur.

Undersøkelsen er godkjent av Norsk Samfunnsvitenskapelig Datatjeneste (NSD), og alle sensorene ble informert om forskningsprosjektets formål, personvern og rettigheter. Jeg transkriberte intervjuene for åpen koding i kategorier forankret i dataene (Hsieh & Shannon, 2005). Dette var en iterativ prosess der jeg benyttet koder, forklarende tekst og illustrative sitater. Under den åpne kodingen ble tekstavsnitt med relevans for problemstillingene identifisert og begrepsfestet (kodet), mens jeg brukte forklarende tekst og illustrative for å eksemplifisere begrepene i teksten.

Resultater

En holistisk (helhetlig) vurderingstradisjon har lenge vært dominerende innen humanistiske og samfunnsvitenskapelige fag, mens bruken av sensurveiledning kan oppfattes som et steg i retning av en analytisk vurderingspraksis med verbale beskrivelser av vurderingskriteriene. Jeg undersøker nå forskningsspørsmålene i tur og orden: de to første med data fra de sju intervjuene og det siste med utgangspunkt i sensurveiledningen (dokumentanalyse) med støtte i intervjudata. Ettersom utvalget kun består av de sju sensorene, validerer jeg undersøkelsen ved korte sitater fra alle. Det første spørsmålet er altså i hvilken grad, eventuelt på hvilken måte, sensurveiledningen blir benyttet.

> Jeg tror jeg brukte den som taus kunnskap, om du skjønner meg. [. . .] Kanskje besitter vi så mye erfaring og taus kunnskap at vi har det i oss, også blir det et ekstra pålegg som nesten forstyrrer litt (R1).

> Jeg leste over på forhånd før jeg leste oppgavene. Så brukte jeg den ganske mye underveis, jevnlig når jeg var usikker. [. . .] (R2).

> Jeg la vekt på fire punkter, og det ble mye enklere da jeg gjorde det på den måten (R3).

> Betyr det at du laget din egen sensurveiledning da? (I).

> Jeg har faktisk gjort det! [. . .] Hadde denne strukturen vært bakt inn i oppgaveteksten, så ville det vært mye lettere fordi da hadde man et felles grunnlag å sensurere ut fra (R3).

> Jeg vurderer vel mer studentene opp mot hverandre enn å bruke sensurveiledningen sånn veldig detaljert. [. . .] (R4).

> Du kalibrerer dine standarder ved bruk av eksempler, da? (I).

> Ja, studenteksempler (R4).

> Hvor mye annerledes ville resultatet blitt uten sensurveiledningen? (I).

> Tror det ville blitt stort sett det samme (R4).

> Jeg forsøkte å ta utgangspunkt i sensurveiledningen. Men det som sto om en god besvarelse, fikk jeg egentlig ikke så mye ut av, så jeg tok mer utgangspunkt i det jeg mente var en god besvarelse. [. . .] (R5).

> [. . .] det tror jeg gikk litt ut fra teften og mageføelsen. Du ser oppgaven, og dette holder, eller dette er ikke godt nok ut fra en helhetlig vurdering (R6).

> Jeg la alle besvarelsene i en bunke, og så jobbet jeg meg ut fra midten, [. . .] altså en form for relativ vurdering. Etter å ha sortert i tre bunker, sjekket jeg grensen mellom E og F (R7).

> Hvor lik eller forskjellig ville sensuren vært uten sensurveiledning? (I).

> Den ville ikke vært så forskjellig, tenker jeg (R7).

Kun én av sensorene benyttet sensurveiledningen aktivt underveis, mens en annen forsøkte, men ga opp og gikk over til egne kriterier etter hvert. Flere trekker fram betydningen av taus kunnskap og bruken av relativ vurdering som vurderingsprinsipp. Sensurveiledningen ser

generelt ikke ut til å ha spilt en sentral rolle underveis, og det er interessant å registrere at de fleste valgte løsninger som ikke var i tråd med anbefalingene. Helhetlig og relativ vurdering var utbredt, og sensurveiledningen ble mest benyttet av sensorer med liten erfaring. En sensor valgte å lage sin egen sensurveiledning og mente at den selvvalgte strukturen med hell kunne vært bakt inn i selve eksamensoppgaven. De to som uttalte seg eksplisitt om saken, mente resultatet ville blitt omtrent det samme uten sensurveiledningen.

Det andre forskningsspørsmålet retter seg mot nytteverdien av sensurveiledning. Hensikten er nå å undersøke i hvilken grad veiledningen ble opplevd som tilstrekkelig av sensorene:

> Det er vanskelig å lage en sensurveiledning som treffer alle sensorene, fordi noen av oss har jo vært inne som undervisere i emnet og kjenner emnet godt. [. . .] Jeg er usikker på om dem skjønner det ordentlig? Har vi samme opplevelse rett og slett? (R1).

> Jeg er jo vant til å sette meg ned med en kollega, at vi bytter oppgaver også. Det savnet jeg denne gangen. Det gjør jeg alltid i alle emner, så argumenterer vi rundt det. Så hvis sensurveiledningen skal erstatte det, da må den sannelig være lang, tenker jeg (R2).

> Det med enesensur er livsfarlig. Jeg føler meg så mye tryggere når vi var to. [. . .] For da kan vi drøfte hvorfor vi forstår oppgaven på den eller den måten, noe man utvikler over tid sammen med kolleger som bidrag til fellesforståelsen (R3).

> Jeg finner veldig mye verdi i denne samsensuren, og kommer det opp spørsmål så diskuterer vi det. [. . .] Det jeg fikk mest nytte av ved siden av den tause kunnskapen, var den diskusjonen vi hadde i gruppene (R4).

> For meg som er helt fersk, hadde det vært veldig mye bedre å sensurere sammen med noen andre, og jeg skjønner egentlig ikke at de kan sette meg på sensur. For det forutsetter jo egentlig bare min bakgrunnskunnskap [. . .] Det synes jeg egentlig kan bli litt urettferdig [. . .] (R5).

> Hvis du er to, får du en second opinion, en form for konsensus om strek-karakterene som en sensorveiledning aldri klarer å fange opp. [. . .] Den beste sensurveiledningen har du i en kollega, en medsensor. [. . .] Det ville gjort meg tryggere og vært en rettssikkerhetsventil for studentene. Jeg har aldri erfart at sensurveiledningen har bidratt til en mer pålitelig og konsistent vurdering (R6).

> Bidrar sensurveiledningen til mer pålitelig og konsistent vurdering, eller er dette bare en intensjon? (I).

> Det er et godt spørsmål. Vanskelig å uttale seg om, tenker jeg. Alle har en masse taus kunnskap. Vi vurderer ut fra det vi selv har erfart (R7).

Sitatene dokumenterer en slående samstemthet rundt ønsket om en kollegialt basert sensur. Rett nok ble sensorene organisert i par, men ordningen ble ikke benyttet av alle. Enesensur blir omtalt som utilstrekkelig, og flere peker på verdien av et bredere kunnskaps- og erfaringsgrunnlag. Som en av respondentene påpeker, er sensurkompetanse noe som blir utviklet over tid sammen med kolleger. Flere er kritiske til egen kompetanse og er usikker på om nødvendige kvalifikasjoner er til stede, og det framgår tydelig av intervjuene at ønsket om samarbeid ikke er begrunnet i gammel vane.

Kandidatenes rettssikkerhet blir også trukket fram som et vesentlig tema fordi sensur innebærer ansvarsfullt arbeid med konsekvenser for videre studier, arbeidsliv og karrieremuligheter.

Det tredje forskningsspørsmålet retter seg mot graden av samsvar (kongruens) mellom eksamensoppgavene og sensurveiledningen. Dersom sensurveiledningen skal fylle sin funksjon, forventes den å tydeliggjøre oppgavesettets krav til prestasjon ut fra faglig innhold og nivå (kriterier og standarder). I dagligtalen gir kriterier referanse til innhold og nivå samtidig, men analytisk dreier det seg om to ulike dimensjoner, som enklest lar seg visualisere i form av en vurderingsrubrikk med en horisontal og vertikal akse. Ut fra ideen om samstemt undervisning (Biggs, 1996) forventes sentrale prinsipper fra konstruktivistisk læringsteori å bli omsatt også i vurderingsdesignet.

Eksamensoppgaven og sensurveiledningen overlapper imidlertid på en slik måte at det oppstår et innholdsmessig sprik (inkongruens) mellom dem. Oppgave 1 etterspør blant annet «hvordan det står til med ungdoms psykiske helsetilstand [. . .]» etterfulgt av spørsmål om «hvilke tiltak som kan gjøres for å ruste unge mennesker til å håndtere livet best mulig. [. . .]» Veiledningen krever imidlertid svar på hvordan ungdommers psykiske helsetilstand kan forstås med utgangspunkt i «pensumrelevant litteratur», mens den et annet sted benytter uttrykket «relevant pensumlitteratur». Her er det språklige nyanser som i verste fall kan bidra til uklarhet om kildebruk. Enda et eksempel er krav om å utvise «breddekunnskap» og «dybdekunnskap» for ulike formål i besvarelsene, som heller ikke er presisert i oppgavetekstene. En sensor kommenterte sin opplevelse av situasjonen på følgende vis: «Jeg føler at det introduseres en ny oppgave i selve sensurveiledningen» (R3).

Sensurveiledningens satte videre krav om drøfting basert på eksempler, noe som heller ikke er nevnt i oppgavetekstene, jamfør: «Drøfting skal bygge på 1–3 selvvalgte teorier innen temaet livsmestring, selvoppfatning og motivasjon» (fra oppgave 1), mens sensurveiledningen setter et utvidet krav: «Kandidaten skal basere sin drøfting på relevante teorier og eksempler.» Dette ble drøftet på sensurmøtet, der det ble vedtatt å trekke i karakteren for besvarelser uten eksempler:

> Det ble enighet om at en skulle trekke en karakter dersom det ikke ble brukt eksempler. [. . .] Jeg har skrevet i en begrunnelse at du fikk C fordi du ikke hadde med eksempler. Jeg ble jo faktisk formet av det de sa da! [. . .] Der en selv har bidratt til uklarhet, bør ikke studentene straffes for det. Det var veldig urettferdig og etisk ikke riktig da! (R5).

> Jeg reagerte på at det var veldig få som benyttet eksempler i oppgavene. Jeg har valgt å se bort fra dette fordi det ikke står nevnt i oppgaven. Hvordan i all verden kan man da kreve det av studentene? Du kan ikke kreve det av studentene når det ikke står i teksten, synes jeg da! Jeg føler at det introduseres ting som kompliserer mer enn det klargjør (R3).

Tilsvarende misforhold oppsto ved omtale av krav til nivå. Sensurveiledningen er her strukturert under to hovedpunkter: «En god prestasjon betinger: [. . .]» og «Minstekrav for å bestå eksamen». En god prestasjon betinger at kandidaten utviser «svært høy grad av selvstendighet», mens karakteren A betinger «høy grad av selvstendighet», som kan oppfattes som lavere krav. Det foreligger imidlertid ingen referanse til generelle eller spesielle karakterbeskrivelser, læringsutbyttebeskrivelser og emnedesign, som kan ansees som overordnete, autoritative kilder for oppgavedesign og veiledning. Ikke overraskende uttrykte flere sensorer frustrasjon over situasjonen, spesielt bruken av udefinerte begreper og selvmotsigende utsagn som svekket dokumentets troverdighet og brukervennlighet:

> Det er veldig mye tolkning, som jeg mener er basert på tilfeldigheter, og det blir jo litt feil» (R5).

> Det er så mye floskler [. . .] Essensielt sett er dette bare en serie med ord som ikke sier noen ting! Jeg mener at den hjelper ikke meg i det hele tatt, faktisk! (R3).

Nynorskoversettelsene er preget av en blanding mellom nynorsk og bokmål med en rekke feil, vel egnet til sterke reaksjoner ut fra et språklig ståsted. Sensurveiledningen fikk en blandet mottakelse blant sensorene, men generelt etterlater intervjuene inntrykk av mye frustrasjon og usikkerhet.

Analyse

Vurderingsforskeren Susan Orr hevder at trender innen vurdering er forankret i et positivistisk kunnskapsparadigme med vekt på eksplisittgjøring av kriterier: «For the positivist researcher all student work has a correct mark and the aim is to explore ways to maximize the chances that the student is awarded this mark» (Orr, 2007, s. 646). Orr beskriver dette som en «tekno-rasjonalistisk» tilnærming og postulerer et epistemologisk alternativ basert på en tolkningsbasert praksis. I tråd med dette hevder Shay (2005) at vurdering av komplekse oppgaver skjer best i tolkningsbaserte fellesskap, og Sadler har redegjort for prinsipielle utfordringer ved kodifisering av standarder (Sadler, 2014).

Bruken av sensurveiledning er tuftet på et positivistisk kunnskapsparadigme der transparens i kriterier og prosedyrer står sentralt, uten at det underliggende kunnskapssynet blir problematisert: «The concept of transparency implies that the total explicitness is attainable, which the role of tacit practice within assessment militates against» (Orr, 2007, s. 646). Sensurveiledningen representerte tolkningsutfordringer som resulterte i frustrasjon og usikkerhet blant sensorene, som på sin side ønsket mer kollegialt samarbeid: «Det jeg fikk mest nytte av ved siden av den tause kunnskapen var den diskusjonen vi hadde i gruppene» (R4). Et begrepskjennetegn ved taus kunnskap er at den ikke lar seg artikulere fullt ut. Den blir etablert gjennom praksis og dokumentert ved bruk, men utvikler seg annerledes sammenliknet med teoretisk kunnskap. Den tause kunnskapen speiler sensorenes praksisteori på ulike tidspunkter i karrieren, men kan utvikles fram mot et profesjonelt skjønn.

En sensurveiledning kan oppfattes som «påstandskunnskap», som er lite tilgjengelig for nybegynnere, ettersom taus kunnskap har øving som viktigste kilde til kunnskapsvekst. Deltakelse i praksisfellesskap med mer erfarne kolleger er et eksempel på en slik læringsarena. Ifølge en kjent kunnskapsteoretiker vet vi mer enn vi kan uttrykke med ord, og kunnskapen artikuleres på andre måter enn via verbalspråket (Polanyi, 1998). Derfor inngår sensurveiledningen i et misforhold mellom mål og midler ved sensur. Mens målet er pålitelig og rettferdig sensur, krever prosessen høy grad av skjønn og profesjonell dømmekraft, som igjen forutsetter tolkning og erfaring, slik Sadler skriver:

> At the very heart of all grading processes [. . .] lies the professional judgment of the university teacher as to the standards that are employed. Criteria, whichever way they are understood, are simply nominated characteristics or properties, and these properties are rarely exhibited as either unambiguously present on the one hand, or completely absent on the other. It is always a matter of degree (Sadler, 2003, s. 9).

Ut fra et sosiokulturelt perspektiv gir kriterier og standarder mening i lys av tolkningsrammer forankret i praksis der gjenkjennelse er viktigere enn beskrivelser: «Jeg gjenkjenner en A eller B. Det er taus kunnskap der som jeg ikke hadde da jeg sensurerte første gangen» (R4). Fordi besvarelsene er så ulike, vil ingen beskrivelse kunne erstatte det profesjonelle skjønnet:

> The expectation is that students come to know the standard as shaped by the assessors and context of assessment, not only what is represented in the written materials of assessment (Ajjawi et al., 2019, s. 6).

> «Standards' frameworks» are dynamic; they are constructed and reconstructed through involvement in communities and practices including engagement with student work, moderation and external examiners' feedback (Bloxham et al., 2016, s. 478).

Det faktum at flere av sensorene hadde liten eller ingen erfaring med sensur, økte sannsynligheten for misforståelser og feil, men dette kunne i beste fall bli avdekket ved sensurklage. Uavhengig av bakgrunn ble sensurveiledningen oppfattet som uferdig og utilstrekkelig. De yngre sensorene brukte den i noen grad, mens de eldre oftest støttet seg på egen erfaring. Dette harmonerer med funn i en studie av bedømmer-reliabilitet i essaybesvarelser (Asmyhr, 2011), der forfatteren konkluderte slik:

> Sensorene kommenterer at til tross for at eksplisitte kriterier var tilgjengelig, ble ikke disse anvendt i særlig grad. I stedet kunne vi identifiserte et holistisk utgangspunkt med subjektive og tause kriterier som basis for vurderingen (Asmyhr, 2011, s. 17).

Som tekst framstår sensurveiledningen som et kulturelt betinget og uferdig produkt: «Vi følger en mal som ble benyttet da jeg var emneansvarlig, og jeg benyttet den som ble benyttet før meg igjen. Å lage sensurveiledningen ble opplevd som krevende: «[. . .] spesielt å skille mellom de ulike karakterene. Det er stadig utfordringer, kanskje også fordi jeg benyttet en tidligere mal, og ikke fikk gjort meg tilstrekkelig erfaringer selv» (R1). Arbeid med sensurveiledningen var privatisert og ikke gjenstand for kritisk gransking: «Vi snakker aldri om det, med studentene, med kollegene, vi bare gjør det! Det er ikke det at vi tar lett på det, men vi snakker ikke om det for å oppnå en felles forståelse» (R1).

Sensurveiledning som tema inngår ikke i universitetets kvalitetssystem, og instituttet har ikke initiert forsknings- eller utviklingsarbeid på dette området, noe som også bekrefter funn fra en nasjonal undersøkelse: «Sensorordninger er forankret i utdanningssystemenes tradisjoner, historie og institusjonelle strukturer. Derfor tas de gjerne for gitt og diskuteres sjelden» (Frølich et al., 2009, s. 21). Dette er oppsiktsvekkende i lys av forarbeidene til kvalitetsreformen, der det heter at «[. . .] ekstern sensur representerer omtrent den eneste kvalitetssikringen av eksamen» (NOU 2000: 14, s. 610). I en tid med pålegg om kriteriebasert vurdering er det et paradoks at det ikke gis bedre veiledning i hvordan prinsippet best kan omsettes i praksis.

Diskusjon

Læringsutbyttebeskrivelsen var i vårt tilfelle ikke nevnt i sensurveiledningen, og dermed kan en bare spekulere om sammenhengen mellom de to dokumentene. Læringsutbytte som fenomen og begrep er dessuten problematisk fordi beskrivelsen ifølge retningslinjene beskriver minstekrav for bestått, uten konkretisering av mer avanserte kompetansenivå (Gynnild, 2017a). Myndighetenes generelle tiltro til sensurveiledning synes imidlertid urokket, uavhengig av fagdisiplin, kunnskapens art og sensorenes tolkning av kriteriene, jamfør følgende sitat fra Stortingsmelding 16 (2016–2017):

> Systematisk bruk av sensorveiledning [. . .] sikrer at ulike sensorer legger de samme retningslinjene til grunn for vurderingen, og at vurderingen gjøres opp mot det fastsatte læringsutbyttet [. . .] (Meld. St. 16 (2016–2017), s. 58–59).

Svak bedømmer-reliabilitet er imidlertid godt dokumentert, og forskning viser at sensurveiledning ikke nødvendigvis avhjelper situasjonen. (Gynnild, 2019). I realfag og teknisk-naturvitenskapelige emner er situasjonen noe annerledes ettersom kunnskapen oftest er av en mer eksakt karakter med riktige og gale svar. Innen tolkningsbaserte kunnskapsområder er bruken av profesjonelt skjønn og taus kunnskap nødvendige forutsetninger i forbindelse med helhetlig vurdering. Myndighetenes pålegg om kriteriebasert vurdering har imidlertid styrket initiativ for å bruke analytisk vurdering på bekostning av helhetlig og skjønnsbasert vurderingspraksis også innen mer kreative disipliner.

Et springende punkt er om sensurveiledning kan oppfattes som samme sak uavhengig av type emne og studieprogram. Det er nærliggende å anta at bruken av sensurveiledninger er enklere i teknisk-naturvitenskapelige emner sammenliknet med humaniora og samfunnsfag. En NOKUT-rapport fra studiefagene anatomi, fysiologi og biologi i sykepleierutdanning tjener til å illustrere dette. Her ble bedømmersamsvaret høyt ved at en «stor majoritet» av sensorparene ga samme poengsum til kandidatene på nesten alle oppgavene. Rapporten konkluderer med at dette i all hovedsak skyldtes oppgavesettets art og at sensorveiledningen var «god og tydelig» (Tokstad & Hamberg, 2017).

Nasjonale myndigheter nyanserer ikke bruken av sensurveiledning ut fra disiplinær egenart, og det underliggende, positivistiske kunnskapssynet blir ikke problematisert. Dette resulterer i at sensurutfordringenes egenart og kompleksitet underkommuniseres. Et eksempel på dette er forholdet mellom analytisk og holistisk vurdering, der sistnevnte prinsipp bygger på en tradisjon med større vekt på erfaring og kunnskapsdeling i kollegiale fellesskap, slik følgende sitat viser:

> Assessment consists in the exercise of an applied skill, and there are core aspects of this knowledge practice that cannot be captured by a mere propositional description of them, thus making them unavailable for publication (Bloxham et al., 2011a, s. 657).

Et eksempel på utfordringer ved bruken av sensurveiledningen i vårt tilfelle beskrives som følger:

> Sensorveiledningene er vel ikke spesifikke på annet enn det som er A og E. Det som er vanskelig er alt som er imellom, og det vet vi ingenting om. Der man trenger beskrivelser, der har man ingenting. Jeg savner en «bruksanvisning» for karakterene B, C og D! (R6).

Flere internasjonale forskere bestrider imidlertid at taus, eller implisitt kunnskap, fullt ut lar seg verbalisere (Kilpert & Shay, 2012; O'Donovan et al., 2008; Orr, 2007; Shay, 2005):

> The «hidden» and inexpressible nature of this tacit knowledge is compounded by the complex nature of work being assessed [. . .] which allows for a wide range of satisfactory student responses. [. . .] This requires tutors to use their judgement, based on their tacit knowledge [. . .] This is an 'interpretivist' view of assessment which recognises the power of the local context [. . .] (Bloxham et al., 2011a, s. 657).

Nasjonale retningslinjer for sensur benytter ikke begreper som «taus kunnskap» og «profesjonelt skjønn», trolig fordi disse assosieres med upålitelighet. Sadler hevder imidlertid at en standard er et abstrakt som ikke lar seg beskrive, men utgjør en del av et implisitt kunnskapsrepertoar. Etter dette kan «kriterium» kanskje oppfattes som et terskelbegrep, som hemmer eller forløser ny innsikt (Land et al., 2005). Hvis «terskelen» til ny innsikt ikke blir passert, forblir sensorene lett i etablerte tanke- og handlingsmønstre ut fra autoritetstro og lojalitet. Forestillingen om at alle fenomen kan verbaliseres og uttrykkes om påstandskunnskap er,

som vi flere ganger har vært inne på, ikke riktig. Evalueringsforskeren Sadler beskriver dette med mulig vidtrekkende konsekvenser:

> [. . .] achievement standards are, in essence, abstract concepts [and] cannot be reduced to set procedures that can be applied by non-experts. [. . .] They cannot be written down as detailed verbal descriptions, categories or lists which can then be used by students and assessors alike (Sadler, 2009a, s. 820).

Sadler, som tidlig i karrieren argumenterte for bruken av sensurveiledninger og karakterbeskrivelser, har senere radikalt endret sin forståelse og foreslått en annerledes praksis, som forklart nedenfor:

> [. . .] the starting point is an exemplar of a standard, not a verbal description of a standard. In determining the quality of complex student work, recognition – not definition – is the primary act. The purpose of a description is to account for a particular judgment. [. . .] The words form the necessary link between an immediate concrete referent and the abstract concept of quality (Sadler, 2009a, s. 821).

En NIFU-rapport fra 2009 konkluderte da også med at enesensur kontra sensur i fellesskap utgjør et hovedskille i kvaliteten (Frølich et al., 2009), noe som også bekreftes i en studie fra Storbritannia (Hannan & Silver, 2006). Respondentene i denne undersøkelsen støtter dette synet, og en mulig konsekvens kan være et annerledes kvalifiserende regime for sensuroppdrag. Håndverksfagene, som har likhetstrekk med kravene til sensurarbeid, kombinerer teori og praksis som obligatoriske søyler i utdanningen. Krav om to sensorer kunne gi et utmerket utgangspunkt for gode koplinger mellom erfarne og mindre erfarne sensorer på tvers av institusjoner. Plattformer for digital kommunikasjon har samtidig dramatisk redusert tids- og ressursbehov ved sensur på tvers av institusjoner.

Kalibrering av faglige standarder skjer best ved bruk av flere sensorer, men i vårt tilfelle var ikke dette et formelt krav. Bruken av ekstern sensor har i tillegg verdi for nasjonal kalibrering av standarder, som i sin tur kan bidra til mer enhetlig vurderingspraksis. Interbedømmer-reliabilitet er imidlertid i seg selv ikke et tilstrekkelig kvalitetskriterium. Et karakteristisk trekk ved internasjonale universiteter er bruken av en felles policy for vurdering og sensur. En empirisk studie ved fem internasjonale universiteter identifiserte følgende fem kriterier: pålitelighet, gyldighet, autentisitet, rettferdighet og transparens (Gynnild, 2017b). Disse har relevans i ulike faser av vurderingsarbeidet, spesielt for faglærere og tilsynssensorer, men de fleste av disse kriteriene er ikke nevnt i nasjonale retningslinjer. Sensurveiledning innen kreative disipliner løper også en risiko for forventningsstyrt sensur, som reduserer tolkningsrommet for kandidatene ved det faktum at besvarelser kan være ulike, men av samme kvalitet, noe som ikke er lett forenlig med ønsket om «tydelige» sensurveiledninger. Bruken av sensurveiledning er kun ett blant mange eksempler på en felleseuropeisk utvikling karakterisert av topptung administrativ styring i sektoren. En mulig forklaring ligger i bruken av New Public Management, som vektlegger ledelse og administrasjon som virkemidler for effektivisering av offentlig sektor (Bleiklie, 1998), mens forskningsbasert kunnskap i mindre grad ser ut til å prege initiativ for å fremme god kvalitet innen tema som er undersøkt i denne artikkelen.

Konklusjon

Denne undersøkelsen har aktualisert forholdet mellom generiske og emnespesifikke retningslinjer ved vurdering. Sensurveiledningen hadde ingen referanser til læringsutbyttebeskrivelsen,

og de generiske karakterbeskrivelsene ble tilpasset på måter som skapte uklarhet for sensorene. Relasjonen mellom eksamensoppgaver og sensurveiledning bidro trolig også til en urettferdig vurdering fordi noen av sensorene benyttet kriterier uten forankring i oppgavetekstene. Sensurveiledning ble benyttet på ulik måte og i ulik grad, men felles for alle sensorene var et ønske om mer kollegialt samarbeid. Dette ønsket var faglig begrunnet som bidrag til en mest mulig rettferdig og pålitelig sensur.

Relasjonene mellom eksamensoppgaver og sensurveiledning framstår som både en prinsipiell og praktisk utfordring. Et aktuelt spørsmål er om kriteriene i større grad kunne vært en integrert del av oppgavetekstene som en felles referanse både under skriving og sensur. Det virker uforståelig at kandidatene blir bedømt med delvis andre kriterier enn det som har vært gjort kjent for dem. Læring om læring og vurdering er tross alt en viktig kompetanse, og kanskje kunne omfanget av sensurklager vært redusert dersom kandidatene i større grad hadde innsikt i vurderingskriteriene.

Hensikten med studien er ikke å rette allmenn kritikk mot sensurveiledning som fenomen fordi dette kan være et nyttig redskap i egnete kontekster. Undersøkelsen har imidlertid avdekket utfordringer som oppstår når en sensurveiledning med et positivistisk utgangspunkt blir benyttet i et emne som åpenbart faller inn under et tolkningsbasert kunnskapsdomene. Ønsket om mer kollegialt basert sensur var et sterkt ønske, som illustrerer vurdering som en argumenterende praksis.

Referanser

Ajjawi, R., Bearman, M. & Boud, D. (2019). Performing standards: a critical perspective on the contemporary use of standards in assessment. *Teaching in Higher Education*, 1–14. doi: https://doi.org/10.1080/13562517.2019.1678579

Asmyhr, M. (2011). Om vurdering av essaybesvarelser i høyere utdanning – en studie av vurderer-reliabilitet. *Uniped*, *34*(4), 17–33. Hentet fra: http://www.idunn.no/uniped/2011/04/om_vurdering_av_essaybesvarelser_i_hoeyere_utdanning_en_st

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, *32*(3), 347–364. doi: https://doi.org/10.1007/bf00138871

Bjølseth, G., Havnes, A. & Lauvås, P. (2011). Lavt sensorsamsvar – kan det bedres? *Uniped*, *34*(4), 4–16. Hentet fra: http://www.idunn.no/uniped/2011/04/lavt_sensorsamsvar_kan_det_bedres

Bleiklie, I. (1998). Justifying the Evaluative State: New Public Management ideals in higher education. *European Journal of Education*, *33*(3), 299–316.

Bloxham, S. (2009). Marking and moderation in the UK: false assumptions and wasted resources. *Assessment and Evaluation in Higher Education*, *34*(2), 209–220. doi: https://doi.org/10.1080/02602930801955978

Bloxham, S., Boyd, P. & Orr, S. (2011a). Mark my words: the role of assessment criteria in UK higher education grading practices. *Studies in higher education (Dorchester-on-Thames)*, *36*(6), 655–670. doi: https://doi.org/10.1080/03075071003777716

Bloxham, S., Boyd, P. & Orr, S. (2011b). Mark my words: the role of assessment criteria in UK higher education grading practices. *Studies in Higher Education*, *36*(6), 655–670. doi: https://doi.org/10.1080/03075071003777716

Bloxham, S., den-Outer, B., Hudson, J. & Price, M. (2016). Let's stop the pretence of consistent marking: exploring the multiple limitations of assessment criteria. *Assessment & Evaluation in Higher Education*, *41*(3), 466–481. doi: https://doi.org/10.1080/02602938.2015.1024607

Eriksen, S. K. & Rasch, B. E. (2009). En eller to sensorer? *Tidsskrift for samfunnsforskning*, *50*(3), 293–312. Hentet fra: http://www.idunn.no/tfs/2009/03/art02

Frølich, N., Opheim, V., Brandt, S. & Prøitz, T. S. (2009). *Hva er viktige kvaliteter ved god sensur? En kartlegging av bruk av ekstern sensor på lavere grad med fokus på kvalitet, økonomi, vurdering og*

læring. Rapport 21, NIFU STEP. Hentet fra: https://nifu.brage.unit.no/nifu-xmlui/bitstream/handle/11250/279814/NIFUrapport2009-21.pdf?sequence=1&isAllowed=y

Gynnild, V. (2017a). Læringsmål eller læringsutbyttebeskrivelse? En empirisk, konstruktiv studie av begrepsbruken. *Norsk Pedagogisk Tidsskrift*, *101*(03), 225–238. doi: https://doi.org/10.18261/issn.1504-2987-2017-03-04

Gynnild, V. (2017b). Which are the Key Principles of Assessment? A Case Study of Policy Documents. *ICERI proceedings*, 0452–0457. Hentet fra: http://library.iated.org/publication_series/ICERI

Gynnild, V. (2019). «Blind» klagesensur ved Universitetet i Oslo: Pålitelig og rettferdig eller et steg i feil retning? *Uniped*, *42*(4), 340–354. Hentet fra: https://doi.org/10.18261/issn.1893-8981-2019-04-02

Hannan, A. & Silver, H. (2006). On being an external examiner. *Studies in Higher Education*, *31*(1), 57–69. doi: https://doi.org/10.1080/03075070500392300

Hendry, G., Armstrong, S. & Bromberger, N. (2012). Implementing standards-based assessment effectively: incorporating discussion of exemplars into classroom teaching. *Assessment & Evaluation in Higher Education*, *37*(2), 149–161. doi: https://doi.org/10.1080/02602938.2010.515014

Hsieh, H.-F. & Shannon, S. E. (2005). Three Approaches to Qualitative Content Analysis. *Qual Health Res*, *15*(9), 1277–1288. doi: https://doi.org/10.1177/1049732305276687

Kilpert, L. & Shay, S. (2012). Kindling fires: examining the potential for cumulative learning in a Journalism curriculum. *Teaching in Higher Education*, 1–13. doi: https://doi.org/10.1080/13562517.2012.678326

Land, R., Cousin, G., Meyer, J. H. F. & Davies, P. (2005). Threshold concepts and troublesome knowledge (3)*: implications for course design and evaluation. I C. Rust (red.), *Improving Student Learning Diversity and Inclusivity*. Oxford Centre for Staff and Learning Development.

Lauvås, P. & Jakobsen, A. (2002). *Exit eksamen – eller?: former for summativ evaluering i høgre utdanning*. Oslo: Cappelen akademisk forlag.

Lauvås, P. (2018). *Vurdering i skolen*. Cappelen Damm akademisk.

Meld. St. 16 (2016–2017). *Kultur for kvalitet i høyere utdanning*. Hentet fra: www.regjeringen.no

NOU 2000: 14. *Frihet med ansvar. Om høgre utdanning og forskning i Norge*. Hentet fra: http://www.regjeringen.no/nb/dep/kd/dok/nouer/2000/nou-2000-14.html

NOU 2020: 3. *Ny lov om universiteter og høyskoler*. Hentet fra: https://www.regjeringen.no/no/dokumenter/nou-2020-3/id2690294/

O'Donovan, B., Price, M., & Rust, C. (2008). Developing student understanding of assessment standards: a nested hierarchy of approaches. *Teaching in Higher Education*, *13*(2), 205–217. doi: https://doi.org/10.1080/13562510801923344

Orr, S. (2007). Assessment moderation: constructing the marks and constructing the students. *Assessment & Evaluation in Higher Education*, *32*(6), 645–656. doi: https://doi.org/10.1080/02602930601117068

Polanyi, M. (1998). The Tacit Dimension. In L. Prusack (Ed.), *Knowledge Organization.* Butterworth Heinemann.

Prop. 59 L (2013–2014). *Endringer i universitets- og høyskoleloven*. Hentet fra: https://www.regjeringen.no/no/dokumenter/Prop-59-L-20132014/id754603/?q=Prop%2059%20L%20(2013-2014)

Sadler, D. R. (2003). *How criteria-based grading misses the point?* Paper presented at the ETL Conference, Queensland College of Art, Griffith University, Australia.

Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment & Evaluation in Higher Education*, *30*(2), 175–194. Hentet fra: http://www.informaworld.com/10.1080/0260293042000264262

Sadler, D. R. (2009a). Grade integrity and the representation of academic achievement. *Studies in Higher Education*, *34*(7), 807–826. Hentet fra: http://www.informaworld.com/10.1080/03075070802706553

Sadler, D. R. (2009b). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education*, *34*(2), 159–179. Hentet fra: http://www.informaworld.com/10.1080/02602930801956059

Sadler, D. R. (2014). The futility of attempting to codify academic achievement standards. *Higher Education*, *67*(3), 273–288. doi: https://doi.org/10.1007/s10734-013-9649-1

Shay, S. (2005). The assessment of complex tasks: a double reading. *Studies in higher education (Dorchester-on-Thames)*, *30*(6), 663–679. doi: https://doi.org/10.1080/03075070500339988

Thomsen, O. B. (1975). *Embedsstudiernes universitet*. København: Akademisk forlag.

Tokstad, K. & Hamberg, S. (2017). *Nasjonal deleksamen – et pilotprosjekt og en mulighetsstudie*. Hentet fra: https://www.nokut.no/globalassets/nokut/rapporter/nasjonal-deleksamen7/nasjonal_deleksamen_et_pilotprosjekt_og_en_mulighetsstudie_2017.pdf